ECONOMICS OF THE

PUBLIC SECTOR

FOURTH EDITION

JOSEPH E. STIGLITZ
JAY K. ROSENGARD

ECONOMICS OF THE PUBLIC SECTOR

FOURTH EDITION

FOURTH EDITION

ECONOMICS OF THE PUBLIC SECTOR

JOSEPH E. STIGLITZ | JAY K. ROSENGARD

To our first teachers,
Nat and Charlotte
Jordan and Betty

BRIEF CONTENTS

CONTENTS

PREFACE

It has been more than a decade since the last revision of this textbook, and much has happened in the intervening years—two recessions, one from which we have yet to fully recover, two wars, a major health care reform, numerous tax reforms, and budget battles, one of which actually led to a government shutdown. Much has happened also in the development of the economics of the public sector. This edition incorporates this changing economic and intellectual landscape. I have been fortunate enough to be joined by Jay Rosengard, a long-term practitioner of the subject and teacher at Harvard's Kennedy School of Government, in writing this edition.

When the last edition published, I commented that I had been lucky to have been an active participant in many of these changes, as member and chairman of the President's Council of Economic Advisers. That edition was written from the unique perspective of a public sector economist who had the chance to be involved in the decision-making process not only in the United States, but also in many other countries. Following my stint at the White House, I worked as chief economist and senior vice president of the World Bank, which is involved in advising developing countries concerning their public sector policies. Since then, I have continued with my passionate involvement in debates about public sector policies, as an adviser to many governments and as member or chair of several international commissions. I have been able to participate in debates around the world on the central questions with which this book is concerned: What should be the role of government? How should it design its programs in areas ranging from health and education, to Social Security and welfare? How should tax systems be designed to promote economic efficiency and be consistent with basic views of fairness?

In one sense, my experiences confirmed many of the views and approaches I had developed in previous editions. Indeed, it gave me great pleasure to see the extent to which the ideas and perspectives, many of which seemed so new when they were presented in the first edition of this book, were being integrated into thinking about policy, not only in the United States, but throughout the world. I have become increasingly convinced that the kind of analysis presented in this book can—and has—significantly improve the formation of policies in the public sector.

The economics of the public sector is a subject that is always in flux. While there are some general principles that are as applicable today as they were two decades ago, new issues have risen to the top of the policy agenda and old issues have waned in importance. Debates today often hinge on different questions than they did even a decade ago. Even the language in which some of the debates are couched has changed. We have tried in this edition not only to incorporate the many changes in expenditure policies and tax laws, but also to reflect some of these changing approaches and themes. There is, for instance, an increased emphasis on understanding why government is often inefficient, and on improving the efficiency of government—to use the phrase popularized by the Clinton administration—to "reinvent government." Growing inequality, especially in the United States, but also in most other countries around the world, has rightly become a subject of concern, and this book reflects these concerns, with special attention given to the distributive consequences of different policies.

Our major aim in writing this edition remains the same as when I wrote the first: the belief that an understanding of the issues addressed in this book is central to any democratic society. Among the most important of these are the appropriate balance between the public and private sectors, the ways in which the public and private sectors can complement each other, and how governments can more effectively meet their objectives. Issues in public sector economics often become highly charged politically, but we tried to present the analysis in an impartial manner, with a clear delineation between the analysis of the consequences of any policy and the value judgments associated with assessing the desirability of the policy. We have tried to be clear about what economic theory and empirical research had to say *on all sides* of the debate, identifying where—and why—there is frequent uncertainty about the outcomes of certain policies, and clarifying why disagreements about the desirability of different policies persist. In this edition, we continue with the commitment to present to the student a balanced account of these often heated debates. The favorable reception of the previous editions by instructors of a wide variety of political persuasions suggests that we have succeeded

in doing so. And the publication of many foreign editions of the book, in countries ranging from Russia, China, Japan, Germany, Italy, and Spain, to Latvia, Turkey, and the Czech Republic has shown that the approach has met with favor not just in the United States, but in countries facing quite different circumstances and problems.

My experiences, at the White House, in the World Bank, and in dialogues around the world have made me even more convinced of the importance of this endeavor. Democracies can only succeed if there are meaningful public debates on the central public policy issues of the day. Too often, in too many countries, good policies flounder because of a lack of widespread understanding of basic economic issues. Writing an undergraduate textbook such as this thus present both a great challenge and a great opportunity: the challenge to present complex and complicated ideas in simple enough terms that they can be understood by someone with a relatively limited background in economics (at the most, a single year of a principles course); and the opportunity, if one succeeds in doing so, to influence the ways in which public policy debates are approached.

Public sector issues are some of the most exciting in all of economics. Health, defense, education, Social Security, welfare programs, and tax reform all receive steady attention in the news media, and economic analysis brings special insights to the debates. Should education be publicly provided? What is the long-term outlook for our Social Security program? How do current proposals for tax reform match our knowledge of who really bears the tax burden? What determines the efficiency and equity consequences of various taxes? These kinds of questions breathe life into the course, which is why we give them careful attention.

Examining specific tax and expenditure programs offers an additional benefit: it underscores the importance of design features. One of the lessons we learned in the past decade is that good intentions are not enough. There are numerous examples where legislation has not been successful in achieving its objectives, and in which there are often unintended consequences. For instance, the 2001 and 2003 tax reforms, counter to their intentions, may have actually led to less investment. We use examples like these not only to enliven the course, but also to instill in students the important habit of testing theories against the complex environment in which public sector decisions are enacted and implemented.

The organization of this book is based on the principle of flexibility. The sequence we follow is to introduce in Parts 1 and 2, the fundamental questions, institutional details, and a review of the microeconomic theory underlying the role of the public sector. Part 3 develops the theory of public expenditures, including public goods, public choice, and bureaucracy, while Part 4 applies the theory of the five largest areas of public

expenditure in the United States: health, defense, education, Social Security, and welfare programs. Parts 5 and 6 repeat this pattern, presenting the theory of taxation and its analysis, respectively. Part 7 takes up two additional topics: issues concerning state and local taxation and expenditure and fiscal federalism; and issues concerning fiscal policy, with particular emphasis on the relation between microeconomic analysis and macroeconomic performance. The ups and downs of the deficit are among the major changes we confront in this book. In the first two editions, deficits were at the center of attention, but then, in the third edition, the deficits changed to surpluses. As I noted then, there was a risk that this change would be temporary, and so it was, and deficits are once again part of the economic debate. In this edition, we try to come to an understanding of these marked fluctuations in the U.S. deficits.

A perfectly workable alternative to this sequence would be to cover taxation before expenditures. Parts 5 and 6 have been carefully developed so that instructors wishing to go straight to taxation after Part 1 can do so without losing continuity. Further tips on how courses can be organized, as well as lecture notes, test questions, and coverage of advanced topics that instructors may wish to include in their lectures are in the *Instructor's Manual*.

The list of those to whom I am indebted is a long one. My teachers at Amherst College, James Nelson and Arnold Collery, not only stimulated my interest in economics and in the particular subject of this course, but also laid the foundations for my later studies. They also showed me, by example, what good teaching meant. I hope that some of what I learned from them is reflected in this book. At M.I.T., Dan Holland and E. Cary Brown introduced me to the formal study of public economics. Again, I hope some of the blend of policy, theory, and institutional detail that marked their work is reflected here. The insights of my colleagues and collaborators at the institutions at which I have worked (M.I.T., Yale University, Stanford University, Princeton University, Oxford University, Cambridge University, and the National Bureau of Economic Research) and the government agencies (Council of Economic Advisers, Treasury, Labor, Interior, Energy, Agency for International Development, State of Louisiana, State of Texas) and international organizations (World Bank, Interamerican Development Bank, Organization of Economic Cooperation and Development) for which I have worked and consulted have also proved invaluable. I should mention Henry Aaron (Brookings Institution), Alan J. Auerbach (Berkeley), Greg Ballantine (former Assistant Secretary of the Treasury for Tax Policy), William Baumol (Princeton University), Charles T. Clotfelter (Duke University), Partha Dasgupta (Cambridge University), Peter A. Diamond (M.I.T.), Avinash Dixit (Princeton University), Martin Feldstein

In that effort, I was aided by a group of outstanding graduate research assistants from University of Maryland: Amy Harris, Kosali Ilayperuma, Steven Karon, and Diana Stech. I also owe a special thanks to Janet McCubbin for her contributions on the first draft of the Third Edition. Janet read each word of the manuscript, and her invaluable feedback put the revision on a firm footing from the get-go.

For this fourth edition, I also wish to thank Eamon Kircher-Allen for his assistance in shepherding this edition to completion.

Finally, we are deeply indebted to the fine people at W. W. Norton and Company, a truly outstanding publishing firm, who brought this project to fruition. For this edition, I owe special thanks to Jack Repcheck for his excellent guidance, as well as the rest of the superb team at Norton: Sujin Hong, Theresia Kowara, Vanessa Nuttry, Marina Rozova, Carson Russell, and Stefani Wallace. I also remain eternally grateful to Ed Parsons, Claire Acher, Kate Barry, Joan Benham, Margaret Farley, Roseanne Fox, and Mark Henderson for their excellent work on earlier editions.

J.E.S.
Fall 2014

ROLE AND SIZE OF THE PUBLIC SECTOR

At the center of any country's political life are some basic economic questions: How does the government affect the economy? What is the appropriate role and size of government? Why are some economic activities undertaken in the public sector and others in the private sector? Should government do more than it is currently doing, or less? Should government change what it is doing, and how it is doing it?

To answer these questions, we must begin by understanding what the government does today. How have governments changed over time? How do the size and scope of government in one country compare with those of other countries? What might explain these differences?

Part One provides this context for public sector economics. Chapter 1 gives an overall perspective on the economic role of government. It sets forth the basic questions that are addressed by public sector economists, and explains some of the reasons why there are disagreements among them about appropriate policies. Chapter 2 addresses challenges in measuring the size of the public sector and provides comparative data on the magnitude of the public sector around the world today.

1

DEFINING PUBLIC SECTOR RESPONSIBILITIES

FOCUS QUESTIONS

1. What are the central questions concerning the economics of the public sector?

2. What are the differing views concerning the economic role of government? How have they changed over the years and what has given rise to those changes?

3. How do economists go about studying the economics of the public sector?

4. What are the principal sources of disagreement among economists about appropriate policies that government should pursue?

From birth to death, our lives are affected in countless ways by the activities of government. For example, in the United States:

- We are born in hospitals that are publicly subsidized, if not publicly owned. Our arrival is then publicly recorded (on our birth certificates), entitling us to a set of privileges and obligations as American citizens.
- Most of us (almost 90 percent) attend public schools.
- Virtually all of us, at some time in our lives, receive money from the government, through programs such as student loans, unemployment or disability payments, antipoverty programs, Social Security, and Medicare.
- All of us pay taxes to the government—sales taxes; taxes on such commodities as gasoline, liquor, telephones, air travel, perfumes, and tires; property taxes; income taxes; and Social Security (payroll) taxes.
- More than a sixth of the work force is employed by the government, and for the rest, the government has a significant impact on employment conditions.
- In many areas of production—be it cars, sneakers, or computers—profits and employment opportunities are greatly affected by whether

the government allows foreign competitors to sell goods in America without a tariff or quota.

- What we eat and drink, where we can live, and what kinds of houses we can live in are all regulated by government agencies.

- We travel on public roads and publicly subsidized railroads. In most communities our garbage is collected and our sewage is disposed of by a public agency; in some communities the water we drink is provided by public water companies.

- Our legal structure provides a framework within which individuals and firms can sign contracts with one another. When there is a dispute between two individuals, the two may turn to the courts to adjudicate the dispute.

- Without environmental regulations, many of our major cities would be choked with pollution, the water of our lakes and rivers would be undrinkable, and we could neither swim nor fish in them.

- Without safety regulations, such as those requiring seat belts, highway fatalities would be even higher than they are.

THE ECONOMIC ROLE OF GOVERNMENT

Why does government engage in some economic activities and not others? Why has the scope of its activities changed over the past hundred years, and why does it have different roles in different countries? Does the government do too much? Does it do well what it attempts to do? Could it perform its economic role more efficiently? These are the central questions with which the economics of the public sector is concerned. To address them, we will first consider the economic role of government in modern economies, how ideas about the role of government have emerged, and the changing role of government in the twenty-first century.

THE MIXED ECONOMY OF THE UNITED STATES

The United States has what is called a **mixed economy**, where many economic activities are undertaken by private firms, while others are undertaken by the government. In addition, the government alters the behavior of the private sector through a variety of regulations, taxes, and subsidies.

By contrast, in the former Soviet Union, most economic activities were undertaken by the central government. Today, only North Korea and Cuba give the government such primacy. In many western European economies, national governments have had a larger role in economic activity than in the United States. For instance, the government of France once participated in a range of economic activities, including the production of cars, electricity, and airplanes. Since the 1980s, however, **privatization**—converting government enterprises into private firms—has been the trend in Europe, although the economic role of government generally remains larger there than in the United States.[1] Some of these privatizations have been far less successful than hoped. In the United States, President Bush proposed partially privatizing Social Security, but in the aftermath of the 2008 financial crisis, support for such efforts vanished, as Americans realized the magnitude of the losses—and the insecurity that they would have faced—had that initiative proceeded. Indeed, around the world, the crisis brought on the most significant *increase* in the role of the government in more than half a century, with governments taking over or providing massive subsidies to the financial sector and a host of other industries.

The origins of the mixed economy of the United States lie in the origins of the country itself. In formulating the United States Constitution, the founders of the republic had to address explicitly key issues concerning the economic role of the new government. The Constitution assigned the federal government certain responsibilities, such as running the post office and printing money. It provided the foundations for what we now call "intellectual property rights" by giving the government the right to grant patents and issue copyrights to encourage innovation and creativity. It gave the federal government certain rights to levy taxes, although those did not include taxes on exports, income, or net wealth. Most importantly, for the future evolution of the country, Article 1, Section 8, Clause 3 gave the federal government the right to regulate interstate commerce. Because so much of economic activity involves goods produced in one state and sold in another, this clause, interpreted broadly, has been used to justify much of the federal government's regulatory activities.

Throughout the history of the United States, the economic role of the government has undergone important changes. For instance, one hundred years ago some highways and all railroads were private; today, there are no major private roads and most interstate railroad passenger travel is by Amtrak, a publicly owned and subsidized enterprise. It is because mixed economies constantly face the problem of defining the appropriate

---

[1] For more on the case of France, see H. Dumez and A. Jeunemaitre, "Privatization in France: 1983–1993," in *Industrial Privatization in Western Europe: Pressures, Problems, and Paradoxes*, ed. Vincent Wright (London and New York: Pinter Publishers, 1994), pp. 83–105, 194.

boundaries between government and private activities that the study of the economics of the public sector in these countries is both so important and so interesting.

DIFFERENT PERSPECTIVES ON THE ROLE OF GOVERNMENT

To better understand contemporary perspectives on the economic role of government, it can be helpful to consider the different perspectives that have evolved in the past.[2] Some of the central ideas of the eighteenth and nineteenth centuries have been critical to economic history in the twentieth century, and continue to be important today.

One dominant view in the eighteenth century, which was particularly persuasive among French economists, was that the government should actively promote trade and industry. Advocates of this view were called **mercantilists**. It was partly in response to the mercantilists that Adam Smith (who is often viewed as the founder of modern economics) wrote *The Wealth of Nations* (1776), in which he argued for a limited role for government. Smith attempted to show how competition and the profit motive would lead individuals—in pursuing their own private interests—to serve the public interest. The profit motive would lead individuals, competing against one another, to supply the goods other individuals wanted. Only firms that produced what was wanted and at as low a price as possible would survive. Smith argued that the economy was led, as if by an invisible hand, to produce what was desired—and in the best possible way.

Adam Smith's ideas had a powerful influence both on governments and on economists. Many of the most important nineteenth-century economists, such as the Englishmen John Stuart Mill and Nassau Senior, promulgated the doctrine known as **laissez faire**. In their view, the government should leave the private sector alone; it should not attempt to regulate or control private enterprise. Unfettered competition would serve the best interests of society.

Not all nineteenth-century social thinkers were persuaded by Smith's reasoning. The grave inequalities in income that they saw around them, the squalor in which much of the working classes lived, and the unemployment that workers frequently faced concerned them. While nineteenth-century writers like Charles Dickens attempted to portray the plight of the working classes in novels, social theorists, such as Karl Marx, Jean Charles Léonard de Sismondi, and Robert Owen, developed theories that not only attempted

---

[2] See A. O. Hirschman, *Shifting Involvements: Private Interest and Public Action* (Princeton, NJ: Princeton University Press, 1982). Hirschman has put forth an interesting theory attempting to explain the constant changes in views on the appropriate role of the government.

to explain what they saw but also suggested ways in which society might be reorganized. Many attributed the evils in society to the private ownership of capital; what Adam Smith saw as a virtue, they saw as a vice. Marx, if not the deepest of the social thinkers, was certainly the most influential among those who advocated a greater role for the state in controlling the means of production. Still others, such as Owen, saw the solution neither in the state nor in private enterprise, but in smaller groups of individuals getting together and acting cooperatively for their mutual interest.

On one hand, private ownership of capital and unfettered free enterprise; on the other, government control of the means of production—these contrary principles were to become a driving force for international politics and economics in the twentieth century, embodied in the Cold War. Today, the countries of the former Soviet Union and the Eastern bloc are in the midst of a monumental transition to market systems—a fundamental transformation of government's role in those economies. In the United States, the economic role of government has also changed, but the changes have arisen more gradually, in response to economic events throughout the century. There is now widespread agreement that markets and private enterprises are at the heart of a successful economy, but that government plays an important role as a complement to the market.

The precise nature of that role, however, remains a source of contention. It differs both between countries and within nations over time, depending largely on society's expectations for government and what the members of a society are willing to pay to meet these expectations, sometimes referred to as a "social compact" or "social contract." For example, the citizens of countries in northern Europe generally expect their governments to provide health, education, and social services that are provided largely by the private sector in the United States, and are willing to pay relatively higher taxes to finance these public services. Several of these countries have succeeded in creating public health care systems that deliver better health outcomes at much smaller costs than the largely private American system. The debate over the appropriate role of the government took a sharp turn in 2008, when it became evident that only the government could save the economy from an economic crisis that had largely been created by private markets.

AN IMPETUS FOR GOVERNMENT ACTION: MARKET FAILURES

Prior to the 2008 crisis, the Great Depression—in which the unemployment rate reached 25 percent and national output fell by about a third from its peak in 1929—was the event that most fundamentally changed

attitudes toward government. There was a justified widespread view that markets had failed in an important way, and there were enormous pressures for government to do something about this market failure. The great English economist John Maynard Keynes, writing in the midst of the Great Depression, argued forcefully that the government not only should do something about economic slumps, but also that it could. The belief that governments should and could stabilize the level of economic activity was eventually embedded in legislation in the United States, in the **Full Employment Act of 1946**, which at the same time established the Council of Economic Advisers, to counsel the President on how best to accomplish these objectives.

The economy's seeming inability to provide jobs was not the only problem that drew attention. The depression brought to the fore problems that, in less severe form, had been there for a long time. Many individuals lost virtually all their money when banks failed and the stock market crashed. Many elderly people were pushed into dire poverty. Many farmers found that the prices they received for their products were so low that they could not make their mortgage payments, and defaults became commonplace.

In response to the depression, the federal government not only took a more active role in attempting to stabilize the level of economic activity, but also passed legislation designed to alleviate many of the specific problems: unemployment insurance, Social Security, federal insurance for depositors, federal programs aimed at supporting agricultural prices, and a host of other programs aimed at a variety of social and economic objectives. Together, these programs are referred to as the New Deal.

After World War II, the country experienced an unprecedented level of prosperity. However, it became clear that not everyone was enjoying the fruits of that prosperity. Many individuals, by the condition of their birth, seemed to be condemned to a life of squalor and poverty; they received inadequate education, and their prospects for obtaining good jobs were bleak.

These inequities provided the impetus for many of the government programs that were enacted in the 1960s, when President Lyndon B. Johnson declared his "War on Poverty." Whereas some programs were aimed at providing a "safety net" for the needy—for instance, programs to provide food and medical care to the poor—others, such as job retraining programs and Head Start, which offers preschool education for underprivileged children, were directed at improving the economic opportunities of the disadvantaged.

Could government actions alleviate these problems? How was success to be gauged? The fact that a particular program did not live up to

the hopes of its most enthusiastic supporters did not, of course, mean that it was a failure. Medicaid, which provides medical assistance to the indigent, was successful in narrowing the differences in access to medical care between the poor and the rich, but the gap in life expectancy between these two groups was not eliminated. Medicare, which provides medical care for the elderly, relieved the elderly and their families of much of the anxiety concerning the financing of their medical expenses, but it left in its wake a national problem of how to finance rapidly increasing medical expenditures. Even though the Social Security program provided the aged with an unprecedented level of economic security, it too has run into financial problems that some critics say have cast doubt on whether future generations will be able to enjoy the same benefits.

Fifty years after the War on Poverty began, poverty has not been eradicated from America. Government programs have significantly reduced poverty from what it otherwise would have been, but both critics and supporters of the government's programs agree that we still have considerable challenges to overcome if we are to eliminate poverty, and that good intentions often have unintended negative consequences. Many programs designed to alleviate the perceived inadequacies of the market economy have had effects markedly different from those their proponents anticipated. Urban renewal programs designed to improve the quality of life in inner cities have, in some instances, resulted in the replacement of low-quality housing with high-quality housing that poor people cannot afford, thus forcing them to live in even worse conditions. Homelessness has become an increasing concern. Although many programs designed to promote integration of public schools have succeeded, because of residential segregation, public schools in some districts are no better integrated than private schools. A disproportionate share of the benefits of farm programs has accrued to large farms; government programs have not enabled many of the small farms to survive.

Supporters of continued government efforts claim that critics exaggerate the failures of government programs. They argue that the lesson to be learned is not that the government should abandon its efforts to solve the major social and economic problems facing the nation, but that greater care must be taken in the appropriate design of government programs. In other words, the limitations of government, or "government failures," should not prevent the government from trying to mitigate market failures.

More recently, attention has shifted to two other market failures: excessive volatility, evidenced by the crisis of 2008 and by more than a hundred other crises around the world since 1980 when the era of deregulation began, and growing inequality, accompanied by a decline

in economic opportunity. These problems are prevalent in many other countries, but they have become a special focus of attention in the United States, which now has the highest level of inequality among the advanced industrial countries and one of the lowest levels of "opportunity"; that is, an American child's life prospects are more dependent on the income and education of his or her parents than in other advanced countries. Even worse, some of the wealth at the top (in particular, that associated with financial markets) has come as a result of exploitation of those at the bottom, through predatory and discriminatory lending, abusive credit card practices, and exploitation of market power. These pervasive concerns about the market suggest that there is an important role for government; but in some quarters, there is the worry that government has not only failed to "correct" the market failures, but it may also have actually contributed to the problems.

ACHIEVING BALANCE BETWEEN THE PUBLIC AND PRIVATE SECTORS

Markets often fail, but governments often do not succeed in correcting the failures of the market. Today, economists, in ascertaining the appropriate role of government, attempt to incorporate an understanding of the limitations of both government and markets. There is agreement that there are many problems that the market does not adequately address; more generally, the market is fully efficient only under fairly restrictive assumptions (see Chapters 3 and 4).

THE MIXED ECONOMY

- The United States is a mixed economy, in which both the public and private sectors play an important role.
- The roles played by government—and views concerning what they should be—have changed markedly over time.
- An important motivation for government's undertaking certain activities is actual or perceived failures of the market.
- There has been increasing recognition of the limitations of government—of "government failures" as well as market failures.

The recognition of the limitations of government, however, implies that government should direct its energies only at areas in which market failures are most significant *and* where there is evidence that government intervention can make a significant difference. Among American economists today, the dominant view is that *limited* government intervention could alleviate (but not solve) the worst problems; thus, the government should take an active role in maintaining full employment and alleviating the worst aspects of poverty, but private enterprise should play the central role in the economy. The prevalent view attempts to find ways for government and markets to work together, each strengthening

the other. For instance, ensuring that governments rely more heavily on markets and market-like mechanisms.

Controversy remains, though, over how limited or how active the government should be, with views differing according to how serious one considers the failures of the market to be and how effective one believes government is in remedying them. Typically, economists who have served Democratic administrations believe that there is a larger role for government to play; those who have served Republican administrations are more doubtful. Part of the disagreement arises from the importance they attach to market failures; part of it arises from the importance they attach to inequality—even when markets are efficient, even pro-market economists agree that markets may lead to an unacceptably high level of poverty. However, much of the difference arises from politics: their assessment of the ability or likelihood that government will effectively deal with the market failure, without creating problems of its own. Some of those who agree that the market produces "too much" inequality believe that the costs of even the most efficient efforts to reduce poverty are too great, whereas others believe that, in practice, government efforts will prove ineffective.

THE EMERGING CONSENSUS

As important as they are, the differences in views of government's economic role are far smaller than the differences a hundred years ago, when socialists advocated a dominant role for government and laissez-faire economists advocated no role for government at all. Contemporary rethinking of the role of government has been reflected in two concurrent initiatives, **deregulation** and privatization.

The first, begun in the United States under President Carter, reduced the role of government in regulating the economy. For instance, the government stopped regulating prices for airlines and long-distance trucking. While there is recognition that regulations have a cost, there is increasing awareness that *not regulating* may have even greater costs. Regulations have continued to grow, partly in response to the growing recognition of market failures, such as those associated with environmental degradation and the near collapse of the banking system twice in the past three decades. The Clinton administration, in its "Reinventing Government" initiative, sought a balance: while recognizing the need for regulation, it also recognized that some regulations were overly burdensome, their benefits were less than their costs, and there might be more effective ways of obtaining the desired objectives. Major reforms were instituted in such areas as banking, telecommunications, and electricity. In some of these

areas, such as telecommunications, it was hoped that, with new technologies, the scope for competition would be larger than had previously been thought. Parallel reforms occurred throughout the world.

In some cases, including the United States, the enthusiasm for deregulation seems to have been carried too far. The economic crisis in East Asia in 1997—as the savings and loan debacle in the United States had done a decade earlier and the global economic crisis did in 2008—brought home the importance of financial market regulation. These crises ultimately end up costing workers, businesses, and even taxpayers dearly. So, too, the hoped-for competition in telecommunications did not emerge.

Like the Clinton administration, the Obama administration claims to be striving to find an appropriate balance. Critics worry whether, like the Clinton administration, it has come too much under the influence of the special interests that benefited from deregulation or of the ideologies that place excessive confidence in the marketplace.

The second initiative, privatization, sought to turn over to the private sector activities previously undertaken by government. The privatization movement was much stronger in Europe, where telephones, railroads, airlines, and public utilities were all privatized. In the United States, because government ran few enterprises, there was much less scope for privatization. Perhaps the most important, and controversial, privatization was that of the United States Enrichment Corporation, the government agency responsible for enriching uranium. (Low-enriched uranium is used in nuclear power plants; highly enriched uranium is used to make atomic bombs. The same process and plants are used to make both.) The privatization, which was approved in 1997 and completed in 1998, raised profound implications for U.S. national security. For instance, it complicated subsequent nuclear disarmament discussions because of conflicts of interest between the privatized firm and national security. To many, this privatization appeared to be a case of the ideology of privatization gone amok—government had lost the sense of balance between the private and public sector required to make a mixed economy work.

As recently as 2005, however, there was a major privatization effort in the United States: to privatize a substantial part of the public old age retirement program (Social Security), following similar efforts in Chile, the United Kingdom, and other countries. The global financial crisis exposed major problems with these initiatives, including increased insecurity for the elderly. Other privatization efforts, such as roads in Mexico and railroads in the United Kingdom, similarly encountered major problems. With the 2008 crisis, many governments were forced to take a much more active role in the economy, in some cases nationalizing, or renationalizing, private enterprises (especially banks). In many resource-rich countries, private oil

and mining companies had driven such unfair bargains that governments demanded new contracts, or again, in some cases, the nationalization, or even renationalization, of the oil or gas fields or mines.

THINKING LIKE A PUBLIC SECTOR ECONOMIST

Economists study *scarcity*—how societies make choices concerning the use of limited resources. They inquire into four central economic questions:

1. What is to be produced?

2. How is it to be produced?

3. For whom is it to be produced?

4. How are these decisions made?

Like all economists, public sector economists are concerned with these fundamental questions of choice, but their focus is the choices made within the public sector, the role of the government, and the ways government affects the decisions made in the private sector.

1. *What is to be produced?* How much of our resources should be devoted to the production of public goods, such as defense and highways, and how much of our resources should we devote to the production of private goods, such as cars, TV sets, and video games? We often depict this choice in terms of the **production possibilities schedule**, which traces the various amounts of two goods that can be produced efficiently with a given technology and resources. In our case, the two goods are public goods and private goods. Figure 1.1 gives the various possible combinations of public goods and private goods that society can produce.

   Society can spend more on public goods, such as national defense, but only by reducing what is available for private consumption. Thus, in moving from *G* to *E* along the production possibilities schedule, public goods are increased, but private goods are decreased. A point such as *I*, which is below the production possibilities schedule, is said to be *inefficient:* society could get more public goods and more private goods. A point such as *N,* which is above the production possibilities schedule, is said to be *infeasible:* it is not possible, given current resources and technology, to have that quantity of public goods and that quantity of private goods at the same time.

FIGURE 1.1

SOCIETY'S PRODUCTION POSSIBILITIES SCHEDULE
This depicts the maximum level of private goods that society can enjoy for each level of public goods. If society wishes to enjoy more public goods, it must give up some private goods.
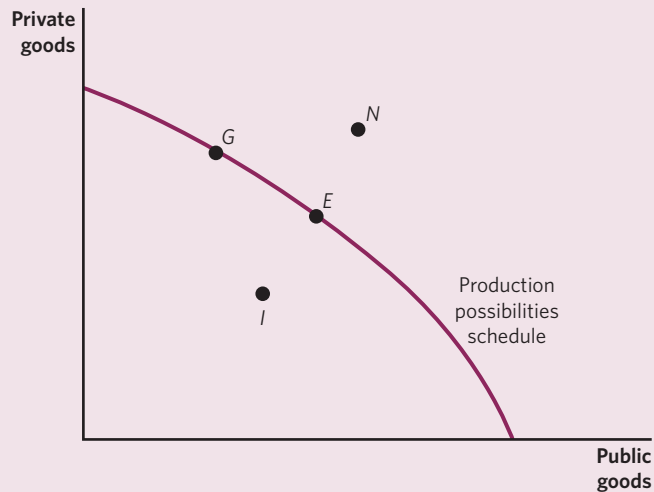
2. *How should it be produced?* Under this question are subsumed such decisions as whether to produce privately or publicly, to use more capital and less labor or vice versa, or to employ energy-efficient technologies.

Other issues are also subsumed under this second question. Government policy affects how firms produce the goods they produce: environmental protection legislation restricts pollution by firms; payroll taxes that firms must pay on the workers they employ may make labor more expensive and thus discourage firms from using production techniques that require much labor.

3. *For whom is it to be produced (the question of distribution)?* Government decisions about taxation or welfare programs affect how much income different individuals have to spend. Similarly, the government must decide what public goods to produce. Some groups will benefit from the production of one public good, others from another.

4. *How are choices made?* In the public sector, choices are made *collectively*. Collective choices are the choices that a society must make together—for instance, choices concerning its legal structure, the size of its military establishment, its expenditures on other public goods, and so on. Texts in other fields of economics focus on how individuals make their decisions concerning consumption, how firms make their decisions concerning production, and how the price system works to ensure that the goods demanded by consumers are produced by firms.

Collective decision making is far more complicated, for individuals often disagree about what is desirable. After all, just as some individuals like chocolate ice cream and some like vanilla ice cream, some individuals get greater enjoyment out of public parks than do others. But whereas with private goods the individual who likes chocolate ice cream can simply buy chocolate ice cream and the individual who likes vanilla ice cream can buy vanilla ice cream, with public goods we must make decisions together. Anyone who has lived in a family knows something about the difficulties of collective decision making (should we go to the movies or go bowling?).

Public decision making is far more complex. Increasingly, though, we have come to understand that decision making in the private sector, especially in large corporations, is far more complex than depicted in simplistic models of firms that have a single owner. Within the corporation, there are large differences in views about what should be done, partially motivated by judgments about the consequences of different actions (how well will a product sell?), but also by differences in "values"—the extent, for instance, to which the firm should focus on the short term or the long. One of the objectives of public sector economics is to study how collective choices (or, as they are sometimes called, social choices) are made in democratic societies.

The recognition of this divergence of views is important in itself. It should make us wary of expressions such as "It is in the public interest" or "We are concerned with the good of society." Different policies may be good for different individuals. One should carefully specify who will benefit from a given policy and who will be harmed by it.

KEY ECONOMIC QUESTIONS

- What is produced?
    Public or private goods?
- How is it produced?
    Within the public sector or the private?
- For whom should it be produced?
    Taxes affect amount different individuals have to spend.
    Different government programs benefit different groups.
- How are these decisions made?
    How are collective decisions, such as those concerning the supply of publicly provided goods and taxes, made?

ANALYZING THE PUBLIC SECTOR

In addressing each of the fundamental economic questions, there are four general stages of analysis: *describing* what the government does, *analyzing* the consequences of government action, *evaluating* alternative policies, and *interpreting* the political forces that underlie the decisions government makes.

1. *Knowing what activities the public sector engages in and how these are organized.* The complexity of the government's operations is so great that it is difficult to assess what its total expenditures are and what they go for. The budget of the U.S. federal government alone is a document that is more than 1,000 pages long, and within the budget, activities are not easily compartmentalized. Some activities are undertaken in several different departments or agencies. Research, for instance, is funded through the Department of Defense, the National Science Foundation, the National Institutes of Health, and the National Aeronautics and Space Administration, among others. Furthermore, a department like the Department of Health and Human Services undertakes a myriad of activities, some of which are only vaguely related to others.

   Further, taxes and expenditures occur at several different levels: in some places, individuals pay not only federal and state taxes but also separate taxes to their school district, their township, their county, the jurisdictions that provide their water and sewage, and their public library.

2. *Understanding and, insofar as possible, anticipating the full consequences of these governmental activities.* When a tax is imposed on a corporation, who bears the tax? At least part of the tax will be passed on to consumers through higher prices, or on to employees as wages fall. What are the consequences of the government's changing the age of retirement for Social Security? Of a tax credit or deduction for college tuition? Will universities respond by raising tuition so a college education will be hardly more affordable than before? Will a tax on the global income of American corporations (eliminating the provision that allows them to postpone taxes until they bring their income back to the United States) reduce their incentive to outsource jobs and increase government revenue? Or will it simply encourage corporations to move their headquarters abroad?

   The consequences of government policies are often too complicated to predict accurately, and even after a policy has been introduced, there is often controversy about what its effects are. This book attempts not only to present all sides of some of the major controversies, but also to explain why such disagreements have persisted, and why they are difficult to resolve.

3. *Evaluating alternative policies.* To do this, we need not only to know the consequences of alternative policies, but also to develop criteria for evaluation. First we must understand the objectives of government policy, and then we must ascertain the extent to which a particular proposal meets (or is likely to meet) those criteria.

Many government programs have multiple objectives. For example, the United States has a program to clean up hazardous waste sites, not only to protect health, but also because such sites may be an impediment to economic development. Some policies are better at achieving one objective; others may be better at achieving others. We need a framework for decision making in such situations: How do we think systematically about the trade-offs in evaluating alternative policies?

> ## ANALYZING THE PUBLIC SECTOR
>
> - Knowing what activities the public sector engages in and how they are organized
> - Understanding and anticipating the full consequences of these government activities
> - Evaluating alternative policies
> - Interpreting the political process

4. *Interpreting the political process.* Collective decisions such as whether to subsidize farmers or to build a supercollider, or how much to spend on education, get made through political processes. How can we explain which alternatives are chosen? Economists identify the various groups that benefit or lose from a government program and analyze the incentives facing these groups to attempt to mobilize the political process to promote outcomes favorable to them. They also ask how the structure of government—the "rules of the game" (the rules by which Congress works, whether the President can veto specific items within a bill or only the bill as a whole, and so on)—affects the outcomes. In many quarters, there is a concern that the Supreme Court decision in *Citizens United v. Federal Election Commission*, 558 U.S. 310 (2010), which seemingly gave corporations unbridled opportunities to make campaign contributions, may have tilted the political playing field toward corporations. Then they try to push the question further: What determines how the rules of the game are chosen? In addressing these questions, economics and political science merge. Economists, however, bring a distinct perspective to the analysis: they emphasize the importance of economic incentives in the behavior of participants in the political process, and therefore of economic self-interest in determining outcomes.

ECONOMIC MODELS

A central part of the analysis of the economics of the public sector is understanding the consequences of different policies. Economists, however, sometimes disagree over what those consequences will be. The standard way that science has found to test competing theories is to

MUSGRAVE'S THREE BRANCHES

Richard Musgrave, one of the great public finance economists of the twentieth century, thought of the government as having three economic branches. The first was the *stabilization* branch; its responsibility was to ensure that the economy remained at full employment with stable prices. How this was to be done was the principal subject of courses in macroeconomics. The second branch was the *allocation* branch. Here, the government intervened in how the economy allocated its resources. It did this directly, by buying goods such as defense and education, and indirectly, through taxes and subsidies, which encouraged some activities and discouraged others. The third branch, the *distribution* branch, was concerned with how the goods that were produced by society were distributed among its members. This branch was concerned with issues such as equity and the trade-offs between equity and efficiency. The economics of the public sector focuses on the latter two branches, though the issues arise in other economic courses as well, such as those that deal with regulation.

Today, we recognize that government activities in all three branches are intertwined and cannot be neatly compartmentalized in the way that Musgrave envisaged. Still, his "three branches" provide a convenient way of looking at the myriad of activities in which the government is engaged.

carry out an experiment. With luck, the results of the experiment will bear out the predictions of only one theory while discrediting others. However, economists ordinarily do not have the possibility of doing controlled experiments. Instead, what economists can observe are the uncontrolled experiments that are being done for us in different markets and in different time periods; the historical evidence, unfortunately, often does not permit us to resolve disagreements about how the economy behaves.

To analyze the consequences of various policies, economists make use of what are called **economic models**. Just as a model airplane attempts to replicate the basic features of an airplane, so too a model of the economy attempts to depict the basic features of the economy. The actual economy is obviously extremely complex; to see what is going on, and to make predictions about what the consequences of a particular change in policy will be, one needs to separate out the essential from the inessential features. The features on which one decides to focus in constructing a model depend on the questions one wishes to address. The fact that models make simplifying assumptions—that they leave out many details—is a virtue, not a vice. An analogy may be useful. In going on a long road trip, you may use several maps. One map, depicting the interstate highway system, provides an overview, enabling you to

see how to get from the general area where you are to the general area where you wish to go. You then use detailed maps to see how to get from your point of origin to the expressway, and from the expressway to your final destination. If the interstate highway map showed every street and road in the country, it would be so large that its usefulness would be limited; the extra detail, though important for some purposes, would simply get in the way.

All analysis involves the use of models, of simple hypotheses concerning how individuals and firms will respond to various changes in government policy, and how these responses will interact to determine the total impact on the economy. Everyone—politicians as well as economists—uses models in discussing the effects of alternative policies. The difference is that economists attempt to be *explicit* about their assumptions, and to be sure that their assumptions are consistent with one another and with the available evidence.

NORMATIVE VERSUS POSITIVE ECONOMICS

In their analysis, economists also try carefully to identify the points in their analysis at which values enter. When they describe the economy and construct models that predict either how the economy will change or the effects of different policies, they are engaged in what is called **positive economics**. When they attempt to evaluate alternative policies, weighing the various benefits and costs, they are engaged in what is called **normative economics**. Positive economics is concerned with what "is," with describing how the economy functions; normative economics deals with what "should be," with making judgments about the desirability of various courses of action. Normative economics makes use of positive economics. We cannot make judgments about whether a policy is desirable unless we have a clear picture of its consequences. Good normative economics also tries to be explicit about precisely what values or objectives it is incorporating. It tries to couch its statements in the form "If these are your objectives . . . , then this is the best possible policy."

Consider the positive and normative aspects of a proposal to levy a $1-per-case tax on beer. Positive economics would describe the effect the tax would have on the price of beer—would the price rise by the full $1, or would producers absorb some of the tax? On the basis of that analysis, economists would go on to predict how much beer consumption would be reduced, and who would be affected by the tax. They might find, for instance, that because lower-income individuals spend a larger fraction of their income on beer, these people would be affected proportionately more.

Studies may have indicated that there is a systematic relationship between the quantity of beer consumed and road accidents. Using this information, economists might attempt to estimate how the beer tax would affect the number of accidents. These steps are all part of describing the full consequences of the tax, without making judgments. In the end, however, the question is, *should* the tax be adopted? This is a normative question; in responding to it, economists will weigh the benefits of the tax revenue, the distortions it induces in consumption, the inequities caused by the fact that proportionately more of the tax is borne by lower-income individuals, and the lives saved in road accidents. Furthermore, in evaluating the tax, economists will also want to compare it with other ways of raising similar amounts of revenue.

This example is typical of many such situations that we face in economic policy analysis. Through positive economic analysis, we identify some gainers (the roads are safer) and some losers (consumers who pay higher prices, producers who have lower profits, workers who lose their jobs). Normative economics is concerned with developing systematic procedures by which we can compare the gains of those who are better off with the losses of those who are worse off, to arrive at some overall judgment concerning the desirability of the proposal.

The distinction between normative statements and positive statements arises not only in discussions of particular policy changes but also in discussions of political processes. For instance, economists are concerned with describing the consequences of the majority voting system in the United States, in which the proposal that gets the majority of votes wins. A major group of economists, led by Nobel Prize winner James Buchanan of George Mason University, has focused on *describing* the impact of political processes on social choices (hence, these economists are often referred to as the *social choice school*). What will be the consequences—in terms of patterns or levels of taxation or expenditure, or the speed with which these change in response to changed circumstances—of requiring a two-thirds majority for increments in public expenditures exceeding a certain amount? What will be the consequences of increasing politicians' pay? Of restricting private contributions to political campaigns? Of imposing campaign spending limits, or a variety of other proposals for reforming the financing and conduct of political campaigns? Of public support for political campaigns?

But economists are also concerned with *evaluating* alternative political processes. Are some political processes better, in some senses, than others? Are they more likely to produce consistent choices? Are some political processes more likely than others to yield equitable or efficient outcomes?

DISAGREEMENTS AMONG ECONOMISTS

Unanimity is rare in the central questions of policy debate. Some individuals think affirmative action or bilingual education is desirable; some do not. Some think that the income tax should be more progressive (i.e., that wealthy individuals should pay a higher percentage of their income in taxes, while poor individuals should pay a lower percentage); some believe it should be less progressive. Some agree with the decision to provide a tax credit for college tuition; some believe the money could have been spent in better ways, including ways that are more effective in providing education for the poor. Some believe that capital gains should be taxed like any other form of income; others think capital gains should receive preferential treatment. One of the central concerns of policy analysis is to identify these sources of disagreement.

Disagreements arise in two broad areas. Economists disagree about the consequences of policies (about the positive analysis) and about values (about the normative analysis).

DIFFERENCES IN VIEWS ON HOW THE ECONOMY BEHAVES

As we have seen, the first question economists ask in analyzing any policy is, what are its full consequences? In answering this question, they have to predict how households and firms will react. In 1696, for example, England imposed a tax on windows, under the Act of Making Good the Deficiency of the Clipped Money. At the time windows were a luxury, and the houses of the wealthy had more windows than those of the poor. The window tax could be thought of as a rough substitute for an income tax, which the government did not have the authority to impose. The government should have asked, how much do people value light in their houses? One could imagine a policy debate among the king's advisers about what fraction of the population would value light so little that, rather than pay the tax, they simply would survive with windowless houses. At the time, there were no statistical studies on which the king could rely. (In fact, many people did not value light highly, so the government raised less revenue than anticipated, and more homes were darker than anticipated.)

Today, economists often disagree about the best model for describing the economy, and even after agreeing about the nature of the economy, they may disagree about quantitative magnitudes. For instance, they may agree that increased taxes discourage work, but disagree about the size of the effect.

A standard model that many economists employ assumes that there is perfect information and perfect competition—every firm or individual is so small that the prices it pays for what it buys and receives for what it sells do not depend at all on what it does. Although most economists recognize that information and competition are both imperfect, some believe that the model of perfect information and perfect competition provides a close enough approximation to reality to be useful; others believe that—at least for some purposes, such as the health care market—the deviations are large, and that policy must be based on models that explicitly incorporate imperfect information and competition.

We cannot resolve these disagreements, but what we can do is to show how and when different views lead to different conclusions.

Even when economists agree about the kind of response a particular policy will elicit, they may disagree about the magnitude of the response. This is one of the sources of dispute about the consequences of President Obama's Affordable Care Act of 2010. Most economists believe that providing health insurance to more people will lead individuals who previously did not have insurance to consume more health care—one of the motivations of the program is that many of those without health insurance are getting inadequate care. However, there is disagreement about how much more they will consume. There is even some disagreement over whether total consumption of health care services by the uncovered will decrease: many of these people previously got medical care, but only in an inefficient way (e.g., in high-cost emergency rooms) and only after their illnesses had been allowed to fester, increasing the overall cost of treatment. There is also disagreement over how much of the expense of increased coverage will be offset by initiatives to improve efficiency in the delivery of health services. The answers to these questions dramatically affect the projected cost of this new law.

Although a central concern of modern economics is ascertaining the magnitude of the response of, say, investment, to an investment tax credit, of consumption to a change in the income tax rate, of savings to an increase in the interest rate, and so on, it is an unfortunate fact that various studies, using different bodies of data and different statistical techniques, come up with different conclusions. As economists obtain more data and develop better techniques for analyzing the limited available data, some of these disagreements may be resolved.

DISAGREEMENT OVER VALUES

Whereas the two previous sources of disagreement—concerning the best model for describing the economy and about quantitative magnitudes, such as the size of the response of savings to interest rates—arise within positive economics, the final source of disagreement lies within normative economics. Even if there is agreement about the full consequences of some policy, there may be disagreement about whether the policy is desirable. As has already been noted, there are frequently trade-offs: a policy may increase national output but also increase inequality; it may increase employment but also increase inflation; it may benefit one group

PUBLIC SECTOR ECONOMICS AND THE GLOBAL ECONOMIC CRISIS

As the world strives to recover from the recent global economic crisis, disagreements about how economies function and how we want them to function (positive and normative economics) have intensified. The impact of the crisis has been deep and sustained: the United States experienced its worst recession since the Great Depression of the 1930s; the U.S. unemployment rate remains persistently high, well in excess of levels that would represent full employment; the global economy contracted in 2009 for the first time since the International Monetary Fund (IMF) began collecting data in 1970. The public sector response to this crisis has also been unprecedented: in addition to injections of substantial liquidity by the world's central banks and widespread publicly financed recapitalization of many of the world's largest financial institutions, countries have also enacted large fiscal stimulus packages, totaling 13.5 percent of 2007 GDP in China (4 trillion yuan), 6.8 percent in the United States (the Bush Economic Stimulus Act of 2008 of $168 billion and the Obama American Recovery and Reinvestment Act

of 2009 of $787 billion), and 5.5 percent in Japan (¥27 trillion). Although the stimulus measures did not restore the economy to full employment, U.S. unemployment peaked at a far lower level than it otherwise would have, and without government help, the banking system may well have collapsed, bringing on an even deeper downturn. In Europe, concerns over looming budget deficits resulted in widespread cutbacks, and after a shallow recovery, Europe sank back into recession, with unemployment reaching record levels.

The global crisis has resulted in a major re-examination of the role of government. There is a broad consensus that financial markets took on too much risk on their own and imposed huge costs on the economy, and that markets on their own recover too slowly. Strong government financial regulation had succeeded in preventing financial crisis for decades, after they were enacted in the 1930s, in response to the Great Depression, but these regulations were stripped away beginning in the 1980s. And government intervention, even if imperfectly designed, has played a role in the recovery.

but make another group worse off. Any policy, in other words, may have some desirable consequences and some undesirable consequences. Individuals may weigh these consequences in different ways, some attaching more importance to price stability than to unemployment, others attaching more importance to growth than to inequality.

On questions of values, there is no more unanimity among economists than there is among philosophers. This book presents the major views and assesses some of the criticisms that have been leveled against each.

REVIEW AND PRACTICE

SUMMARY

1. In mixed economies, such as the United States, economic activity is carried on by both private enterprise and the government.

2. Since the time of Adam Smith, economic theory has emphasized the role of private markets in the efficient supply of goods. However, economists and others have come to recognize important limitations in the ability of the private sector to produce efficient outcomes and meet certain basic social needs. The attempt to correct these market failures has led to the growth of government's role in the market economy.

3. The government, however, also has its limitations when intervening to mitigate market failures. These government failures sometimes result in government programs with unintended adverse consequences.

4. The United States has a federal government structure, with certain activities primarily the responsibility of states and localities (e.g., education) and other activities primarily the responsibility of the federal government (e.g., defense).

5. Economics is the study of scarcity—how resources are allocated among competing uses. Public sector economics focuses on choices between the public and private sectors and choices within the public sector. It is concerned with four basic issues: what gets produced, how it gets produced, for whom it gets produced, and the processes by which these decisions are made.

6. In studying the public sector, positive economics looks at the scope of government activity and the consequences of various government policies. Normative economics attempts to evaluate alternative policies that might be pursued.

7. Disagreements about the desirability of policies are based on disagreements about the appropriate assumptions for describing the economy, such as how competitive the economy actually is, disagreements about how strongly the economy will respond to policy initiatives, and disagreements about values.

KEY CONCEPTS

Deregulation

Economic models

Full Employment Act of 1946

Laissez faire

Mercantilists

Mixed economy

Normative economics

Positive economics

Privatization

Production possibilities schedule

QUESTIONS AND PROBLEMS

1. Consider the following discussion of a program of price supports for farmers:

    a. The objective of our farm program is to ensure that all farmers have a reasonable standard of living. The way it does this is to ensure that farmers receive fair prices for their commodities. It is no more right that farmers should produce for substandard prices than that workers should work for substandard wages.

    b. Our farm program has been a failure. The benefits of the price subsidies accrue largely to large farmers (because they produce more). Many farmers still have incomes below the poverty line. The high prices have induced increased production, which has meant high costs for the government. Acreage restrictions have had only limited effect because farmers have kept their best land in production. Direct grants to farmers would be preferable to our price support program.

    Which of the statements in this discussion are normative, and which are positive? (The fact that you disagree with a normative statement or that you think a particular "positive" statement is inaccurate does not change the nature of the statement.)

    Identify the sources of disagreement: Are they due to differences in values and objectives? To differences in perceptions about the nature of the economy? Or to a failure on one (or the other) side of the debate to take into account the full consequences of the government's action?

2. For each of the following programs, identify one or more "unintended" consequences:

    a. Rent control

    b. Minimum wages

    c. Medicare (free hospital care to the aged)

    d. Improved highways making suburbs more accessible to the city

    e. Forced integration of central-city schools

    f. Agricultural price supports

    g. Lowering the speed limit to 55 miles an hour to save gasoline

    h. Providing health insurance to children who currently are underinsured

    i. Banning advertising of cigarettes (Hint: Consider the consequences of increased life spans for the Social Security system.)

    j. National testing standards for schools

3. There has been considerable concern that our Social Security (old age and survivors insurance) program is not adequately financed: with expected birth rates, death rates, and increases in payroll tax collections, the current level of benefits can be sustained only with increases in tax rates. Some believe that the appropriate response is to reduce the current level of benefits, others that the appropriate response is to increase taxes in the future. Still others, worried about the effects of even higher tax rates but believing that lowering the benefits of those currently receiving Social Security would be unfair, argue that benefits should be cut in the future.

    In this discussion, separate the positive statements from the normative statements. To what extent are the disagreements attributable to differences in views of the economy?

2

MEASURING PUBLIC SECTOR SIZE

A central topic of debate in the United States, and in other mixed economies, is the appropriate size of the public sector. Some believe that the public sector is too large. They are skeptical of government's ability to solve social and economic problems because of the kinds of government failures we discussed in Chapter 1—for example, government's limited control over private market responses. Or they may believe in limited government on philosophical grounds, because of a fear that big government undermines economic and political freedom.[1] Others believe that the public sector is too small. In their view, greater government spending could solve the problems of blighted inner cities and inadequate schools.

Whatever view you take, there is no doubt that the U.S. government today is far larger than it was before World War I. In 2010, tax revenues (and other nontax receipts)[2] collected at all levels of government were $3.6 trillion, or 25 percent of total U.S. production, and government

---

[1] A leading proponent of this view, a form of libertarianism, is Robert Nozick. His ideas are summarized in the preface of his *Anarchy, State, and Utopia* (Oxford, UK: Basil Blackwell, 1974). See also Milton Friedman, *Capitalism and Freedom* (Chicago: University of Chicago Press, 1962).
[2] Nontax receipts include, for instance, fees the government receives for various services.

expenditures were 34 percent.[3] By contrast, in 1913, prior to World War I, taxes and government expenditures were less than 10 percent of total production. How do we account for this dramatic change in the size of government? What does the government spend all this money on?

This chapter gives an overview of the scope of the U.S. public sector and how it has broadened over time. It also shows the ways in which government actions affect private markets. Chapter 4 takes up the economic rationale for government intervention in markets. These chapters will not resolve the debate over whether the U.S. public sector is too big or too small, but they provide a basis for formulating a reasonable position on this issue and provide a context for comparison with other countries.

WHAT OR WHO IS THE GOVERNMENT?

Throughout this chapter we have referred to "the government." But what precisely is the government? We all have some idea about what institutions are included: Congress and state and local legislatures, the president and state governors and mayors, the courts, and a host of alphabet-soup agencies, such as the FTC (Federal Trade Commission) and the IRS (Internal Revenue Service). The United States has a *federal* governmental structure—that is, governmental activities take place at several levels: federal, state, and local. The federal government is responsible for national defense, the post office, the printing of money, and the regulation of interstate and international commerce, whereas the states and localities have traditionally been responsible for education, police and fire protection, and the provision of other local services, such as libraries, sewage, and garbage collection. Even though the Constitution asserts that all rights not explicitly delegated to the federal government reside with the states and the people, commonly referred to as *subsidiarity*, the Constitution has proven to be a sufficiently flexible document that the exact boundaries are ambiguous. Although education is primarily a local responsibility, the federal government has become increasingly involved in its support. The constitutional provision giving the federal government the right to control interstate business has provided the basis for federal regulation of almost all businesses, as almost all businesses are involved in one way or another, in interstate commerce.

---

[3] Office of Management and Budget, *Budget of the U.S. Government, Fiscal Year 2012, Historical Tables*, Tables 15.1 and 1.2.

At the local level, there are frequently several separate governmental structures, each having the power to levy taxes and the responsibility for administering certain programs. In addition to townships and counties, there are school districts, sewage districts, and library districts. In 2007, there were 90,000 such governmental entities in the United States, down from 155,000 in 1942.[4]

The boundaries between what are public institutions and what are not are often unclear. When the government sets up a corporation, a public enterprise, is that enterprise part of the "government"? For instance, Amtrak, which was set up by the federal government to run the nation's interstate passenger railway services, receives subsidies from the federal government but otherwise is run like a private enterprise. Matters become even more complicated when the government is a major stockholder in a company, but not the only stockholder. For instance, prior to 1987 the British government owned up to 50 percent of the shares of British Petroleum.

What distinguishes the institutions that we have labeled as "government" from private institutions? There are two important differences. First, in a democracy the individuals who are responsible for running public institutions are elected, or are appointed by someone who is elected (or appointed by someone who is appointed by someone who is elected . . .). The "legitimacy" of the person holding the position is derived directly or indirectly from the electoral process. In contrast, those who are responsible for administering General Motors are chosen by the shareholders of General Motors, and those who are responsible for administering private foundations (such as the Rockefeller and Ford foundations) are chosen by a self-perpetuating board of trustees.

Second, the government is endowed with certain rights of compulsion that private institutions do not have. The government has the right to force you to pay taxes (and if you fail, it can confiscate your property and/or imprison you). The government has the right to seize your property for public use provided it pays you just compensation (this is called the *right of eminent domain*).

Not only do private institutions and individuals lack these rights, but the government actually restricts the rights of individuals to give to others similar powers of compulsion. For instance, the government does not allow you to sell yourself into slavery.

In contrast, all private exchanges are voluntary. I may need your property to construct an office building, but I cannot force you to sell it. I may think that some deal is advantageous to both of us, but I cannot force you to engage in the deal.

---

[4] U.S. Census Bureau, *Statistical Abstract of the United States, 2011,* Table 426.

Government is thus fundamentally different from other institutions in our society. It has strengths—its ability to use compulsion means that it may be able to do some things that private institutions cannot do. But it also has weaknesses, as we discuss in greater detail in later chapters. Understanding these strengths and weaknesses is an essential part of assessing what should be the role of the government in our mixed economy, and of determining how government can most effectively fulfill that role.

TYPES OF GOVERNMENT ACTIVITY

A primary role of government is to provide the legal framework within which all economic transactions occur. Beyond that, the activities of government fall into four categories: (1) the production of goods and services; (2) the regulation and subsidization or taxation of private production; (3) the purchase of goods and services, from missiles to the services of street cleaners; and (4) the redistribution of income—that is, payments like unemployment benefits to particular groups of individuals that enable them to spend more than they could otherwise. Payments that transfer money from one individual to another—but not in return for the provision of goods or services—are called **transfer payments**.

These four categories—production, regulation, purchase, and redistribution—are simply a convenient way of grouping the vast array of government activities. However, they do not correspond to the way the federal government organizes its budget or divides responsibilities among its various departments—Commerce, Health and Human Services, Interior, and so on. Moreover, government activities are undertaken at the state and local levels as well as at the federal level, with the relative importance of state, local, and federal expenditures of various types having changed over time.

A final complication is that the nature of some government expenditures is ambiguous. For example, government subsidies to small farmers could be considered a production subsidy or a redistributive (transfer) payment. Pension payments to military retirees are often counted as transfer payments, but they are more appropriately treated as part of the cost of national defense, just as the pension costs of a private firm are counted among its labor costs.

Thus, the task of constructing a quantitative description of the government's activities is a formidable one.

PROVIDING A LEGAL SYSTEM

An important activity of the government, but one that accounts for very little expenditure, is the establishment of the legal framework within which firms and individuals can engage in economic interactions. Economists and philosophers often try to imagine what life would be like in the complete absence of government. Without laws defining property rights, for instance, only the exercise of force would stop one individual from stealing from another. Without the ability to protect property, individuals would have little incentive to accumulate assets. Needless to say, economic activities would be severely restricted.

The U.S. legal system does much more than just protect property rights. It enforces contracts between individuals. It also imposes restrictions on the kinds of contracts that are legally enforceable. Our bankruptcy laws limit the liability of investors. Product liability laws have an important effect on the quality of goods produced. Antitrust laws attempt to encourage competition among firms: they restrict mergers, acquisitions, and unfair business practices.

The effects of our criminal justice system are pervasive, but expenditures for running it are relatively small: about 5 percent of total government expenditures.[5]

GOVERNMENT PRODUCTION

The U.S. government undertakes certain types of production directly. Much of this is similar to corresponding activities carried out by private firms. For instance, both private and government enterprises produce and sell electricity (the most famous of the latter is, perhaps, the Tennessee Valley Authority). In addition, under the Constitution, the federal government takes responsibility for running the postal service and for printing money.[6]

At the local level, many communities provide water and collect garbage, services that in other communities are provided by private firms. Most elementary and secondary school students go to public schools— schools run by the government—although others go to private schools,

---

[5] Based on 2007 data, the most recent available. See Office of Management and Budget, *Budget of the U.S. Government, Fiscal Year 2012, Historical Tables*, Table 3.1; and Bureau of Justice Statistics, "Direct Expenditure by Level of Government, 1982–2007," in *Justice Expenditure and Employment Extracts, 2011.*

[6] Although the U.S. Postal Service has a monopoly on the delivery of first-class mail, private carriers, such as United Parcel Service, Federal Express, and others, play a major role in the delivery of parcels and express mail.

some of which are run by nonprofit organizations like churches and a few run on a for-profit basis.

Comparing the public and private sectors in various countries, we see that some industries frequently fall within the public sector, whereas other industries seldom do. Agriculture and retail trade are seldom in the public sector. On the other hand, in most countries, at least part of the radio and TV broadcasting industry lies in the public sector. In many countries, the banking system is at least partially owned and operated by the government; in the United States it is closely regulated but privately owned.[7]

The line between public and private production shifts over time. During the past two decades in Europe, many countries have converted public enterprises into private enterprises, a process called *privatization*. (The process of converting private enterprises to government enterprises is called **nationalization**.) For instance, the British government has privatized enterprises in industries ranging from telecommunications to energy, automobiles, aerospace, and steel. In France, a wave of privatization, which began in 1986, included the privatization of enterprises that had been nationalized earlier in the decade when the socialist party was in power.

The distinction between public and private production has also become more blurred. Public enterprises are sometimes corporatized or commercialized, so they are managed like private enterprises although they are still government owned, as are many of Singapore's government-linked companies (GLCs). This can also be a transitional step in the privatization process. There are also many hybrid models characterized by cooperation between the public and private sectors, such as public–private partnerships (PPP) and private sector participation (PSP) in the provision of public goods and services.[8]

For technical reasons, the best way to measure the size of government production is to look at employment, as in Figure 2.1. In 2009, public employees, including public education and the armed forces, represented 16.9 percent of total employment. This was almost double the percentage in 1929 when it was 8.9 percent of the labor force. The figure shows

---

[7] The Federal Reserve Banks, which are responsible for the management of the banking system, are publicly owned. Most of their profits are turned over to the U.S. Treasury. In 2009, their comprehensive income prior to distribution was $53.4 billion. (*The Federal Reserve Banks Combined Financial Statements as of and for the Years Ended December 31, 2009 and 2008 and Independent Auditors' Report*.)

[8] For more on privatization, see William L. Megginson et al., "The Financial and Operating Performance of Newly Privatized Firms: An International Empirical Analysis," *Journal of Finance* 49, no. 2 (1994), and Pierre Guislain, *The Privatization Challenge: A Strategic, Legal, and Institutional Analysis of International Experience* (Washington, DC: World Bank, 1997). For more on alternatives to privatization, see José A. Gómez-Ibáñez, "Alternatives to Infrastructure Privatization Revisited: Public Enterprise Reform from the 1960s to the 1980s," Policy Research Working Paper 4391 (Washington, DC: World Bank, 2007), and Darrin Grimsey and Mervyn K. Lewis, *Public Private Partnerships: The Worldwide Revolution in Infrastructure Provision and Project Finance* (Cheltenham, UK: Edward Elgar, 2004).

FIGURE 2.1

GOVERNMENT EMPLOYMENT AS A PERCENTAGE OF TOTAL EMPLOYMENT, 1929–2009

Government employment as a percentage of all employment provides a view of the government's role as producer. Employment is defined here as the number of full-time equivalent employees.
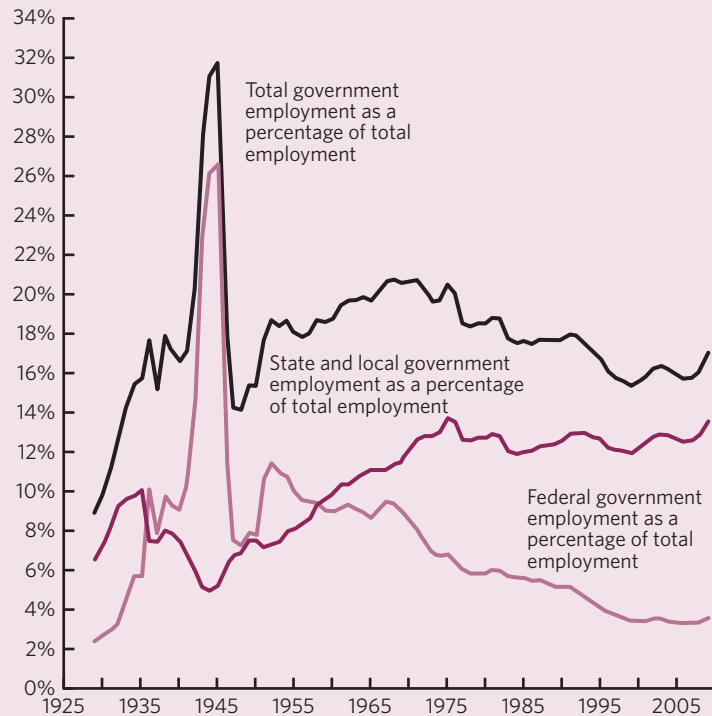
a marked increase in the ratio of public employment from 1929 through 1936 (both in the Hoover and Roosevelt administrations), a burst of public employment during World War II, and a return to prewar levels by 1947. Although there was a slight decrease in the pace of growth during the Eisenhower years, employment in the public sector did not begin to decline until the Nixon and Ford administrations. This decline continued until the current global economic crisis. In fact, by 2007 the federal government's share of total employment had fallen to 3.2 percent, the same as it was in 1932, before the New Deal.

It is also important to note the variations in the relative roles played by the federal, state, and local governments, as suggested by the bottom line in Figure 2.1. Comparing it to the top line, we see that total government employment and federal government employment do not always move together. Federal employment as a percentage of total employment declined in the early 1970s, but this decline was offset by the rise in employment at the state and local level. It is important to bear this in mind: reductions in federal expenditures or employment do not of

themselves necessarily imply a reduction in government expenditures or employment. More of a burden may simply be placed on states and localities.

GOVERNMENT'S INFLUENCE ON PRIVATE PRODUCTION

In industries in which the government is neither a producer nor a consumer, it may nevertheless have a pervasive effect on the decisions of private producers. This influence is exercised through subsidies and taxes—both direct and indirect—and through regulations. There are many motives for such government influence. There may be dissatisfaction with particular actions of firms, such as pollution. There may be concern about the monopoly power of some firms. Special interest groups may convince Congress that they are particularly deserving of help. Private markets may fail to provide certain goods and services that are felt to be important.

SUBSIDIES AND TAXES   Government subsidizes private production in three broad ways: direct payments to producers, indirect payments through the tax system, and other hidden expenditures. The most extensive of the U.S. government's subsidy programs is for agriculture. Direct payments to farmers rose precipitously during the 1980s, from $1.3 billion in 1980 to a peak of $16.7 billion in 1987, when they began a steady decline until they reached $7.3 billion in 1995. In 1987, direct payments amounted to 37 percent of income from wheat, 40 percent of income from rice, and 20 percent of income from all crops. At least one of every five dollars in farm income was a gift from the government. Despite passage of the "Freedom to Farm" bill of 1996, which was designed to further reduce reliance of U.S. farmers on government payments, subsequent legislation and volatile markets have resulted in marked fluctuation of agricultural subsidies over the past decade, ranging from a low of $11.9 billion in 2007 to a high of $24.4 billion just two years earlier.[9]

The tax system also sometimes serves to subsidize production. If the government gives a grant to a producer to assist in buying a machine, it appears as an expenditure. But suppose the government allows the producer to take a tax credit on the expenditures on machines—that is, if the

---

[9] United States Department of Agriculture, Economic Research Service, *Economic Indicators of the Farm Sector: National Financial Summary, 1992*, Tables 14 and 22, January 1994; *Survey of Current Business*, Table B-10, June 1997, p. D-24; *Direct Government Payments by Program, United States, 1933–95*; and *Direct Government Payments by Program, United States, 1996–2009*.

producer buys a $100 machine with a 7 percent tax credit, it will get a $7 tax credit, which reduces by $7 the taxes it otherwise would have paid. Although it is not accounted as such, for all intents and purposes the tax credit is equivalent to government expenditure, and is thus referred to as *tax expenditure*. The value of federal tax expenditures has become very large in recent years, amounting to 30 percent of direct expenditures for fiscal year 2010.[10]

Finally, many government subsidies show up neither in the statistics on government expenditures nor in those on tax expenditures. For instance, when the government restricts the importation of some foreign good or imposes a tariff on its importation, this raises the prices of that good in the United States. American producers of competing goods are helped. In effect, there is a subsidy to American producers, paid not by the government but directly by consumers.

GOVERNMENT CREDIT   A special type of subsidy is government provision of credit below market interest rates, in the form of low-interest loans and loan guarantees. Government subsidies tend to lead to the expansion of the subsidized industry, by lowering its cost of doing business. This is as true for subsidies to credit as it is for other forms of subsidies. Although such subsidies were once hidden, the Credit Reform Act of 1990 required the government to treat as expenditures any difference between the interest rates it paid and the interest rates it charged (taking into account the probability that the borrower might not repay).

In addition to loan subsidies, other government programs affect the allocation of credit, and thus of productive resources. In the United States, the subsidies are often used to buy particular goods and services. For instance, government-sponsored enterprises (GSEs) encourage lending to enable people to buy homes and go to school. In 1993, the Clinton administration began lending funds directly to college students.

REGULATING BUSINESS   Government regulates business activity in an attempt to protect workers, consumers, and the environment; to prevent anticompetitive practices; and to prevent discrimination.

The Occupational Safety and Health Administration attempts to ensure that workers' places of employment meet certain minimal standards. The National Labor Relations Board attempts to ensure that management and unions deal fairly with each other. The Federal Trade Commission attempts, among other things, to protect consumers from misleading advertising. The Environmental Protection Agency attempts

---

[10] *Budget of the United States Government, Fiscal Year 2012, Analytical Perspectives*, Table 17-1, and *Historical Tables*, Table 1.1.

to protect certain vital parts of our environment by regulating, for instance, emissions from automobiles and toxic waste disposal.

In addition to these broad categories, some regulations apply to specific industries. The banking industry is regulated both by the Federal Reserve Board and the Comptroller of the Currency. Trucking is regulated by the Federal Highway Administration. The airlines are regulated by the Federal Aviation Administration. The telecommunications industry is regulated by the Federal Communications Commission. The securities industry is regulated by the Securities and Exchange Commission. Beginning in the late 1970s, there was a concerted effort to reduce the extent of federal regulation. As noted earlier, the process of reducing or eliminating regulations is referred to as deregulation. There has been significant deregulation in the airline, natural gas, electricity, trucking, and telecommunications industries.

Although the overall trend has entailed reduced regulation, there have been some instances of tightened regulation: the massive failure of the savings and loan associations in the 1980s was attributed in part to lax banking regulation, and legislation enacted in 1989 provided for heightened scrutiny. Likewise, repeal in 1999 of the Glass–Steagall Act of 1933, which had separated commercial and investment banking, contributed to the recent global economic crisis, so the Dodd–Frank Wall Street Reform and Consumer Protection Act was passed in 2010 to strengthen financial sector regulation. In other cases, the focus has been on changing regulation to reflect changing circumstances. For example, several cases of deaths from food poisoning reinforced the importance of food safety regulations, but there was increasing recognition that the visual inspection system (did the meat smell and look rotten?) that had been employed since the beginning of the twentieth century needed to be replaced with a more scientific process.

Today, there is a move toward "smart regulation" that tries to balance the need for government oversight with its potential burden. It began with President Clinton's issuance of an executive order in 1993 calling for a cost–benefit review of government regulations. The initiative was called "Reinventing Government" or "National Performance Review," and focused on making government agencies more client oriented and employing more market-like regulatory mechanisms. This approach was reaffirmed and extended by a subsequent executive order issued by President Obama in 2011. Both directives stress the principle that a regulation's benefits must justify its costs and "impose the least burden on society, consistent with obtaining regulatory objectives."[11]

---

[11] See President Clinton's Excutive Order 12866 of September 30, 1993, and President Obama's Executive Order 13563 of January 18, 2011 (the source of the quotation in the text), for specific instructions on improving regulation and regulatory review, including acknowledgement of the difficulty in quantifying many costs and benefits, as well as the importance of a participatory process when reviewing government regulations.

Federal outlays for the regulatory agencies are minuscule relative to the rest of the budget. However, these expenditures do not give an accurate view of the impact of the federal regulatory agencies. The extent to which these agencies influence virtually every aspect of business practices goes well beyond the simple measure of government expenditures. Many regulations have effects that are similar to those of taxes and subsidies. For example, regulations on utility prices may reduce prices for certain users below the free-market level, while raising the price to other users.

GOVERNMENT PURCHASES OF GOODS AND SERVICES

Every year the government buys billions of dollars' worth of goods and services to support our national defense, maintain a network of highways, and provide education, police protection, fire protection, and parks. These purchases of goods and services amount to nearly one-fifth of the total production in the United States. In 2010, total government purchases were $3 trillion. Of these purchases, 17 percent was for investments in infrastructure such as roads and bridges that increase the economy's future productivity.[12]

What we characterize as government purchases are amounts spent for goods and services made available to the public, such as national defense, public schools, and highways. Government payments to the aged through the Medicare program to finance their hospital expenses or to the poor through the food stamp program are categorized as transfer payments, not as direct government purchases. They are discussed in the next section, on government redistribution of income.

GOVERNMENT REDISTRIBUTION OF INCOME

The government takes an active role in redistributing income; that is, in taking money away from some individuals and giving it to others. There are two major categories of explicit redistribution programs: public assistance programs, which provide benefits to those poor enough to qualify; and social insurance, which provides benefits to the retired, disabled, unemployed, and sick.

As we saw earlier, outlays for explicit redistribution programs are called *transfer payments*. These expenditures are qualitatively different from government spending on, say, roads or bombers. Transfer payments are simply

---

[12] The investment numbers do not include investments in people—human capital, for either education or health—but they do include investments in military aircraft and other hardware that enhance the country's future defense capabilities, in transportation infrastructure such as highways and airports, and in natural resource ventures such as pollution control facilities.

changes in who has the right to consume goods. In contrast, a government out-lay for a road or a bomber reduces the amount of other goods such as private consumption goods that society can enjoy. Transfer payments affect the way in which society's total income is divided among its members, but transfers do not affect the total amount of private goods that can be enjoyed (neglecting here losses of output due to distorted incentives associated with transfers).

PUBLIC ASSISTANCE PROGRAMS Public assistance programs take two forms. Some provide cash, whereas others provide payment only for specific services or commodities. The latter are referred to as **in-kind benefits**. Of the cash programs, the largest are Temporary Assistance to Needy Families (TANF), the program that replaced the long-standing Aid to Families with Dependent Children (AFDC) in 1997, and Supple-mental Security Income (SSI), which provides cash to the poor who are aged, blind, or disabled. The largest in-kind public assistance program is Medicaid, which covers the medical costs of the poor, and accounts for about one half of total public assistance.

Table 2.1 lists the main public assistance programs with their date of enactment and their benefits. (In-kind benefits are valued at government cost in the table; we will see later that this may be different from their value to the recipients.) The table shows that most benefits are in-kind, not cash,

TABLE 2.1 SELECTED GOVERNMENT PUBLIC ASSISTANCE PROGRAMS (IN MILLIONS OF 2008 DOLLARS).

PROGRAM	DATE ENACTED	1990 OUTLAY	2000 OUTLAY	2008 OUTLAY
CASH BENEFITS				
AFDC/TANF from 1998	1935	32,109	22,950	18,874
SSI	1972	27,897	39,423	44,062
Earned income tax credit*	1975	9,877	40,196	50,669
Total		69,884	102,569	113,605
IN-KIND BENEFITS				
Medicaid	1965	130,828	255,169	352,170
Food stamps/Supplemental Nutrition Assistance Program (SNAP) from 1998	1964	24,217	18,128	36,442
National School Lunch Program	1946	5,379	6,837	8,265
Total		160,423	280,133	396,877
TOTAL		230,307	382,702	510,482
In-kind benefits share of total		69.7%	73.2%	77.7%

*Includes only the refunded portion.

SOURCES: Census Bureau, *Statistical Abstract of the United States, 2011*, Tables 538 and 568, and *Statistical Abstract of the United States, 1992*, Table 565; Internal Revenue Service, *Statistics of Income Bulletin, Historical Table*, Table 4.

<div style="border: 2px solid #9b3c5c; padding: 1em;">

GOVERNMENT ACTIVITIES

- Providing a legal system—required if a market economy is to function
- Producing goods—defense, education, mail
- Affecting what the private sector produces, through subsidies, taxes, credit, and regulation
- Purchasing goods and services from the private sector, which are then supplied by the government to firms and households
- Redistributing income

</div>

and that this share has risen from roughly two-thirds to more than three-quarters since 1990.

SOCIAL INSURANCE PROGRAMS **Social insurance** differs from public assistance in that an individual's entitlements are partly dependent on his or her contributions, which can be viewed as insurance premiums. To the extent that what individuals receive is commensurate with their contributions, social insurance can be viewed as a government "production activity," not a redistribution activity. Because what some receive is far in excess of what they contribute (on an actuarial basis), however, a large element of redistribution is involved in government social insurance programs.

The largest of these programs is the Old-Age, Survivors, and Disability Insurance Program (OASDI, the proper name for Social Security). It provides income not only for the retired, but also for their survivors (in particular, widows and widowers) and the disabled. The other major social insurance programs are unemployment insurance, which provides temporary benefits after an individual loses a job; workers' compensation, which provides compensation for workers injured at work; and Medicare. The Medicare program, providing medical services to the aged, has (like Medicaid) grown rapidly since it was first introduced in 1965, and now is the second largest social insurance program. Figure 2.2 gives the relative size of the various social insurance and public assistance programs.

The Social Security and Medicare programs are sometimes referred to as **middle-class entitlement programs**, because the main beneficiaries are the middle class, and benefits are provided not on the basis of need but because the beneficiaries satisfy certain other eligibility standards (e.g., age). As soon as they satisfy these criteria, they become entitled to receive the benefits.

HIDDEN REDISTRIBUTION PROGRAMS The government affects the distribution of income not only through direct transfers but also through the indirect effects of the tax system and other government programs. One could imagine the government taxing everyone at the same rate but then giving grants to those whose income fell below a certain level. This would have the same effect as taxing the lower-income individuals at a lower rate. Thus, there is a certain arbitrariness in distinguishing between transfer payments through spending programs and the implicit transfers through the tax system.[13]

---

[13] Some of the tax expenditures can be viewed explicitly as forms of social insurance. The fact that unemployment insurance and Social Security are only partially taxed and disability benefits not taxed at all, means that a dollar of direct expenditures for those purposes goes further than it would if subjected to taxation.

A

Workers'
compensation and
temporary disability
(1.7%)

Other
(1.3%)

Unemployment
insurance
(9.7%)

OASDI
(49.8%)

Medicare
(37.5%)

Social insurance total:
$1,333.5 billion

B

Veterans' benefits
(6.8%)

Other
(15.0%)

EITC
(9.1%)

SNAP
(7.5%)

Medicaid and
other
medical care
(52.4%)

Public assistance total:
$732.4 billion

Family assistance
(2.7%)

SSI
(6.5%)
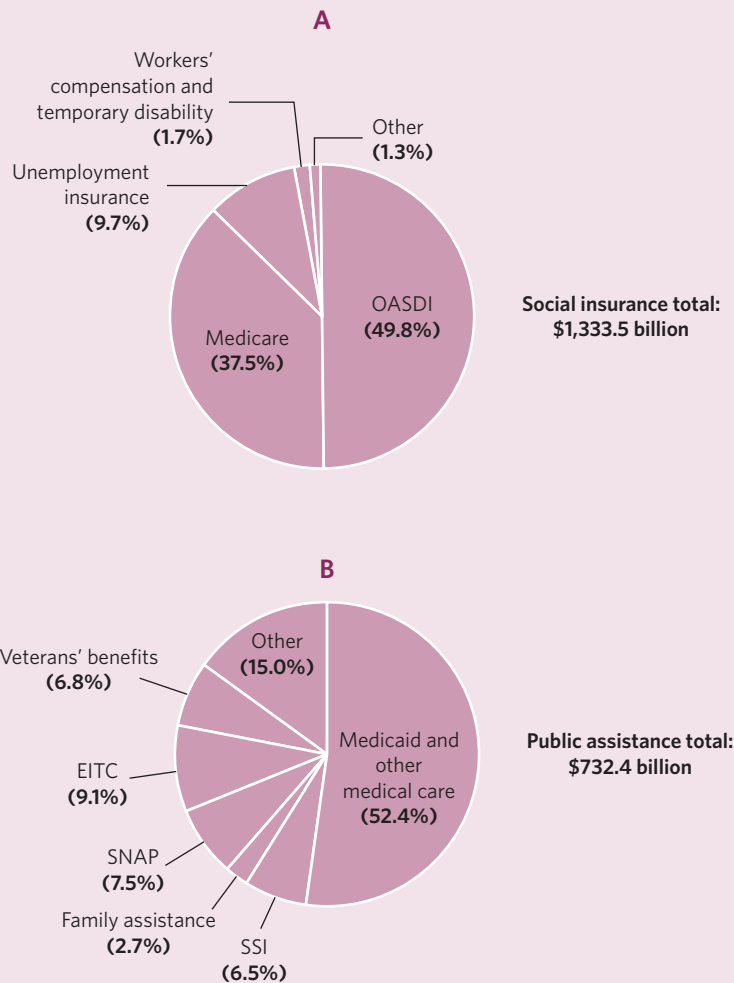
FIGURE 2.2

**RELATIVE IMPORTANCE
OF SOCIAL INSURANCE
AND PUBLIC ASSISTANCE
PROGRAMS, 2009**
Social Security (OASDI) and
Medicare are by far the largest
social insurance expenditures,
as well as the largest overall
transfer programs. Medicaid
is the largest public assistance
expenditure and the third larg-
est overall transfer program.

The major example of a transfer program run through the tax system
is the earned income tax credit (EITC), which actually provides supple-
mental income to low-income earners. Under the Clinton administration,
expenditures for the EITC were expanded in an effort to enhance the
incentives for low-skilled workers to stay off welfare.

The government also redistributes income in the guise of subsidy
programs and quotas. Our agricultural programs in effect redistrib-
ute income to farmers. The oil import quotas of the 1950s redistributed
income to owners of oil reserves. The alleged reason for the quotas was
to ensure the energy independence of the United States; nonetheless, the

redistributive effects were among the primary consequences, and they may indeed provide the true motivation for the legislation.

Spending for goods and services also has its redistributive consequences: subsidies to urban bus transport may help the poor, whereas subsidies to suburban rail lines may help the middle class.

OVERVIEW OF GOVERNMENT EXPENDITURES

We can now put together the discussion of purchases and transfers to get an overview of government expenditures. Figure 2.3 shows the distribution of government expenditures in 2009. In panel A, which combines outlays at all levels of government, we can see that purchases of goods and services

FIGURE 2.3

GOVERNMENT EXPENDITURES BY TYPE, 2009
(A) Today, almost half of all government expenditures at the federal, state, and local levels are transfers, (B) excluding an additional one out of every seven dollars of federal expenditures for grants-in-aid to state and local government.

A

Interest (7.2%)
Subsidies (1.2%)
Purchases (48.3%)
Transfers (43.3%)

Total government expenditures, 2009: $4,998.7 billion

B

Interest (7.3%)
Subsidies (1.7%)
Grants-in-aid to state and local government (14.0%)
Purchases (28.6%)
Transfers (48.4%)

Federal government expenditures, 2009: $3,457.4 billion
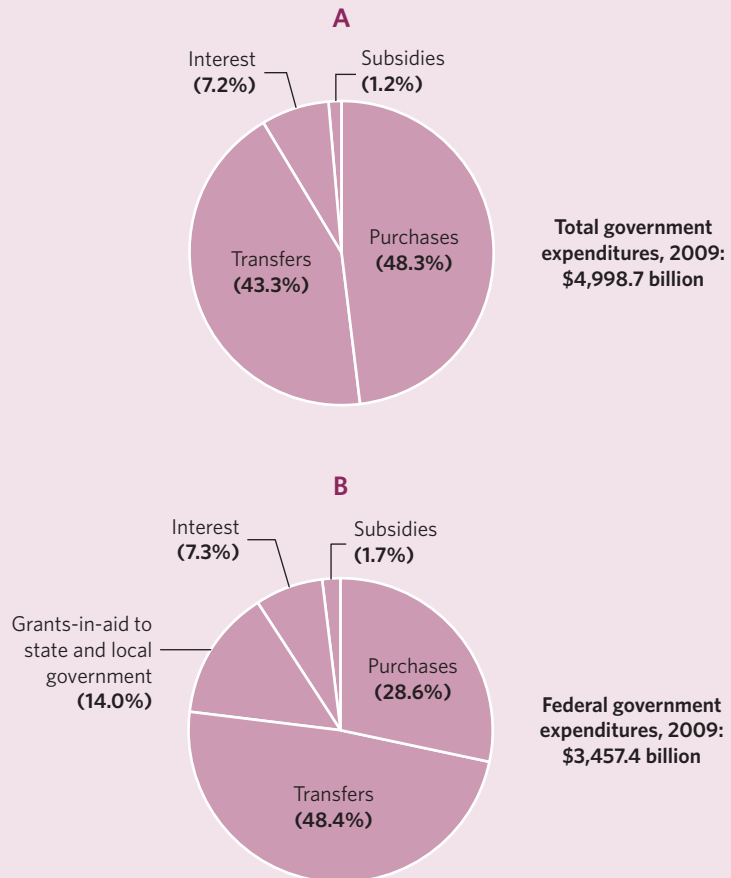
SOURCE: *Economic Report of the President, 2011*, Tables B-82, B-84, and B-85.

primarily for defense and education constitute almost half of expenditures, and transfer payments comprise the bulk of the rest. However, in panel B, we can see that at the federal level, almost two-thirds of all expenditures are for transfer payments to individual, states, and local governments.

Figure 2.4 shows the purpose of the expenditures by broad categories, both for the federal government's expenditures and for all government expenditures. Note that at the federal level, Social Security (OASDI) and defense purchases play major roles. For total government expenditures, education appears as a major category because it is the largest type of expenditure at the state and local level.[14]

---

[14] There is some inherent imprecision in any classification. For instance, veterans' benefits, which are typically not included in defense expenditures, can be thought of as expenditures for previously delivered defense services, and some of the expenditures on space are motivated by defense concerns.

A

Interest (7.2%)
Subsidies (1.2%)
Education purchases (15.1%)
Other transfers (30.0%)
Defense purchases (13.3%)
Other purchases (19.9%)
OASDI (13.3%)

Total government expenditures, 2009: $4,998.7 billion

B

Interest (7.4%)
Subsidies (1.7%)
Grants (14.0%)
Defense purchases (19.2%)
Other transfers (29.2%)
OASDI (19.2%)
Other purchases (9.3%)

Federal government expenditures, 2009: $3,457.4 billion
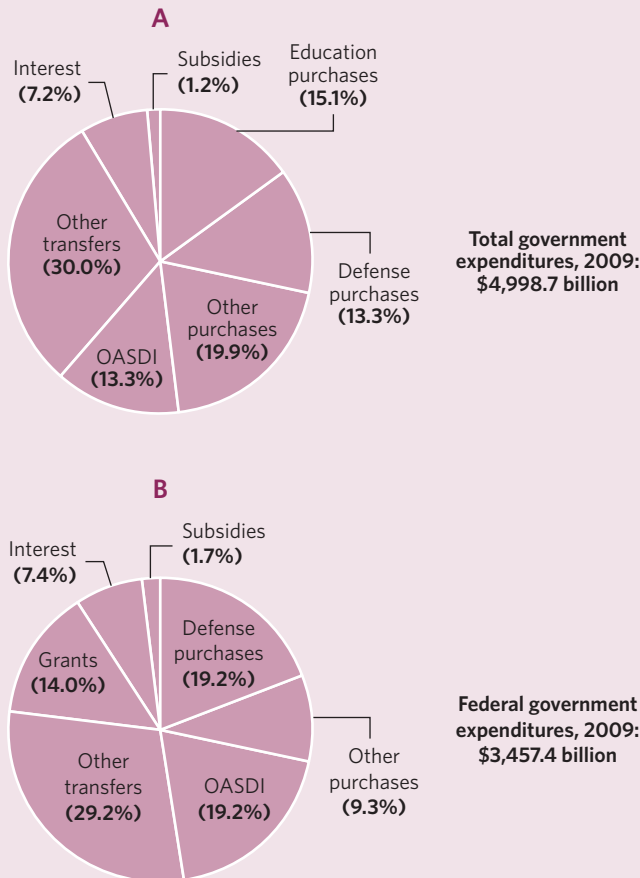
FIGURE 2.4

GOVERNMENT EXPENDITURES BY PURPOSE, 2009

At the federal level, the most important expenditures, other than transfers, are for defense; at the state and local level, the most important expenditures are for education.

SOURCES: U.S. Department of Commerce, Bureau of Economic Analysis, *National Income and Product Accounts*, Tables 3.1, 3.12, 3.9.5, and 3.15.5; *Economic Report of the President, 2011*, Tables B-82, B-84, and B-85.

There have been marked changes in the relative importance of expenditures at different levels of government over the past century. For instance, the federal share of all nondefense government spending grew from slightly less than one-fifth in 1902 to almost two-thirds in 2009.

GAUGING THE SIZE OF THE PUBLIC SECTOR

Because the government's impact on the private economy depends on its regulatory and tax policies as well as on its outlays, no single number can provide an accurate indicator of the government's effect on the economy. Nonetheless, one indicator that economists have found particularly convenient to use is the size of public expenditures relative to the size of the total economy. A standard measure of the size of the total economy is gross domestic product (GDP), which is a measure of the value of all the goods and services produced in the economy during a given year.

GROWTH IN EXPENDITURES AND THEIR CHANGING COMPOSITION

During the past eighty years, public expenditures as a share of GDP have quadrupled. In 1929, they were 10 percent of GDP, whereas in 2010, they represented 40 percent of GDP, as we see in Figure 2.5.[15]

DEFENSE EXPENDITURES   Figure 2.5 also shows that between 1967 and 1979 defense expenditures as a percentage of GDP declined from 10.0 percent to 5.7 percent. They then increased during the Reagan years until 1987, peaking at 7.4 percent, but after the end of the Cold War they once again declined to 3.7 percent of GDP in 2000. They were back up to 5.6 percent in 2010, primarily due to the conflicts in Iraq and Afghanistan.

To avoid the misleading impressions that can be caused by failing to take account of inflation, economists like to express expenditures in dollars of constant value. Thus, if last year the government spent $1 billion on some program, and this year it spends $1.1 billion but overall prices have increased by 10 percent, we say that the current expenditures (measured in last year's prices) are $1 billion—in constant dollars, expenditures have not increased at all.

---

[15] Recall from the earlier discussion the arbitrariness of this measure. For instance, if the government switches from providing aid to education through direct grants to providing it through tax expenditures, these statistics would show a fall in the share of public expenditures.
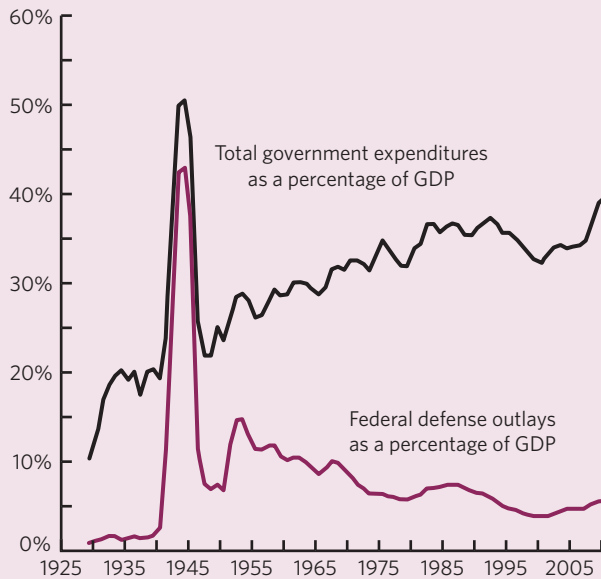
ESTIMATING THE FULL BUDGETARY AND ECONOMIC COSTS OF WAR

Annual defense allocations rarely include the full budgetary costs of war, and certainly exclude the economic costs beyond direct financial outlays. For example, in 2008, five years after the United States invaded Iraq, the Congressional Budget Office estimated the cost of the Iraq war at between $1 trillion and $2 trillion, depending on troop levels and duration of the conflict. This estimate consists of outlays beyond normal defense and veterans affairs expenditures. It includes war zone operations and related logistical support, repair and replacement of some equipment, and a portion of medical expenses for wounded combatants. However, if other costs are included, such as the long-term costs of providing medical care and disability compensation for war veterans, the economic costs of lives lost and disrupted, and the impact of rising oil prices, the 2008 estimate of other analysts rises to $4 trillion. Relatively conservative full-cost accounting estimates in 2011 are between $4 trillion and $6 trillion.

For a more detailed discussion on calculating the full costs of war, see Joseph E. Stiglitz and Linda J. Bilmes, *The Three Trillion Dollar War: The True Cost of the Iraq Conflict* (New York: W. W. Norton & Company, 2008), and http://threetrilliondollarwar.org.

Thus, in constant 2005 dollars, defense expenditures shrank from $518 billion in 1968 to $311 billion in 1977. They then increased over the next decade to $482 billion in 1989 before declining again to $346 billion in 1998. Since then they have risen to a post-Vietnam War high of $612 billion in 2010.

TRANSFER PAYMENTS AND INTEREST   Growth in expenditures for Social Security and Medicare account for much of the rise in public expenditures since 1950—social insurance's share of expenditures increased almost fivefold during this period (see Figure 2.6). Interest payments also increased significantly, owing to the huge deficits that began under President Reagan, as the government spent more than it received. It was not until President Clinton succeeded in passing a deficit reduction act in 1993 that the deficit was brought under control; interest payments actually came down, as interest rates fell by more than the deficit increased.

FIGURE 2.6

GOVERNMENT TRANSFER AND INTEREST PAYMENTS AS A PERCENTAGE OF TOTAL EXPENDITURES, 1950–2009

Social insurance payments (largely OASDI and Medicare) have grown from 5.9 percent of total public expenditures in 1952 to 26.7 percent of expenditures in 2009. Public assistance share of total expenditures has risen from 6.1 percent in 1963 to 14.6 percent in 2009. Net interest payments grew from a low of 4.7 percent of expenditures in 1975 to a high of 9.7 percent in 1995. Although the national debt has grown tremendously over the past decade, interest payments fell to 3.9 percent of expenditures in 2009 as a result of near-zero interest rates on government borrowing.

SOURCES: U.S. Department of Commerce, Bureau of Economic Analysis, *National Income and Product Accounts*, Tables 3.1 and 3.12; and Office of Management and Budget, *Historical Tables*, Table 15.4.
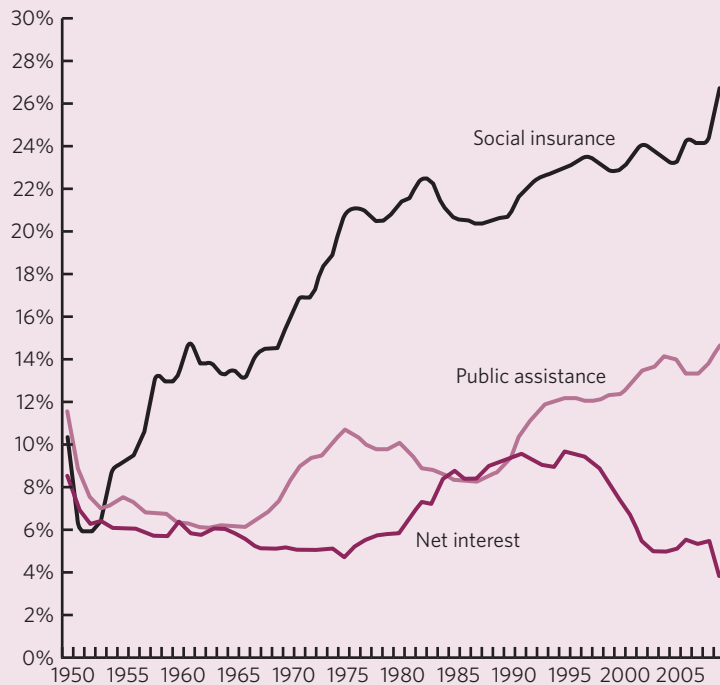
Although public assistance is often blamed for the growth in public expenditures, its share of total government expenditures increased only from 6.1 percent of expenditures in 1963, before President Johnson's War on Poverty, to 14.6 percent in 2009.

A major source of increased expenditure during the past two decades has been health care—Medicaid and Medicare—and there is real concern that these expenditures will continue to soar in coming decades as well.

Figures 2.5 and 2.6 tell one other interesting story. Total nondefense expenditures in general, as well as social insurance expenditures in particular, increased rapidly from the mid-1960s through the mid-1970s, under both Democratic and Republican administrations. The most significant break in the *rate of increase* of expenditures occurred under President Carter (1976–1980). Furthermore, although Presidents Reagan and both Bushes preached that they would cut back the size of government, they failed to do so.

One reason for this is the tremendous inertia in the fiscal system. The full economic consequences of the Medicare program, enacted in 1965, were not felt until many years later. The scope for discretion—for changing directions—within any administration is accordingly limited.

GAUGING THE SIZE OF GOVERNMENT

- The size of the U.S. government today is much larger than it was a hundred years ago.
- During the past eighty years, public expenditures as a share of GDP have grown rapidly. In 1930, they were 9 percent of GDP. In 2010, they represented 36 percent of GDP.
- Growth in expenditures for Social Security, Medicare, Medicaid, and interest account for much of the increase in public expenditures since 1950.
- The size of government relative to the economy is much smaller in the United States than in most European countries.

COMPARISON OF EXPENDITURES ACROSS COUNTRIES

The share of government, measured by total government expenditures as a percentage of GDP, is smaller in the United States than in most other high-income countries (see Figure 2.7), and its relative growth has also been slower than in most other advanced economies. The share of government in most middle- and low-income countries is considerably smaller than it is in high-income countries.

Because defense expenditures play a larger role in the United States, the relative size of nondefense expenditures is particularly low, viewed from this international perspective (see Figure 2.8).[16]

---

[16] Comparisons across countries always need to be treated with caution. Particular problems are raised by the treatment of public enterprises. The fact that tax expenditures are relatively more important here than abroad may result in an understatement of the "effective" relative size of the public sector in the United States. On the other hand, regulations are perhaps less important in the United States than in most other developed countries.

GOVERNMENT REVENUES

Now that we have examined what the government spends its money on, we will briefly survey the methods by which government raises revenue to pay for these expenditures. The government levies a variety of taxes. When the revenues that it receives from taxes are less than its planned expenditures, it must either cut back expenditures or borrow the difference.[17]

TAXES AND THE CONSTITUTION

The issue of taxation was very much in the thoughts of the founders of the republic. Indeed, the American Revolution began as a tax revolt with the Boston Tea Party, which was a protest against the tax on tea, and with the slogan "Taxation without representation is tyranny." The first article of the Constitution provides, "The Congress shall have power to levy and collect Taxes, Duties, Imposts, and Excises, to pay the Debts and provide for the Common Defense and General Welfare of the United States."

Three restrictions were imposed: the government could not levy taxes on exports; "all Duties, Imposts, and Excises" had to be "uniform throughout the United States" (referred to as the *uniformity clause*); and "no capitation or other direct tax shall be laid, unless in proportion to the Census or Enumeration herein before directed to be taken" (referred to as the *apportionment clause*). (A capitation tax is a tax levied on each person. These taxes are also called *head taxes* or *poll taxes*. They are no longer levied by any state.)

The constitutional provision restricting direct taxes proved to be a problem. Congress levied an income tax during the Civil War and reenacted it in 1894 as a tax on very high incomes. However, it was declared unconstitutional by the Supreme Court in 1895. The Court held that the individual income tax was, in part, a direct tax, which the Constitution stipulates must be apportioned among the states according to their population. Widespread criticism of this rule led to a constitutional amendment. The Sixteenth Amendment, ratified in 1913, declares that "Congress shall have the power to levy and collect taxes on incomes, from whatever sources derived, without apportionment among the several states, and without regard to census or enumeration."

---

[17] In many countries when there is a gap between expenditures and revenues, the difference is financed by printing money. This is how the Continental Congress financed the Revolutionary War. (The expression "not worth a continental" arose from the fact that the currency was not highly valued.)

The apportionment provision, however, still may restrict Congress's ability to impose some taxes. Several countries impose national property taxes or wealth taxes, but these are likely to be considered direct taxes and thus precluded in the United States by the apportionment provision.

FEDERAL TAXATION TODAY

The federal government currently relies on five major forms of taxation: (1) the individual **income tax**, (2) **payroll tax** (to finance Social Security and Medicare benefits), (3) **corporate income tax**, (4) **excise tax** (taxes on specific commodities, such as gasoline, cigarettes, airline tickets, and alcohol), and (5) **customs tax** (taxes levied on selected imported goods). Although the individual income tax was the single largest source of tax revenue for the federal government after World War II and accounted for half of government revenues in 2001, its share has dropped in recent years and is now surpassed by social insurance payroll taxes. In 2010, the corporation income tax accounted for another 12.8 percent and customs and excise taxes 4.5 percent of federal government revenue.[18]

Just as there has been a marked shift in the composition of expenditures over the past fifty years, so too has there been a marked change in the source of government revenues. With the two exceptions mentioned previously, the federal government did not impose any income tax on individuals before 1913. The individual income tax accounted for 20 percent or less of government tax revenues in the years preceding World War II, when rates were quadrupled to pay for the war. Instead, the federal government relied heavily on excise, customs, and corporate income taxes.[19] Since then, the individual income tax had been the largest single source of federal revenues until it was recently overtaken by social insurance and retirement receipts, as shown in Figure 2.9. The corporation income tax has played a decreasing role, falling from 36 percent of federal revenues in 1927 to 23 percent in 1960 and 13 percent in 2010.

Between 1789 and 1909, the federal government received almost all its revenues from excise taxes and customs. Today, those taxes are relatively unimportant. On the other hand, the payroll tax, which was introduced by the Social Security Act of 1935, increased from 10 percent of federal revenues in 1953 to 41 percent in 2010.

---

[18] U.S. Department of Commerce, Bureau of Economic Analysis, *National Income and Product Accounts*, Table 3.2.
[19] For a historical summary of the major federal taxes, see Joseph Pechman, *Federal Tax Policy*, 5th ed. (Washington DC: Brookings Institution, 1987), Appendix A.
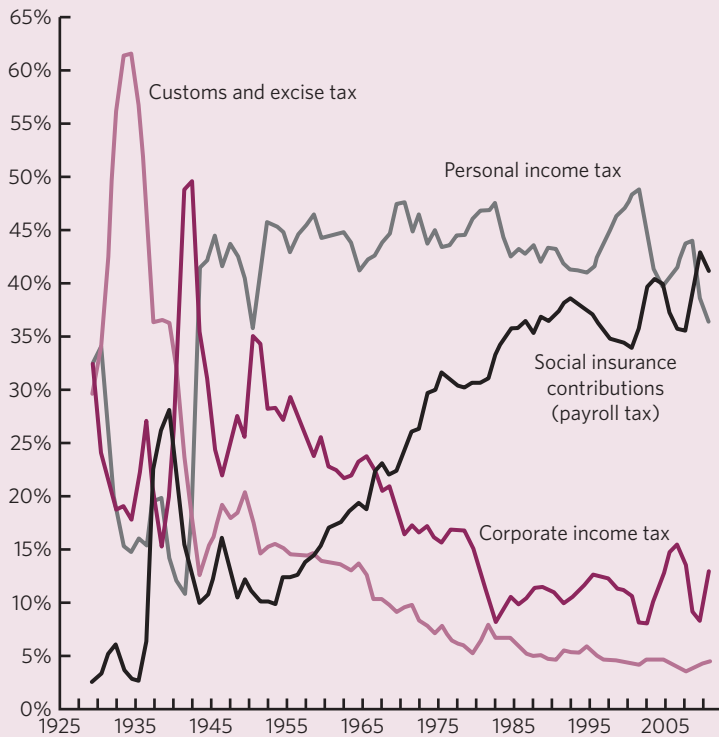
STATE AND LOCAL GOVERNMENT REVENUES

Unlike the federal tax system, state and local tax systems rely heavily on sales and property taxes. As shown in Figure 2.10, until the 1970s property taxes were their major source of revenue. Today, property and sales taxes each contribute about 20 percent of their total revenue. State and local individual income taxes amount to only 15 percent of the total, whereas corporate income taxes are between 2 and 3 percent.

Competition among states for industry discourages the use of some state and local taxes, especially corporate income taxes. The federal government provides substantial aid to state and local governments, much of it directed at specific programs, such as road construction, mass transit, bilingual education, vocational education, and libraries. Over the past decade, federal grants to state and local governments have provided one-fifth of their revenue; this share rose to one-fourth of their revenue in 2009 and 2010 as part of the federal stimulus to combat the nation's deep economic recession.
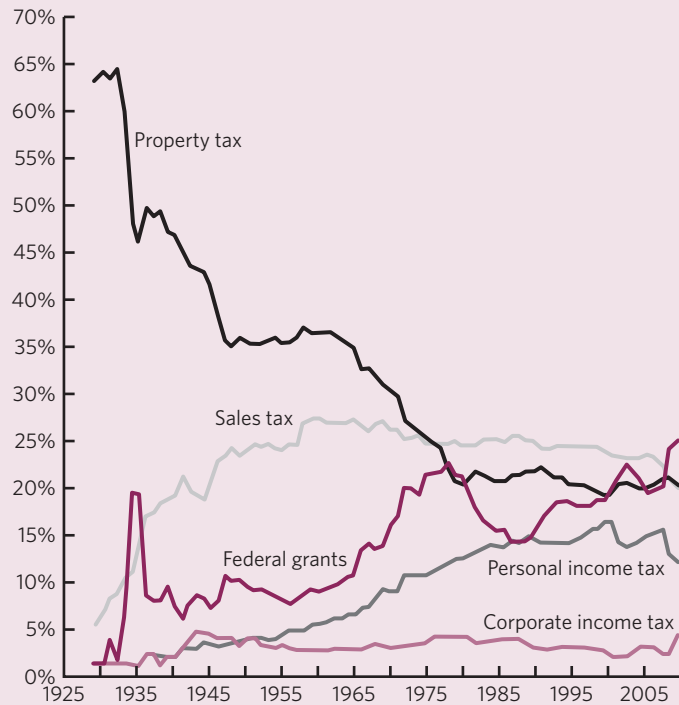
Note: Percentages reflect tax revenues in 2006.

COMPARISON OF TAXATION ACROSS COUNTRIES

As indicated in Figure 2.11, patterns of taxation differ from country to country. Although in most European countries the income tax is less important than in the United States—in 2006, it was almost half of tax revenue in the United States but averaged only 36 percent in Organisation for Economic Cooperation and Development (OECD) countries—taxes on goods and services are more important, averaging almost a third of tax revenue in OECD countries but only 17 percent in the United States. Outside the United States, the **value-added tax** (a tax imposed on the value of the output of a firm less the value of goods and services purchased from other firms) is a major source of revenue. Social Security taxes comprise about the same share of government revenues in the United States and other OECD countries. In contrast to

GOVERNMENT REVENUE AS A PERCENTAGE OF GDP, 2009

The United States has a comparatively diverse set of revenue sources when all levels of government are aggregated, as do the United Kingdom, Italy, and Tunisia.

high-income countries, lower-income countries tend to rely more on indirect than direct taxes, as the small size of their formal economies make it very difficult to collect payroll-based income and social security taxes.[20]

DEFICIT FINANCING

The major source of financing of government expenditures is taxes. But many governments, especially in recent years, have found tax revenues insufficient to pay for their expenditures. A **deficit** in any period is the excess of spending over revenues. A deficit is financed by borrowing. The cumulative value of borrowing by a firm, household, or government is its debt.

---

[20] IMF, *Government Finance Statistics Yearbook 2010*, Tables W3 and W4.

A firm or household that runs a deficit cannot continue to borrow indefinitely, but will be forced into bankruptcy once its debt gets too large. Because of the federal government's ability to tax, and the huge potential revenue sources it can tap, its deficits do not cause the same kinds of problems that large debts incurred by private firms or individuals would. Lenders will continue to willingly finance the federal government's debt, provided the interest rate is high enough.

In the early 1980s, the size of the federal deficit, both in dollar terms and, more importantly, as a fraction of GDP and of the budget, reached all-time highs (for peacetime); see Figure 2.12. The size of the deficits in the 1980s caused great consternation both in and outside Washington. To finance the deficit, the role of the federal government as a *borrower* in U.S. credit markets soared.

FIGURE 2.12

THE FEDERAL BUDGET DEFICIT AS A PERCENTAGE OF EXPENDITURES AND OF GDP, 1929–2010

The deficit increased markedly during the early 1980s, fell, and then increased again during the early 1990s. A federal budget surplus was achieved from 1998 to 2001, but the United States resumed running deficits in 2002, culminating in 10 percent of GDP in 2009.

SOURCES: Office of Management and Budget, *Historical Tables*, Tables 1.1 and 1.2; and U.S. Department of Commerce, Bureau of Economic Analysis, *National Income and Product Accounts*, Table 1.1.5.
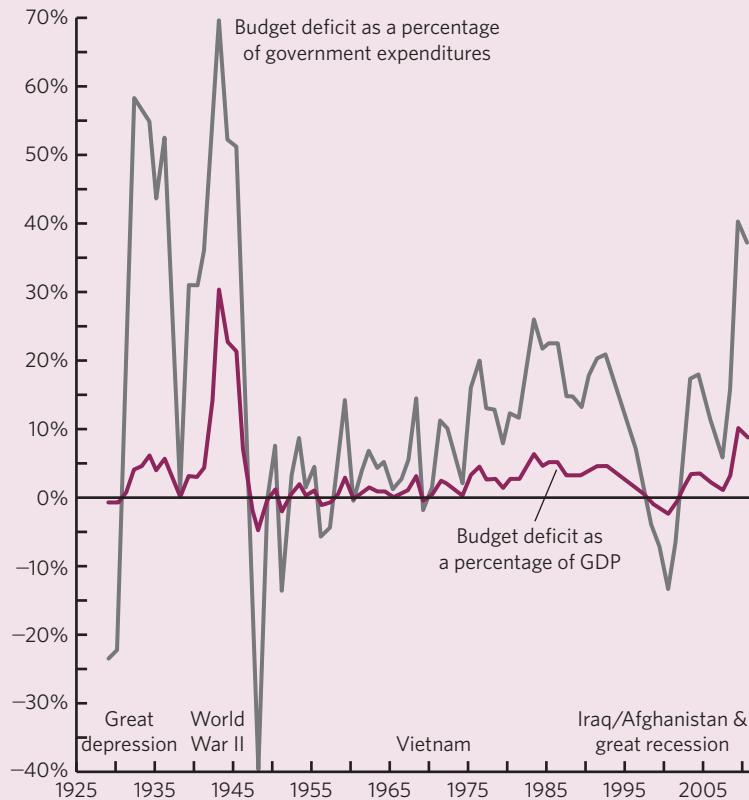
The dollar value of the debt goes up each year by the amount of that year's federal deficit (a federal surplus reduces the federal debt). However, the *real* value of the debt also depends very much on inflation. To see what this means, assume you promise to pay someone $100 next year. If the prices of all goods and services rise by 10 percent, next year that person will be able to purchase with $100 the same goods that he or she could have purchased with $91 this year. The "real value" of what you have to pay has declined by $9.

Figure 2.13 traces the changes since 1940 in the real value of the total gross federal debt as well as the portion of this debt owed to U.S. citizens and foreigners, commonly refered to as the *publicly held federal debt*. In real terms, the increase in the debt after 1980 is dramatic. As a result of the high deficits and the fall in the inflation rate, the period 1981–1989 saw a doubling of the publicly held real debt, from $1.65 trillion in 1981 to $3.30 trillion in 1989 (both amounts measured in 2005 prices). To put it another way, during the Reagan administration, the total *increase* in *real* debt of the federal
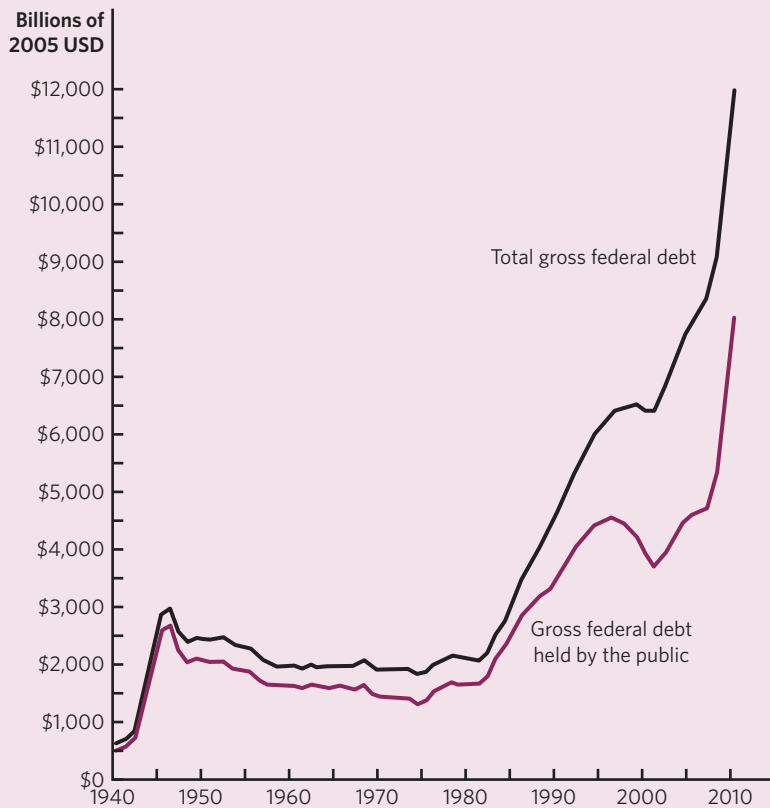
FIGURE 2.13

GROSS FEDERAL DEBT (2005 PRICES), 1940–2010
Federal debt held by the public doubled from 1981 to 1989, and then doubled again over the next two decades.

SOURCE: Office of Management and Budget, *Historical Tables*, Tables 1.3 and 7.1.

government was equal to the total real debt accumulated over the first two hundred years of existence of the United States, including the entire debt required to finance U.S. participation in World War II.

In 1998, the expanding economy, which had grown strongly since 1993, the tax increases enacted in 1993, and the limitations on expenditures that had been imposed for almost a decade, beginning in 1990, finally achieved their long-sought goal: there was a $70 billion surplus. Further surpluses were achieved through the end of the Clinton administration, but subsequent substantial tax cuts and the cost of the conflicts in Iraq and Afghanistan resulted in renewed annual deficits and a rapidly growing federal debt. These trends were exacerbated by the recent recession; in 2009, the deficit was 10 percent of GDP and the publicly held real debt had more than doubled again, totaling $6.80 trillion.

To put the United States in a comparative international perspective (see Figure 2.14), the United Kingdom ran a central government deficit of 10.0 percent of GDP and Japan 9.0 percent in 2009, close to the

FIGURE 2.14

CENTRAL GOVERNMENT DEFICITS AND GENERAL GOVERNMENT DEBT, 2009 (% OF GDP)

The United States ran deficits comparable to those of the United Kingdom and Japan in 2009, and only Australia had a significantly lower debt in 2009. Japan's chronic large annual deficits have resulted in a debt at more than double its GDP, while Chile stands out with remarkably low deficit and debt figures.
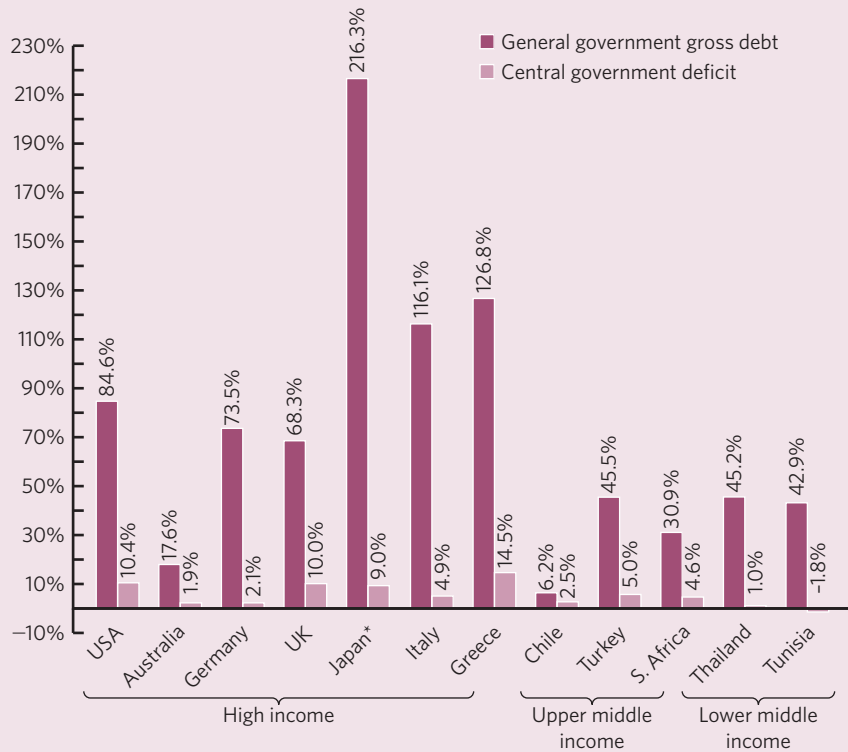
* Japan deficit is for general government.

10.4 percent federal deficit, while both Australia and Germany's deficits were much lower, at 1.9 and 2.1 percent of GDP, respectively. However, at 17.6 percent of GDP, only Australia's general government debt was significantly lower than that of the United States (84.6 percent), the United Kingdom (68.3 percent), and Germany (73.5 percent), whereas Japan's stood at an alarming 216.3 percent of GDP in 2009. In contrast to these high-income countries, Chile, an upper-middle-income country, ran a deficit of just 2.5 percent of GDP and had a debt of only 6.2 percent of GDP in 2009.

PLAYING TRICKS WITH THE DATA ON GOVERNMENT ACTIVITIES

The budgets of the federal, state, and local governments set out their expenditures and receipts. As we have seen, however, budgets provide only a partial view of the size of government and the effect of government on economic activity. As a result, one must treat with caution any comparisons of the size of the public sector either over time or across countries.

Earlier in this chapter, the sections on government subsidies and credits discussed how tax expenditures may result in misleading conclusions concerning not only the size of the public sector but also the composition of its expenditures. If the federal government wishes to hide the size of its subsidies to business, it provides tax credits to businesses. It hides the extent of its subsidies to states and localities by providing "tax expenditures" in the form of tax deductions on the federal individual income tax for most state and local taxes and tax exemption for interest on state and local bonds.

There is a second method by which the budget may be manipulated: by recording the revenues obtained when assets are sold, but not the cost— the reduction in the assets of the government. Such tricks were important in President Reagan's attempt to reduce the deficit. For instance, he accelerated the sale of offshore oil and gas leases.

Speeding tax collections by increasing withholding or by increasing penalties for failing to pay taxes in a timely fashion is another one-time way of reducing a current deficit.

The overall size of the public sector (but not the deficit) can be decreased by setting up independent agencies and enterprises. It makes no real difference whether the post office is a department of the U.S. government or, as is

the case today, a separate "corporation" receiving a subsidy from the federal treasury. If it is a department, however, all its income and all its expenditures will be included in the government budget; if it is a separate enterprise, only the deficit (the difference between its expenditures and income) is recorded. Similarly, placing special government funds off budget, such as those established to pay government pension obligations, also decreases the reported overall size of the public sector. Excluding quasi-fiscal activities of the central bank, such as subsidies for special credit programs or preferred exchange rates, also understates the size of government. The U.S. federal government switched to a unified (consolidated) budget in 1969 to address many such presentational issues.[21]

Although these problems provide considerable room for politicians to select statistics to support their views, the pattern of changes in the level and structure of expenditures and taxation in the United States since World War II has been significant enough that there can be little question about three major observations that have been made in this chapter:

1. The public sector exerts a major influence on the production of goods and the distribution of income in the United States.

2. Social insurance has been the fastest-growing category of government expenditures in the past thirty years. Since 1960, the rapid growth in nondefense expenditure by government was largely accounted for by Social Security, government retirement programs, Medicare, and interest.

3. The individual income tax and social insurance payroll taxes have become the principal sources of federal revenue, whereas the role of the corporation income tax as a revenue source has dwindled.

---

[21] For a detailed discussion of what the United States includes and excludes in its budget presentations, see Office of Management and Budget, "Coverage of the Budget," *Analytical Prespectives, Budget of the United States Government, Fiscal Year 2012.*

REVIEW AND PRACTICE

SUMMARY

1. The government performs many roles:

    a. It provides the basic legal framework within which we live.

    b. It regulates economic activities. It encourages some activities by subsidizing them and discourages others by taxing them.

    c. It produces goods and provides credit, loan guarantees, and insurance.

d. It purchases goods and services, including many that are produced by private firms, such as weapons manufacturers.

e. It redistributes income, transferring income from some individuals to others.

f. It provides social insurance, for retirement, unemployment, disability, and medical care for the aged.

2. The size of the government relative to GDP is much larger now than it was forty years ago. Much of this increase is accounted for by increased payments for social insurance.

3. The relative size of the public sector in the United States is smaller than in most western European countries.

4. The three major areas of government expenditures are defense, education, and transfers. Together, these accounted for 72 percent of governmental expenditures in 2009.

5. The major source of revenue for the federal government is the payroll tax, followed by the individual income tax, corporation tax, and customs and excise taxes.

6. The major sources of revenue for state and local government are the sales tax, the property tax, and the income tax.

7. The Constitution provides the basic framework for the government of the United States. It provides some restrictions on the taxes that can be imposed, but no effective restrictions on what the government can spend its money on.

8. The deficit—the difference between the government's expenditures and revenues—grew enormously, beginning in 1981, with the total real debt accumulated from 1981 to 1988 alone equaling the total real debt accumulated over the first two hundred years of the country's existence. Deficit reduction measures, begun in 1990 and extended in 1993, combined with a growing economy, enabled a surplus to be achieved in 1998. Subsequent tax cuts and the cost of the Iraq and Afghanistan conflicts, together with the recent recession, have resulted in a doubling of the publicly held real debt since then, and a 2009 deficit equal to 10 percent of GDP.

KEY CONCEPTS

Corporate income tax

Customs tax

Deficit

Excise tax

Income tax

In-kind benefits

Middle-class entitlement programs

Nationalization

Payroll tax

Social insurance

Transfer payments

Value-added tax

QUESTIONS AND PROBLEMS

1. To see what is going on, economists often "adjust" the data to reflect changes in the economy. For instance, in the text, we discussed the adjustments in dollar amounts made to correct for inflation. Another adjustment that is frequently made is to take into account the increase in population. What adjustments might you make in looking at education expenditures? At Social Security expenditures?

2. In each of the following areas, give one or more examples (where possible) in which the government is involved as a producer, a regulator, or a purchaser of final goods and services distributed directly to individuals or used within government:

   a. Education

   b. Utilities

   c. Transportation

   d. Credit markets

   e. Insurance markets

   f. Food

   g. Housing

3. In each of the following areas, give an example of a tax expenditure and a conventional expenditure. Explain how the same results could be obtained by converting the tax expenditure into a conventional expenditure.

   a. Medicine

   b. Housing

   c. Education

4. Assume you were President and your planned expenditures exceeded your receipts. Describe some of the tricks you might use to reduce the apparent budget deficit while maintaining current levels of services and transfers (subsidies).

   Assume, on the other hand, that you had run on a platform of keeping the growth in total governmental expenditures down to 3 percent. Once in office, you see, however, that you would like expenditures to rise by 5 percent. How might you do this while appearing to keep your election promises?

FUNDAMENTALS OF WELFARE ECONOMICS

Most economies today are mixed economies, in which there is both a private and a public sector. At the core of the economy are profit-maximizing firms interacting with households in competitive markets. Under certain idealized conditions, a competitive economy is efficient. If those conditions were satisfied, there would be a very limited role for government. To understand the role of the public sector then, we have to understand when markets work well, and when—and in what ways—they do not. That is the objective of this part of the book.

Chapter 3 explains what is entailed by efficiency, and why, under idealized conditions, competitive economies are efficient. Chapter 4 then explains the variety of reasons why and circumstances in which markets may fail to produce efficient outcomes, and why, even if the economy were efficient, there might be a role for government in redistributing income. Chapter 5 builds on this overview of market failure to produce efficient and equitable outcomes by exploring the unique nature of public goods, why these are undersupplied by private markets, and the rationale for publicly provided private goods. Chapter 6 examines another type of market failure in greater depth, namely, externalities.

The hardest choices facing the public sector involve trade-offs, in particular, trade-offs between increased efficiency and a more equitable distribution of income. Chapter 7 provides a conceptual framework for thinking about these trade-offs, some tools that are used by governments in attempting to quantify them, and an analysis of the circumstances in which one can have both more efficiency and more equity.

3

MARKET EFFICIENCY

In most modern industrial economies, primary reliance for the production and distribution of goods lies in the private rather than the public sector. One of the most enduring tenets of economics holds that this form of economic organization leads to an efficient allocation of resources—but if private markets are efficient, why should there be an economic role for government? To answer this question, a precise understanding of the meaning of economic efficiency is needed. That is the aim of this chapter. The next chapter will consider why private markets may fail to achieve efficient outcomes and how government may respond to these market failures.

THE INVISIBLE HAND OF COMPETITIVE MARKETS

In 1776, Adam Smith, in *The Wealth of Nations*—the first major work of modern economics—argued that competition would lead the individual

in the pursuit of his or her private interests (profits) to pursue the public interest, as if by an **invisible hand**:

> [H]e intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it.[1]

The significance of Smith's insight is clarified by a look at the views about the role of government commonly held prior to Smith. There was widespread belief that achieving the best interests of the public (however that might be defined) required an active government. This view was particularly associated with the mercantilist school of the seventeenth and eighteenth centuries, which argued that government should promote industry and trade. Indeed, many European governments had actively promoted the establishment of colonies, and the mercantilists provided a rationale for this.

Some countries (or some citizens within them) had benefited greatly from the active role taken by their government, but other countries, whose governments had been much more passive, had also prospered. And some countries with strong, active governments had not prospered, as their resources were squandered on wars or on a variety of unsuccessful public ventures.

In the face of these seemingly contradictory experiences, Smith addressed himself to the question: Can society ensure that those entrusted with governing actually pursue the public interest? Experience had shown that although at times the policies governments pursued seemed consistent with the public good, at other times, the policies pursued could not by any reasonable stretch of the imagination be reconciled with the public good. Rather, those in the position of governing sometimes seemed to pursue their private interests at the expense of the public interest. Moreover, even well-intentioned leaders often led their countries astray. Smith argued that it was not necessary to rely on government or on any moral sentiments to do good. The public interest, he maintained, is served when each individual simply does what is in his or her own self-interest. Self-interest, Smith argued, is a much more persistent characteristic of human nature than a concern to do good, and therefore provides a more reliable basis for the organization of society. Moreover, individuals are more likely to ascertain with some accuracy what is in their own self-interest than they are to determine what is in the public interest.

---

[1] Adam Smith, *The Wealth of Nations* (New York: Modern Library, 1937); originally published in 1776.

The intuition behind Smith's insight is simple: if there is some commodity or service that individuals value but that is not currently being produced, then they will be willing to pay something for it. Entrepreneurs, in their search for profits, are always looking for such opportunities. If the value of a certain commodity to a consumer exceeds the cost of production, there is a potential for profit, and an entrepreneur will produce the commodity. Similarly, if there is a cheaper way of producing a commodity than that which is presently employed, an entrepreneur who discovers this cheaper method will be able to undercut competing firms and make a profit. The search for profits on the part of enterprises is thus a search for more efficient ways of production, and for new commodities that better serve the needs of consumers.

In this view, no government committee needs to decide whether a commodity should or should not be produced. It will be produced if it meets the market test—that is, if what individuals are willing to pay exceeds the costs of production. Nor does any government oversight committee need to check whether a particular firm is producing efficiently: competition will drive out inefficient producers.

There is widespread consensus among economists that competitive forces do lead to a high degree of efficiency, and that competition does provide an important spur to innovation. However, over the past two hundred years, economists have come to recognize that in some important instances the market does not work as perfectly as the more ardent supporters of the free market suggest. Economies have gone through periods of massive unemployment and idle resources; the Great Depression of the 1930s left many who wanted work unemployed; pollution has choked many of our larger cities; and urban decay has set in on others.

WELFARE ECONOMICS AND PARETO EFFICIENCY

Welfare economics is the branch of economics that focuses on what were termed *normative issues* in Chapter 1. The most fundamental normative issue for welfare economics is the economy's organization—what should be produced, how it should be produced, for whom, and who should make these decisions. In Chapter 1, we noted that the United States and most other economies today are *mixed*, with some decisions made by the government but most left up to the myriad firms and households. But there are many "mixes." How are we to evaluate the alternatives? Most economists embrace a criterion called **Pareto efficiency**, named after the great Italian economist

ON THE PROWL FOR PARETO IMPROVEMENTS

Although finding Pareto improvements is difficult, economists are constantly on the lookout for such opportunities. Two recent proposals illustrate some of the problems that may be encountered.

One proposal concerned offshore oil wells. The federal government leases the land to oil companies in return for a royalty, usually around 16 percent. The oil companies compete for these leases in competitive auctions; the lease goes to the firm offering the highest bid. As oil wells get old, the cost of extraction increases, often to the point at which, with the royalty taken into account, it pays to shut down the well. If the price of oil is $20 a barrel, for instance, and there is a 16 percent royalty, it pays to shut down the well when the cost of extraction exceeds $16.80 ($16.80 plus the $3.20 royalty equals the $20 received). This seems inefficient, as the value of the oil ($20) exceeds the cost of production. Hence, there have been proposals to eliminate royalties on old wells and to allow the oil companies to pay a fixed up-front fee. The government is no worse off (because if the well is shut down it receives no revenue), and, provided the fee is set low enough, the oil company is better off (because if the well is shut down it receives nothing). The oil companies have resisted the proposal: they prefer that the government simply eliminate royalties. Although the proposal is a Pareto improvement over the status quo, they would prefer to garner for themselves more of the potential gains from the increased economic efficiency.

A second proposal involved allowing private companies to construct improved turbines at hydroelectric sites, increasing the energy output. They would be allowed to sell the electricity at market prices. Hydroelectric energy is particularly attractive, since it generates no pollution. There would be no adverse environmental impacts, as the developments would occur only at sites already being used. This too appeared to be a Pareto improvement: economic efficiency would be increased as cheaper hydroelectric power replaced power relying on fossil fuels; the benefits of the improved efficiency would be shared among consumers, investors, and the government; and future generations would be better off as a result of the more favorable environmental impacts. This proposal was opposed by utility companies that currently got electricity from these dams at below-market prices. Although the proposal did not alter the current level of preferential treatment, they were worried that once the principle that electricity from hydroelectric sites could be sold at market prices was established, their preferential treatment would be threatened. Even though the proposal as framed was a Pareto improvement, the utilities saw the long-run consequences of the proposal as a gain in efficiency at the expense of their future welfare.

and sociologist Vilfredo Pareto (1848–1923). Resource allocations that have the property that no one can be made better off without someone being made worse off are said to be *Pareto efficient*, or *Pareto optimal*. Pareto efficiency is what economists normally mean when they talk about efficiency.

Assume, for instance, that the government is contemplating building a bridge. Those who wish to use the bridge are willing to pay more than enough

in tolls to cover the costs of construction and maintenance. The construction of this bridge is likely to be a **Pareto improvement**; that is, a change that makes some individuals better off without making anyone worse off. We say "likely" because there are always others who might be adversely affected by the construction of the bridge. For example, if the bridge changes the traffic flow, some stores might find that their business is decreased, and they are worse off; or an entire neighborhood may be affected by the noise of bridge traffic and the shadows cast by the bridge superstructure.

Frequently, on summer days or at rush hour, large backups develop at tollbooths on toll roads and bridges. If tolls were raised at those times and the proceeds used to finance additional tollbooths or more peak-time toll collectors, everyone might be better off. People would prefer to pay a slightly higher price in return for less waiting. Even this change might not be a Pareto improvement, though: among those waiting in line may be some unemployed individuals who are relatively not concerned about the waste of time but who are concerned about spending more money on tolls.

Economists are always on the lookout for Pareto improvements. The belief that any such improvements should be instituted is referred to as the **Pareto principle**.

"Packages" of changes together may constitute a Pareto improvement, when each change alone might not. Thus, although reducing the tariff on steel would not be a Pareto improvement (because steel producers would be worse off), it might be possible to reduce the tariff on steel, increase income taxes slightly, and use the proceeds to finance a subsidy to the steel industry. Such a combination of changes might make everyone in the country better off, and make those abroad—the foreign exporters of steel—also better off.

PARETO EFFICIENCY AND INDIVIDUALISM

The criterion of Pareto efficiency has an important property that requires comment. It is *individualistic*, in two senses. First, it is concerned only with each individual's welfare, not with the relative well-being of different individuals. It is not concerned explicitly with inequality. Thus, a change that made the rich much better off but left the poor unaffected would still be a Pareto improvement. Some people, however, think that increasing the gap between the rich and the poor is undesirable. They believe that it gives rise, for instance, to undesirable social tensions. Less developed countries often go through periods of rapid growth during which all major segments of society become better off but the income of the rich grows more rapidly than that of the poor. To assess these changes, is it enough simply to say that everyone is better off? There is no agreement on the answer to this question.

Second, it is each individual's perception of his or her own welfare that counts. This is consistent with the general principle of **consumer sovereignty**, which holds that individuals are the best judges of their own needs and wants; that is, of what is in their own best interests.

THE FUNDAMENTAL THEOREMS OF WELFARE ECONOMICS

Two of the most important results of welfare economics describe the relationship between competitive markets and Pareto efficiency. These results are called the **fundamental theorems of welfare economics**. The first theorem tells us that if the economy is *competitive* (and satisfies certain other conditions), it is Pareto efficient.

The second theorem asks the reverse question. There are many Pareto efficient distributions. By transferring wealth from one individual to another, we make the second individual better off and the first worse off. After we make the redistribution of wealth, if we let the forces of competition freely play themselves out, we will obtain a Pareto efficient allocation of resources. This new allocation will be different in many ways from the old. If we take wealth away from those who like chocolate ice cream and give it to those who like vanilla, in the new equilibrium, more vanilla ice cream will be produced and less chocolate, but no one can be made better off in the new equilibrium without making someone else worse off.

Let's say there is a *particular* distribution that we would like to obtain. Assume, for instance, that we care particularly about the aged. The second fundamental theorem of welfare economics says that the *only* thing the government needs to do is redistribute initial wealth. *Every Pareto efficient resource allocation can be obtained through a competitive market process with an initial redistribution of wealth*. Thus, if we do not like the income distribution generated by the competitive market, we need not abandon the use of the competitive market mechanism. All we need do is redistribute the initial wealth, and then leave the rest to the competitive market.

The second fundamental theorem of welfare economics has the remarkable implication that every Pareto efficient allocation can be attained by means of a *decentralized market mechanism*. In a decentralized system, decisions about production and consumption (what goods get produced, how they get produced, and who gets what goods) are carried out by the myriad firms and individuals that make up the economy. In contrast, in a **centralized allocation mechanism**, all such decisions are concentrated in the hands of a single agency—the central planning

agency—or a single individual, who is referred to as the *central planner*. Of course, no economy has even come close to being fully centralized, although under communism in the former Soviet Union and some of the other Eastern bloc countries, economic decision making was much more concentrated than in the United States and other Western economies. Today, only Cuba and North Korea place heavy reliance on central planning.

FUNDAMENTAL THEOREMS OF WELFARE ECONOMICS

- Every competitive economy is Pareto efficient.
- Every Pareto efficient resource allocation can be attained through a competitive market mechanism, with the appropriate initial redistributions.

The second fundamental theorem of welfare economics says that to attain an efficient allocation of resources, with the desired distribution of income, it is not necessary to have a central planner, with all the wisdom an economic theorist or a utopian socialist might attribute to him or her; competitive enterprises, attempting to maximize their profits, can do as well as the best of all possible central planners. This theorem thus provides a major justification for reliance on the market mechanism. Put another way, if the conditions assumed in the second welfare theorem were valid, the study of public finance could be limited to an analysis of the appropriate governmental redistributions of resources.

Why the competitive market, under ideal conditions, leads to a Pareto optimal allocation of resources is one of the primary subjects of study in standard courses in microeconomics. Because we will be concerned with understanding why, under some circumstances, competitive markets do not lead to efficiency, we first need to understand why competition, under ideal conditions, leads to efficiency. Before turning to this, though, it is important to emphasize that these results are *theorems;* that is, logical propositions in which the conclusion (the Pareto efficiency of the economy) follows from the assumptions. The assumptions reflect an ideal competitive model, in which, for instance, there are many small firms and millions of households, each so small that it has no effect on prices; in which all firms and households have perfect information, say, concerning the goods that are available in the market and the prices which are being charged; and in which there is no air or water pollution.[2] The accuracy of these assumptions in portrayal of our economy and the robustness of the results—the extent to which the conclusions change when the assumptions change—are two of the main subjects of debate among economists. In the next chapter, we look at some of the important ways in which markets fail to deliver efficient outcomes; that is, we identify important circumstances in which the ideal conditions underlying the fundamental theorems of welfare economics are not satisfied.

---

[2] There are also a number of technical assumptions.

EFFICIENCY FROM THE PERSPECTIVE OF A SINGLE MARKET[3]

We can see why competition results in economic efficiency using traditional demand and supply curves. The demand curve of an individual gives the amount of the good the individual is willing to demand at each price. The market demand curve simply adds up the demand curves of all individuals: it gives the total quantity of the good that individuals in the economy are willing to purchase, at each price. As Figure 3.1 illustrates, the demand curve is normally downward sloping: as prices increase, individuals demand less of the good. In deciding how much to demand, individuals equate the **marginal (additional) benefit** they receive from consuming an extra unit with the **marginal (additional) cost** of purchasing an extra unit. The marginal cost is just the price they have to pay.

The supply curve of a firm gives the amount of the good the firm is willing to supply at each price. The market supply curve simply adds up the supply curves of all firms: it gives the total quantity of the good that firms in the economy are willing to supply, at each price. As Figure 3.1 illustrates, the supply curve is normally upward sloping: as prices increase, firms are willing to supply more of the good. In deciding how much of a good to produce, competitive firms equate the marginal (additional) benefit they receive from producing an extra unit—which is just

---

[3] This is often called the *partial equilibrium approach*, in contrast to the *general equilibrium approach* that looks at all markets simultaneously. We take the latter approach in the next section.

FIGURE 3.1

EFFICIENCY FROM THE PERSPECTIVE OF A SINGLE MARKET

In deciding how much to demand, individuals equate the marginal benefit they receive from consuming an extra unit with the marginal cost, the price they have to pay. In deciding how much to supply, firms equate the marginal benefit they receive, which is just the price, with the marginal cost. At the market equilibrium, where supply equals demand, the marginal benefit (to consumers) is equal to the marginal cost (to firms)—and each equals the price.
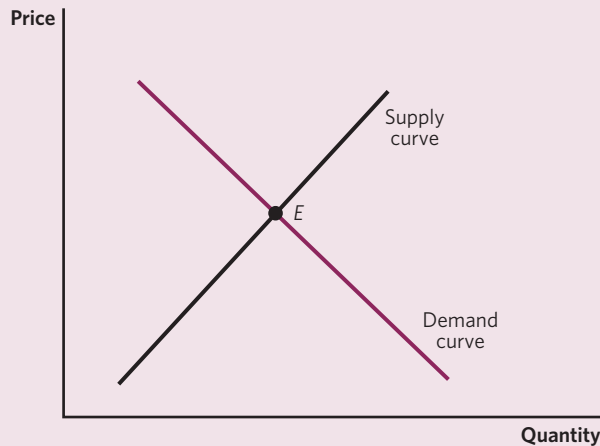
the price they receive—with the marginal (additional) cost of producing an extra unit.

Efficiency requires that the marginal benefit associated with producing one more unit of any good equal its marginal cost—for if the marginal benefit exceeds the marginal cost, society would gain from producing more of the good; if the marginal benefit was less than the marginal cost, society would gain from reducing production of the good.

Market equilibrium occurs at the point at which market demand equals supply, point *E* in Figure 3.1. At this point, the marginal benefit and the marginal cost each equal the price; thus, the marginal benefit equals the marginal cost, which is precisely the condition required for economic efficiency.

ANALYZING ECONOMIC EFFICIENCY

To develop a deeper analysis that goes beyond the basic supply and demand framework just presented, economists consider three aspects of efficiency, all of which are required for Pareto efficiency. First, the economy must achieve **exchange efficiency**; that is, whatever goods are produced have to go to the individuals who value them most. If I like chocolate ice cream and you like vanilla ice cream, I should get the chocolate cone and you the vanilla. Second, there must be **production efficiency**. Given the society's resources, the production of one good cannot be increased without decreasing the production of another. Third, the economy must achieve **product mix efficiency** so that the goods produced correspond to those desired by individuals. If individuals value ice cream a lot relative to apples, and if the cost of producing ice cream is low relative to apples, then more ice cream should be produced. The following sections examine each of these types of efficiency in turn.

THE UTILITY POSSIBILITIES CURVE

In preparation for learning what is entailed by each of the three aspects of Pareto efficiency, the concept of the **utility possibilities curve** is useful. Economists sometimes refer to the benefits that an individual gets from consumption as the *utility* that the individual gets from the combination of goods he or she consumes.[4] If the person gets more goods,

---

[4] The concept of utility is only a useful way of thinking about the benefits that an individual gets from consumption. There is no way of measuring utility other than indirectly by looking at what individuals are willing to pay; no machine can ascertain the number of "utiles," or whatever the unit of measurement of utility might be called, derived from eating a pizza or listening to a CD.

THE UTILITY POSSIBILITIES CURVE

The utility possibilities curve gives the maximum level of utility that one individual (Friday) can achieve, given the level of utility of the other individual (Crusoe). Along the frontier, it is not possible for Crusoe to consume more unless Friday consumes less. Therefore, the utility possibilities curve is downward sloping: the higher Crusoe's utility, the lower the maximum level of Friday's utility.
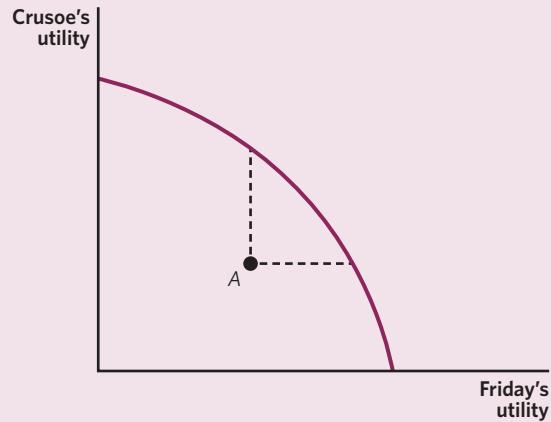
his or her utility has increased. The utility possibilities curve traces out the maximum level of utility that may be achieved by two consumers. Figure 3.2 shows a utility possibilities frontier for Crusoe and Friday, showing Friday's maximum level of utility, given Crusoe's level of utility (and vice versa). Recall the definition of Pareto efficiency: an economy is Pareto efficient if no one can be made better off without making someone else worse off; that is, we cannot increase the utility of Friday without decreasing the utility of Crusoe. Thus, if an economy is Pareto efficient, it must be operating along the utility possibilities frontier. If the economy were operating at a point below the utility possibilities frontier, such as at point *A* in Figure 3.2, it would be possible to increase the utility of Friday or Crusoe without decreasing the utility of the other, or to increase the utility of both.

The first fundamental theorem of welfare economics says that a competitive economy operates along the utility possibilities frontier; the second fundamental theorem of welfare economics says that we can attain any point along the utility possibilities frontier using competitive markets, provided we redistribute initial endowments appropriately.

EXCHANGE EFFICIENCY

Exchange efficiency concerns the distribution of goods. Given a particular set of available goods, exchange efficiency provides that those goods are distributed so no one can be made better off without
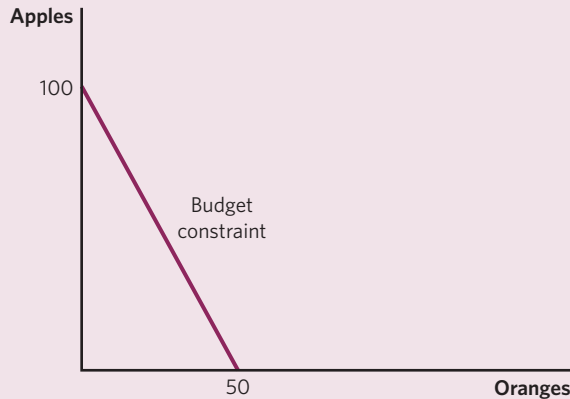
someone else being made worse off. Exchange efficiency thus requires that there is no scope for **trades**, or exchanges that would make both parties better off.

Assume that Crusoe is willing to give up one apple in exchange for one orange, or to get one apple in exchange for giving up one orange. Assume that Friday, on the other hand, is willing to give up three apples if he can get one more orange. At the margin, Friday values oranges more highly than does Crusoe. Clearly, there is room for a deal: if Crusoe gives Friday one of his oranges, and Friday gives Crusoe two of his apples, both are better off. Crusoe would have required only one apple to make him just as well off, but he gets two in exchange for his orange. Friday would have been willing to give up three apples, but he gave up only two, so he is clearly better off.

The amount of one commodity that an individual is willing to give up in exchange for a unit of another commodity is called the **marginal rate of substitution**. As long as Crusoe and Friday's marginal rates of substitution differ, there will be room for a deal. Thus, exchange efficiency requires that all individuals have the same marginal rate of substitution.

We now will see why competitive economies satisfy this condition for exchange efficiency. To do so, we need to review how consumers make their decisions. We begin with the budget constraint—the amount of income a consumer can spend on various goods. Crusoe has $100, which he can divide between apples and oranges. If an apple costs $1 and an orange $2, Crusoe can buy 100 apples or 50 oranges, or combinations in between, as illustrated in Figure 3.3. If Crusoe buys one more orange, he has to give up two apples. Thus, the slope of the budget constraint is equal to the ratio of the prices.

Crusoe chooses the point along the budget constraint that he most prefers. To see what this entails, we introduce a new concept: **Indifference curves** give the combinations of goods among which an individual is indifferent or which yield the same level of utility. Figure 3.4 shows indifference curves for apples and oranges. For example, the indifference curve I_0 gives all the combinations of apples and oranges that the consumer finds just as attractive as 80 apples and 18 oranges (point *A* on the indifference curve). If points *A* and *B* are on the same indifference curve, the consumer is indifferent between the two combinations of apples and oranges represented by the two points. The indifference curve also shows how much of one good (apples) the consumer is willing to give up in return for one more unit of another good (oranges). The amount of one good the individual is willing to give up in return for one more unit of another good is just the marginal rate of substitution, which we defined earlier. Thus, *the slope of the indifference curve equals the marginal rate of substitution*. In Figure 3.4, in moving from point *A* to point *B*, Crusoe gives up one orange, but he is just as well off if he is compensated with nine extra apples. Note that the number of apples that he needs to compensate him for having one less orange is much higher

FIGURE 3.4

THE CONSUMER'S CHOICE PROBLEM

The budget constraint gives the combinations of apples and oranges that Crusoe can buy, given his income and given the price of apples and oranges. The indifference curve gives the combinations of apples and oranges among which Crusoe is indifferent. *A* and *B* are on the same indifference curve; Crusoe is indifferent between them. Other individuals prefer combinations of apples and oranges that are on a higher indifference curve. Thus, point *F* is preferred to either *A* or *B*. Crusoe chooses the point along the budget constraint that he most prefers; that is, the point at which the indifference curve I_0 is tangent to the budget constraint (point *E*).
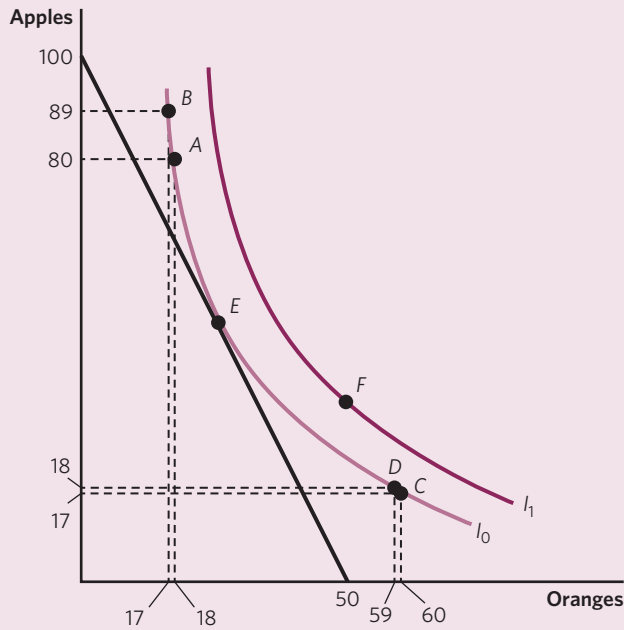
when he moves from *A* to *B* than when he moves from *C* to *D*. When he has
60 oranges, he is much more willing to give up one of his oranges: he needs
only one more apple to compensate him. Thus, the marginal rate of substitu-
tion diminishes as the number of oranges that Crusoe consumes increases.
This explains why the indifference curves have the shape depicted.

Clearly, individuals are better off if they have more apples and oranges;
that is why combinations of goods along a higher indifference curve give a
higher level of utility. Thus, any of the points on I_1 in Figure 3.4 are more
attractive than the points on I_0. By definition, a consumer does not care
at which point along an indifference curve he or she is placed; but the
consumer wants to be along the highest indifference curve possible. Cru-
soe would like to get to any point along the indifference curve I_1, but he
cannot: all these points lie above the budget constraint, and so are not
feasible. The best that Crusoe can do is to choose point *E*, at which the
indifference curve is tangent to the budget constraint.

At the point of tangency, the slope of the indifference curve is identical
to the slope of the budget constraint, but the slope of the indifference curve
is the marginal rate of substitution, and the slope of the budget constraint
is the price ratio. Thus, individuals choose a combination of apples and
oranges at which the marginal rate of substitution is equal to the price ratio.

Because all consumers face the same prices in a competitive economy,
and each sets his or her marginal rate of substitution equal to the price
ratio, they all have the same marginal rate of substitution. Earlier, we
showed that the condition for exchange efficiency was that all individuals
have the same marginal rate of substitution. Thus, *competitive markets
have exchange efficiency*.

Another way to represent exchange efficiency is illustrated in Figure 3.5.
For simplicity, we continue the example of Crusoe and Friday: whatever Cru-
soe does not get, Friday gets. Thus, we can represent all possible allocations
in a box (called an *Edgeworth–Bowley box* after two early–twentieth-
century English mathematical economists) in which the horizontal axis
represents the total supply of oranges and the vertical axis represents the
total supply of apples. In Figure 3.5, what Crusoe gets to consume is mea-
sured from the bottom left corner (*O*), and what Friday gets is measured
from the top right corner (*O'*). At the allocation denoted by the point *E*,
Crusoe gets *OA* oranges and *OB* apples, while Friday gets the remainder
(*O'A'* oranges and *O'B'* apples). We then draw Crusoe's indifference curves,
such as U^c. We have also drawn Friday's indifference curves. His indiffer-
ence curves look perfectly normal if you turn the book upside down.

Let us now fix Crusoe's utility. Pareto efficiency requires us to
maximize Friday's utility, given the level of utility attained by Crusoe.
Thus, we ask, given that Crusoe is on the indifference curve U^c, what is

FIGURE 3.5

EXCHANGE EFFICIENCY
The sides of this Edgeworth–
Bowley box give the available
supplies of apples and oranges.
OA and *OB* give Crusoe's con-
sumption of the two commod-
ities. Friday gets what Crusoe
does not consume; that is, *O'A'*
and *O'B'*. Pareto efficiency
requires the tangency of the
two indifference curves (one
such point is at *E*), where the
marginal rates of substitution of
apples for oranges are equal.
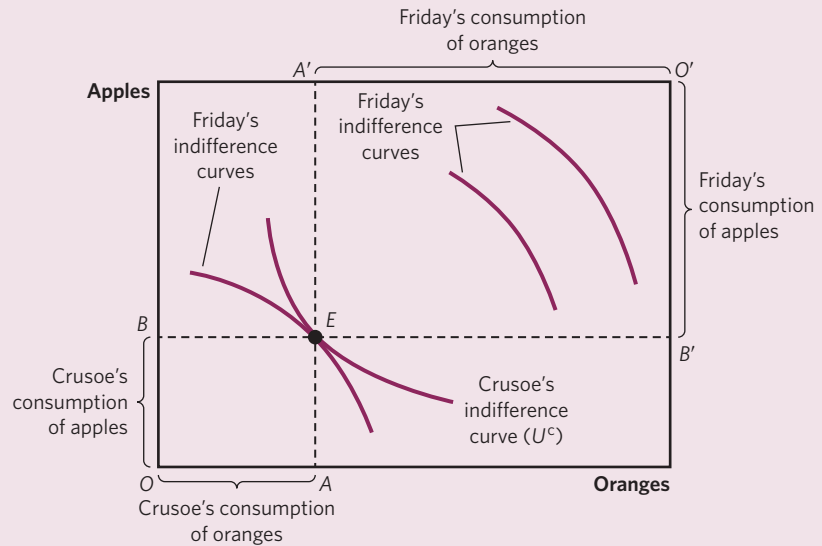
the highest indifference curve that Friday can get to? Remember that
Friday's utility increases as we move down and to the left (Friday is get-
ting more goods, Crusoe fewer goods). Friday attains his highest utility
when his indifference curve is *tangent* to Crusoe's, at *E*. At this point, the
slopes of the indifference curves are the same; that is, their marginal rates
of substitution of apples for oranges are the same.

PRODUCTION EFFICIENCY

If an economy is not productively efficient, it can produce more of one good
without reducing production of other goods. Along the production possibil-
ities frontier in Figure 3.6, the economy cannot produce more of one good
without giving up some of another good, given a fixed set of resources.[5]

The analysis used to determine whether an economy is productively effi-
cient is similar to the one we used above for exchange efficiency. Consider
Figure 3.7. In place of the budget constraint we have an **isocost line**, giving
the different combinations of inputs that cost the firm the same amount.
The slope of the isocost line is the relative price of the two factors. The fig-
ure also shows two **isoquants**. These trace out the different combinations

---

[5] The production possibilities schedule has the shape it does because of the law of diminishing returns.
As we try to produce more and more oranges, it becomes harder and harder to produce an additional
orange. Thus, as we give up apples, we get more oranges, but for each additional apple we give up, we
get fewer and fewer extra oranges.

Apples

Oranges

Land

Isoquants

Q_1

Q_0

Isocost
line

Labor

of inputs—in this case, land and labor—that produce the same quantities of outputs. Thus, isoquants are to the analysis of production what indifference curves are to the analysis of consumption. Economists call the slope of an isoquant the **marginal rate of technical substitution**. In Figure 3.7, the marginal rate of technical substitution is the amount of land required to compensate for a decrease in the input of labor by one unit. When relatively little labor is being used, it is hard to economize further in its use, so if one less worker is used, there must be a large increase in land if output is to remain unchanged. That is why the isoquants have the shape they do. There is a diminishing marginal rate of technical substitution.

Just as exchange efficiency requires that the marginal rate of substitution between any pair of commodities be the same for all individuals, production efficiency requires that the marginal rate of technical substitution be the same for all firms. Assume the marginal rate of substitution between land and labor is 2 in producing apples and 1 in producing oranges. This means that if we reduce labor by one in oranges, we need one more unit of land. If we reduce labor by one in apples, we need two more units of land. Conversely, if we increase labor by one in apples, we need two fewer units of land. Thus, if we take one worker from producing oranges and put him or her to work in apples, and we take one unit of land, and switch it from producing apples to producing oranges, production of oranges is unchanged but production of apples is increased. Whenever the marginal rates of substitution differ, we can switch resources around in a similar way, to increase production.

A firm maximizes the amount of output that it produces at a given level of expenditures on inputs by finding the point at which the isoquant is tangent to the isocost line. At the point of tangency, the slopes of the two curves are the same—the marginal rate of technical substitution is equal to the ratio of the prices of the two inputs. In a competitive economy all firms face the same prices, so all firms using labor and land will set their marginal rate of technical substitution equal to the *same* price ratio. Hence, all will have the same marginal rate of technical substitution—the condition that is required for production efficiency.

In Figure 3.8, we see the same principle diagrammatically, using another Edgeworth–Bowley box. We wish to know how to allocate a fixed supply of inputs to ensure productive efficiency. We represent the fixed supply of the two inputs by a box, with the total available supply of land measured along the vertical axis and the total supply of labor along the horizontal axis. We measure inputs into orange production from the bottom left-hand corner. Thus, point *E* means that the amount *OB* of land is used in orange production, and *OA* of labor. That, in turn, means that the *remaining* inputs are used in the production of apples. Thus, we measure
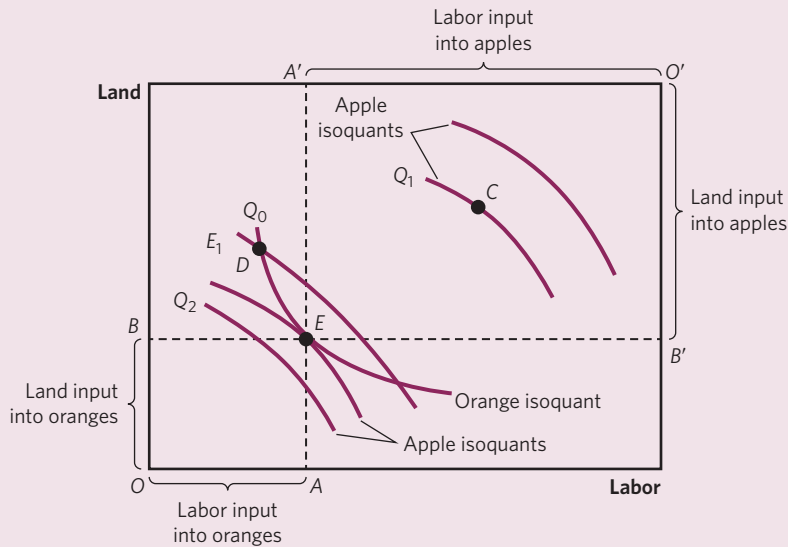
PRODUCTION EFFICIENCY
The sides of this Edgeworth–Bowley box give the available supply of resources—land and labor. Resources used in the production of oranges are given by *OA* and *OB*; resources not used in the production of oranges are used in the production of apples, *O'A'* and *O'B'*. Production efficiency requires the tangency of the isoquants. At tangency points, such as *E*, the marginal rate of substitution of land for labor is the same in the production of apples and oranges.

inputs into apples from the upper right-hand corner. At *E*, the amounts *O'B'* of land and *O'A'* of labor go into apple production.

The isoquants also appear in the figure. Q_0 gives a typical orange isoquant. Remember that the quantities of inputs going into apple production are measured from *O'*. That is why the isoquants for apples have the shape they do; they look perfectly normal if you turn the book upside down. Clearly, production efficiency requires that for any level of production of oranges the output of apples is maximized. As we move down and to the left in the box, more resources are being allocated to apple production; hence, isoquants through those points represent higher levels of apple output. If we fix the output of oranges at the level corresponding to isoquant Q_0, it is clear that the output of apples is maximized by finding the apple isoquant that is tangent to isoquant Q_0. Given that we produce Q_0 of oranges, producing Q_1 of apples (at, say, point *C*) means that some resources are unused. Producing along Q_0, but not at *E* (at, say, point *D*), means that all resources are used, but not efficiently; we can produce the same number of oranges and more apples at *E*. The economy cannot produce more than Q_1 of apples and still produce Q_0 of oranges; producing Q_2 of apples would require producing less than Q_0 of oranges. Only at point *E* are all resources used efficiently and Q_0 of oranges produced. At the point of tangency, the slopes of the isoquants are the same; that is, the marginal rate of substitution of land for labor is the same in the production of apples as it is in the production of oranges.

PRODUCT MIX EFFICIENCY

To choose the best mix of apples or oranges to produce, we need to consider both technical feasibility *and* individuals' preferences. For each level of output of apples, we can determine from the technology the maximum feasible level of output of oranges. This generates the production possibilities schedule. Given the production possibilities schedule, we wish to get to the highest possible level of utility. For simplicity, we assume all individuals have identical tastes. In Figure 3.9, we depict both the production possibilities schedule and the indifference curves between apples and oranges. Utility is maximized at the point of tangency of the indifference curve to the production possibilities schedule. The slope of the production possibilities schedule is called the **marginal rate of transformation**; this tells us how many extra apples we can have if we reduce production of oranges by one. At the point of tangency, E, the slopes of the indifference curve and the production possibilities schedule are the same, that is, the marginal rate of substitution of apples for oranges is equal to the marginal rate of transformation.

Under competition, the marginal rate of transformation will be equal to the relative price of apples to oranges. If, by reducing production of apples by one, firms can increase the production of oranges by, say, one, and sell the oranges for more than the price of apples, profit-maximizing firms will clearly expand production of oranges. We have shown why, under competition, consumers' marginal rates of substitution will equal the price ratio. Because both the marginal rates of substitution and the marginal rate of transformation will equal the price ratio, the marginal rate of transformation must equal consumers' marginal rates of substitution. Hence, under ideal competitive markets, all three conditions required for Pareto efficiency are satisfied.

BASIC CONDITIONS FOR PARETO EFFICIENCY

1. *Exchange efficiency:* The marginal rate of substitution between any two goods must be the same for all individuals.
2. *Production efficiency:* The marginal rate of technical substitution between any two inputs must be the same for all firms.
3. *Product mix efficiency:* The marginal rate of transformation must equal marginal rate of substitution.

Competitive economies satisfy all three conditions.
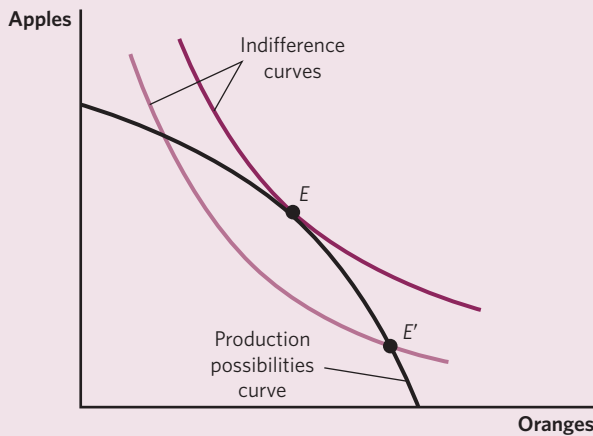
PRODUCT MIX EFFICIENCY REQUIRES THAT THE MARGINAL RATE OF TRANSFORMATION EQUAL CONSUMERS' MARGINAL RATE OF SUBSTITUTION
To reach the highest level of consumers' utility, the indifference curve and the production possibilities schedule must be tangent (point *E*). At any other point, such as *E'*, consumer utility is lower than at *E*.

REVIEW AND PRACTICE

SUMMARY

1. Resource allocations that have the property that no one can be made better off without someone else being made worse off are called Pareto efficient allocations.

2. The Pareto principle is based on individualistic values. Whenever a change can make some individuals better off without making others worse off, it should be adopted. Most public policy choices, however, involve trade-offs, under which some individuals are better off and others are worse off.

3. The principle of consumer sovereignty holds that individuals are the best judges of their own needs and pleasures.

4. Pareto efficiency requires exchange efficiency, production efficiency, and product mix efficiency.

5. The fundamental theorems of welfare economics provide conditions under which a competitive economy is Pareto efficient, and under which every Pareto efficient allocation can be obtained through markets, provided that the appropriate redistribution of initial endowments (incomes) occurs.

6. Exchange efficiency means that, given the set of goods available in the economy, no one can be made better off without someone else being made worse off; it requires that all individuals have the same marginal rate of substitution between any pair of commodities. Competitive markets in which individuals face the same prices always have exchange efficiency.

7. Production efficiency requires that, given the set of resources, the economy not be able to produce more of one commodity without reducing the output of some other commodity; the economy must be operating along its production possibilities curve. Production efficiency requires that all firms have the same marginal rate of technical substitution between any pair of inputs; competitive markets in which firms face the same prices always have production efficiency.

8. Product mix efficiency requires that the marginal rate of transformation—the slope of the production possibilities curve—equal individuals' marginal rate of substitution. Competitive markets have product mix efficiency.

KEY CONCEPTS

Centralized allocation mechanism

Consumer sovereignty

Exchange efficiency

Fundamental theorems of welfare economics

Indifference curves

Invisible hand

Isocost line

Isoquants

Marginal (additional) benefit

Marginal (additional) cost

Marginal rate of substitution

Marginal rate of technical substitution

Marginal rate of transformation

Pareto efficiency

Pareto improvement

Pareto principle

Production efficiency

Product mix efficiency

Trades

Utility possibilities curve

QUESTIONS AND PROBLEMS

1. Explain why an economy in which airlines charge different passengers different prices for the same flight will not have exchange efficiency.

2. Doctors often charge patients different amounts depending on their judgment concerning the patients' ability to pay. What implications does this have for exchange efficiency?

3. Can you think of other common practices and policies that might interfere with exchange efficiency?

4. Explain why a tax that is levied only on the use of capital by corporations will interfere with the production efficiency of the economy. (Compare the marginal rates of technical substitution between corporations and unincorporated enterprises.)

5. Advocates of small businesses often argue that they should receive special tax treatment. Assume that small businesses had to pay only half the Social Security tax that is imposed on large corporations. What effect would that have on production efficiency?

6. Consider an economy that produces two goods, cars and shirts. Explain why if a tax is imposed on the consumption of cars but not on shirts, the economy will not exhibit product mix efficiency.

7. An individual is indifferent among the combinations of public and private goods shown in the following table.

COMBINATION	PUBLIC GOODS	PRIVATE GOODS
A	1	16
B	2	11
C	3	7
D	4	4
E	5	3
F	6	2

Draw the individual's indifference curve. Assuming that the economy can produce one unit of public goods and ten units of private goods, but that it can produce one more unit of public goods by reducing its production of private goods by two units, draw the production possibilities schedule. What is the maximum production of private goods? The maximum production of public goods? Can it produce five units of public goods and one unit of private goods? Which of the feasible combinations maximizes utility?

4

MARKET FAILURE

The last chapter explained why markets play such a central role in our economy: under ideal conditions, they ensure that the economy is Pareto efficient. However, there is often dissatisfaction with markets. Some of the dissatisfaction is of the "grass is always greener on the other side" variety: people like to think that an alternative way of organizing the economy might make them better off. Some of the dissatisfaction is real, though: markets often seem to produce too much of some things, such as air and water pollution, and too little of others, such as support for the arts, research into the nature of matter, or the causes of cancer. Furthermore, markets can lead to situations in which some people have too little income to live on. Over the past fifty years, economists have devoted enormous efforts to understanding the circumstances under which markets yield efficient outcomes, and the circumstances in which they fail to do so.

This chapter looks both at these market failures and at the reasons why governments intervene in markets even when they are efficient.

PROPERTY RIGHTS AND CONTRACT ENFORCEMENT

Chapter 3 explained why markets result in Pareto efficient outcomes. For markets to work, however, there needs to be a government to define property rights and enforce contracts. In some societies, land is held in common: anyone can graze cattle and sheep on it. Because no one has the property right to the land, though, no one has an incentive to ensure

PROPERTY RIGHTS AND MARKET FAILURES: THE TRAGEDY OF THE COMMONS REVISITED

The **tragedy of the commons** is a term that encapsulates how difficult it is to achieve the appropriate level of property rights. Oft-cited examples of this are the common pasture that is ruined by overgrazing or the common lake that is depleted by overfishing. Although the action of each individual is a rational attempt to promote one's own short-term self-interest, the group's collective actions are not in the community's long-term best interests. This dilemma raises the perplexing and hotly debated policy question: How does one manage a resource that formally does not belong to anyone?

For a long time, the common response was either conversion of common resources to private property or the external regulation of these common resources. If a resource is converted to private property, the owner should have both the profit incentives and the protection of property rights to manage the entire resource responsibly. If the resource is regulated by the government, rules could be imposed on individual users for the common good.

However, in 1999, Nobel Prize-winning economist Elinor Ostrom revisited these solutions because they sometimes left the commons worse off. Private control over previously public resources could create the usual problems resulting from monopoly business practices, whereas external oversight could generate inappropriate and poorly implemented regulations. She noted a third response to the tragedy of the commons, one that she documented empirically by drawing on many cases from around the world: utilization of community social capital to devise creative and effective local solutions.

Another twist on this debate is referred to as the **tragedy of the anticommons**, a term coined by Michael Heller of Columbia Law School. In contrast to the problems created by ambiguous ownership of a community resource, this refers to the opposite: excessive private ownership of a community resource that prevents achievement of a desirable outcome for society. The tragedy of the anticommons is now commonly applied to excessive property rights in areas such as biomedical research, for instance, in the patenting of genes (see Chapter 12).

that overgrazing does not occur. In the former communist countries, property rights were not well defined, so people had insufficient incentive to maintain or improve their apartments. In market economies, though, the benefits of such improvements are reflected in the market price of the property.

Similarly, if individuals are to engage in transactions with each other, the contracts they sign must be enforced. Consider a typical loan, in which one person borrows money from another and signs a contract to repay it. Unless such contracts are enforced, no one would be willing to make a loan.

At an even more primitive level, unless there is protection of private property, people will have insufficient incentive to save and invest, as their savings might be taken away.

Government activities aimed at protecting citizens and property, enforcing contracts, and defining property rights can be thought of as providing the foundations on which all market economies rest.

MARKET FAILURES AND THE ROLE OF GOVERNMENT

The first fundamental theorem of welfare economics asserts that the economy is Pareto efficient only under certain circumstances or conditions. Markets are not Pareto efficient under six important conditions, referred to as **market failures**, which provide a rationale for government activity.

1. FAILURE OF COMPETITION

For markets to result in Pareto efficiency, there must be **perfect competition**; that is, there must be a sufficiently large number of firms that each believes it has no effect on prices. However, in some industries—supercomputers, operating systems and chips for PCs, aluminum, cigarettes, and greeting cards, for instance—there are relatively few firms, or one or two firms have a large share of the market. When a single firm supplies the market, economists refer to it as a **monopoly**; when a few firms supply the market, economists refer to them as an **oligopoly**. Even when there are many firms, each may produce a slightly different good, and may thus perceive itself facing a downward-sloping demand curve.

Economists refer to such situations as **monopolistic competition**. In all these situations, competition deviates from the ideal of perfect competition, in which each firm is so small that it believes there is nothing it can do to affect prices.

It is important to recognize that under these circumstances, firms may still seem to be competing actively against each other, and that the market economy may seem to "work" in the sense that goods are being produced that consumers seem to like. The first fundamental theorem of welfare economics—the result that market economies are Pareto efficient—requires more than just that there be *some* competition. As we saw in the last chapter, Pareto efficiency entails stringent conditions, such as exchange, production, and product mix efficiency; these conditions typically are satisfied only if each firm and household believes that it has *no* effect on prices.

There are a variety of reasons why competition may be limited. When average costs of production decline as a firm produces more,[1] a larger firm will have a competitive advantage over a smaller firm. There may even be a **natural monopoly**, a situation in which it is cheaper for a single firm to produce the entire output than for each of several firms to produce part of it. Even when there is not a natural monopoly, it may be efficient for only a few firms to operate. High transportation costs mean that goods sold by a firm at one location are not perfect substitutes for goods sold at another location. Imperfect information may also mean that if a firm raises its price it will not lose all its customers; it only faces a downward-sloping demand curve.

Firms may also engage in strategic behavior to discourage competition. They may threaten to cut prices if potential rivals enter; such threats may both be credible and serve to discourage entry.

Finally, some imperfections of competition arise out of government actions. Governments grant patents—exclusive rights to an invention—to innovators. Although patents are important in providing incentives to innovate, they make competition in the product market less than perfect. Of course, even without patents, the fact that an innovator has some information (knowledge) that is not freely available to others may enable it to establish a dominant market position.

It is easy to see why imperfect competition leads to economic inefficiency. We saw earlier that under competition, firms set output at the Pareto efficient level. They set price equal to the marginal cost of production. Price can be thought of as measuring the marginal benefit of consuming an extra unit of the good. Thus, with competition, marginal benefits equal marginal costs. Under imperfect competition, firms set

---

[1] Declining average costs correspond to increasing returns: doubling inputs more than doubles output.
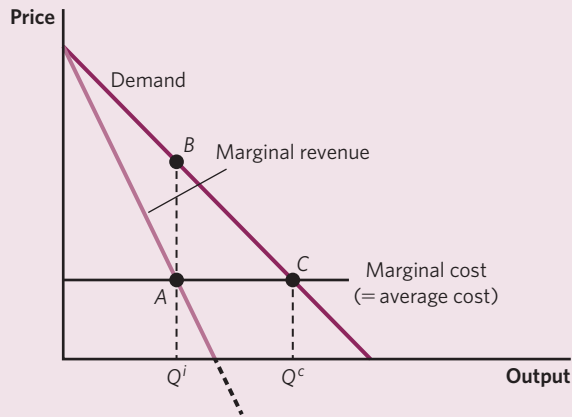
MONOPOLY PRICING
Monopoly output is lower than competitive output, or the output at which profits are zero. There is a resulting welfare loss.

the extra revenue they obtain from selling one unit more—the **marginal revenue**—equal to the marginal cost. With a downward-sloping demand curve, the marginal revenue has two components. When a firm sells an extra unit, it receives the price of the unit, but to sell the extra unit, it must lower the price it charges on that and all previous units—the demand curve is downward sloping. The revenue gained from selling the extra unit is its price minus the revenue forgone because the expansion in sales lowers the price on all units. Thus, marginal revenue is less than price. Figure 4.1 shows the demand curve facing a firm and the marginal revenue, which lies below the demand curve. Competitive equilibrium occurs at Q^c, whereas the imperfect competition equilibrium occurs at Q^i, a much lower level of output. This reduction in output is the inefficiency associated with imperfect competition.

Of course, if there is a natural monopoly, with declining average costs, and with marginal costs below average costs,[2] competition is not viable; if a firm charged a price equal to the marginal cost (as would be the case under competition), it would operate at a loss, as the marginal cost is lower than the average cost. Even then, however, a private monopoly would typically charge more than a government-run monopoly; the private monopoly would seek to maximize profits, whereas the government-run monopoly, which did not receive any subsidy, would only seek to break even.

---

[2] When average costs are declining, marginal costs always lie below average costs; it is the low value of the marginal cost—the cost of producing the last unit—that brings down the average costs.

2. PUBLIC GOODS

Some goods either will not be supplied by the market or, if supplied, will be supplied in insufficient quantity. An example on a large scale is national defense; on a small scale, navigational aids (such as buoys). These are called **pure public goods**. They have two critical properties. First, it costs nothing for an additional individual to enjoy their benefits: formally, there is zero marginal cost for the additional individual enjoying the good. It costs no more to defend a country of one million and one individuals than to defend a country of one million. The costs of a lighthouse do not depend at all on the number of ships that sail past it. Second, it is, in general, difficult or impossible to exclude individuals from the enjoyment of a pure public good. If I put a lighthouse in a rocky channel to enable my ships to navigate safely, it is difficult or impossible for me to exclude other ships entering the channel from its navigational benefits. If our national defense policy is successful in diverting an attack from abroad, everyone benefits; there is no way to exclude any single individual from these benefits.

The market either will not supply, or will not supply enough of, a pure public good. Consider the case of the lighthouse. A large shipowner with many ships might decide that the benefits it receives from a lighthouse exceed the costs; but in calculating how many lighthouses to put in place, it will look only at the benefits it receives, not at the benefits received by others. Thus, there will be some lighthouses for which the total benefits (taking into account *all* of the ships that make use of the lighthouse) exceed the costs but for which the benefits of any single shipowner are less than the costs. Such lighthouses will not be put into place, and that is inefficient. The fact that private markets will not supply, or will supply too little of, public goods provides a rationale for many government activities. Public goods are discussed in detail in the next chapter.

3. EXTERNALITIES

In many cases, the actions of one individual or one firm affect other individuals or firms, such as when one firm imposes a cost on other firms but does not compensate them, or, alternatively, when one firm confers a benefit on other firms but does not reap a reward for providing it. Air and water pollution are examples. When I drive a car that is not equipped with a pollution control device, I lower the quality of the air, and thus impose a cost on others. Similarly, a chemical plant that discharges its chemicals into a nearby stream imposes costs on downstream users of the

water, who may have to spend a considerable amount of money to clean up the water to make it usable.

Instances in which one individual's actions impose a cost on others are referred to as **negative externalities**. Not all externalities are negative, though. There are some important instances of **positive externalities**, in which one individual's actions confer a benefit on others. If I plant a beautiful flower garden in front of my house, my neighbors may benefit from being able to look at it. An apple orchard may confer a positive externality on a neighboring beekeeper. An individual who rehabilitates his or her house in a neighborhood that is in decline may confer a positive externality on the neighbors.

There are a large number of other examples of externalities. An additional car on a crowded highway will add to road congestion, both reducing the speed at which other drivers can travel safely and increasing the probability of an accident. An additional fisherman fishing in a given pond may reduce the amount of fish that others will be able to catch. If several oil wells are drilled in the same oil pool, taking more oil from one of the wells may reduce the amount of oil extracted by the other wells.

The crisis of 2008 made it clear that the financial sector could (and did) impose large externalities on the rest of the economy. In fact, many described what they did using language similar to that associated with environomental externalities: the financial sector was accused of polluting the global economy with toxic mortgages.

Whenever such externalities exist, the resource allocation provided by the market will not be efficient. Because individuals do not bear the full cost of the negative externalities they generate, they will engage in an excessive amount of such activities; conversely, because individuals do not enjoy the full benefits of activities generating positive externalities, they will engage in too little of these. Thus, for example, without government intervention of some kind, the level of pollution would be too high.

Externalities and environmental policy are discussed in detail in Chapter 6.

4. INCOMPLETE MARKETS

Pure public goods and services are not the only goods and services that private markets fail to provide adequately. Whenever private markets fail to provide a good or service even though the cost of providing it is less than what individuals are willing to pay, there is a market failure that we refer to as **incomplete markets** (because a complete market would provide all goods and services for which the cost of provision is less than

what individuals are willing to pay). Some economists believe that private markets have done a particularly poor job of providing insurance and loans, and that this provides a rationale for government activities in these areas.

INSURANCE AND CAPITAL MARKETS   The private market does not provide insurance for many important risks that individuals face, although insurance markets are much better today than they were ninety years ago. The government has undertaken a number of insurance programs, motivated at least in part by this market failure. In 1933, following the bank failures of the Great Depression, the government set up the Federal Deposit Insurance Corporation. Banks pay the corporation annual premiums, which provide insurance for depositors against a loss of savings arising from the insolvency of banks. The government has also been active in providing flood insurance. Following urban riots in the summer of 1967, most private insurance companies refused to write fire insurance in certain inner-city areas, and again the government stepped in.

Similarly, government has provided farmers with crop insurance, partly because of the failure of markets to do so; it provides unemployment insurance; and until Medicare, the government health insurance program for the aged, was introduced in the 1960s, many of the elderly found it difficult to procure health insurance in the market. To protect investors against the effects of inflation, the government has been issuing Treasury inflation-protected securities (TIPS) since 1997.

In recent decades, the government has taken an active role not only in remedying deficiencies in risk markets but also in ameliorating the effects of imperfect capital markets. In 1965 the government passed legislation providing for government guarantees on student loans, making it less difficult for individuals to obtain loans to finance their college education. Today, after it was discovered that private lenders were making massive profits out of student loans and charging excessive interest rates, the government has become a major provider of student loans. This is only one of several government loan programs: the government provides loans to businesses engaged in international trade through the Export–Import Bank, it provides loans for small business through the Small Business Administration, and so forth. In each of these credit markets, there were allegations that access to credit was restricted prior to the introduction of the government program.

The question of why capital and insurance markets are imperfect has been the subject of extensive research during the past four decades. At least three different answers have been put forward; each may have some validity. One focuses on *innovation*: we are used to new products

STUDENT LOANS: INCOMPLETE REFORM OF AN INCOMPLETE MARKET

Government guarantees opened the door to student loans, but the student loan market has not always worked well. Many of the problems arise out of the undue influence of the financial sector and the private education sector in shaping the student loan program. For a long while, although the government guaranteed the loans, the private sector charged interest rates *as if* they were risky, imposing enormous costs on students. By replacing government-guaranteed loans by government loans, students and the government were able to save an estimated $80 billion over ten years. Many for-profit schools entice poorly informed students; they fail to provide them with the skills they need to get a job. However, these schools have successfully repelled attempts to regulate them effectively, or even to deny them access to government loan programs, without which they would not survive. Making matters worse, several pieces of legislation (most recently, in the so-called reform of bankruptcy) made it extraordinarily difficult to discharge student debt, to get a fresh start, *even if the school did not deliver the education and jobs promised.*

There are increasing worries too that the private sector will attempt to "skim the cream," offering loans to the best risks, leaving the government with the highest risks.

With student debt now exceeding a trillion dollars—exceeding even credit card debt—and many young Americans saddled with seemingly crushing debts, there is a growing sense that there must be better ways of financing these important investments.

Australia has explored one such approach, in which the amount students repay depends on their income. *Income-contingent loan programs* are now being experimented with by several countries around the world (see Chapter 14).

constantly coming onto the market; but there are also innovations in how the economy functions—innovations in creating new markets, including inventing new securities and new insurance policies. Indeed, those working in the insurance and securities industries refer to these advances as new products. However, there is often an undersupply of innovations, which is why there is an important role for government in research.

The introduction of many of these new products is related to the second explanation: *transactions costs*. It is costly to run markets, to enforce contracts, and to introduce new insurance policies. An insurance firm may be reluctant to go to the trouble of designing a new insurance policy if it is unsure whether anyone will buy the policy, or if even if the product is successful, whether it will be able to reap the rewards as other competitors come into the market.

The third set of explanations centers around *asymmetries of information* and *enforcement costs*. The insurance company is often less informed

about the nature of some risks than the person purchasing insurance. When the two parties to a transaction have different information of this kind, we say that there is an information asymmetry. Thus, a firm might well wish to buy insurance against the risk that the demand for its product will decline. The insurer, on the other hand, may well reason: I want to estimate the risk, and charge a premium based on that estimate. If I overestimate the risk, the premium will be too high, and the firm will refuse to buy my policy; whereas if I underestimate the risk, the premium will be too low, in which case, the firm will buy my policy, but on average, I will lose money. I am in a heads-you-win-tails-I-lose situation. When information asymmetries like this are large, markets will not exist.

Similarly, in capital markets, lenders worry about getting repaid. They may not be able to tell which borrowers are likely to repay. This is particularly a problem with loans, such as student loans, for which there is no collateral. (In the case of a loan on a house, if the borrower defaults, at least the lender can sell the house and recoup most or all of what it has put out.) The bank finds itself in a dilemma: if it increases the interest rate to reflect the fact that many loans are not repaid, it may find that the default rate actually increases; those who know that they are going to repay refuse to borrow, while those who are not planning to repay care very little about the amount the lender is nominally charging, as in all likelihood they will not pay that amount anyway. The phenomenon is called **adverse selection**. It may turn out that there is no interest rate that the bank can charge for, say, student loans (without a government subsidy) at which it can reap an *expected return* commensurate with what it can obtain on other investments.

This basic principle—that when there are asymmetries of information and enforcement problems markets may not exist—has been shown to provide part of the explanation of many missing markets.[3] We shall examine these problems in greater depth in the context of health insurance, in Chapter 13.

The reasons why markets do not exist may have implications for how governments might go about remedying the market failure. Government, too, faces transactions costs, enforcement problems, and asymmetries of information, although in many instances they are different from those faced by the private sector. Thus, in designing loan programs or interventions in capital markets, governments need to bear in mind that they too are often less informed than the borrower.

---

[3] The literature in this area is extensive. The basic articles are George Akerlof, "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* 84 (1970): 488–500; and Michael Rothschild and Joseph Stiglitz, "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics* 90 (1976): 629–650.

COMPLEMENTARY MARKETS Finally, we turn to the problems associated with the absence of certain complementary markets. Suppose all individuals enjoy only coffee with sugar. Assume, moreover, that without coffee there is no market for sugar. Given that sugar was not produced, an entrepreneur considering whether to produce coffee would not do so, because he would realize that he would have no sales. Likewise, given that coffee was not produced, an entrepreneur considering whether to produce sugar also would not do so, since she too would realize that she would have no sales. If, however, the two entrepreneurs could get together, there would be a good market for coffee and sugar. Each *acting alone* would not be able to pursue the public interest, but acting together they could.

This particular example is deliberately quite simple, and in this case coordination (between the potential sugar producer and the potential coffee producer) might easily be provided by the individuals themselves without government intervention. In many cases, however, particularly in less developed countries, large-scale coordination is required; this may require government planning. Similar arguments have been put forward as justification for public urban renewal programs. Redeveloping a large section of a city requires extensive coordination among factories, retailers, landlords, and other businesses. One of the objectives of government development agencies is to provide that coordination (if markets were complete, the prices provided by the market would perform this "coordination" function).

5. INFORMATION FAILURES

A number of government activities are motivated by imperfect information on the part of consumers, and by the belief that the market, by itself, will supply too little information. For instance, the Truth in Lending Act requires lenders to inform borrowers of the true rate of interest on their loans. The Federal Trade Commission and the Food and Drug Administration have both adopted a number of regulations concerning labeling, disclosure of contents, and the like. At one time, the Federal Trade Commission proposed that used-car dealers be required to disclose whether they had tested various parts of the car, and, if so, what the outcomes of the tests were. These regulations generated a considerable amount of controversy, and under pressure from Congress, the FTC was forced to back down.

Opponents of regulations on information disclosure contend that they are unnecessary (the competitive market provides incentives for firms to

disclose relevant information), irrelevant (consumers pay little attention to the information the law requires firms to disclose), and costly, both to government that must administer them and to the firms that must comply with the regulations. Proponents of these regulations claim that, though they are sometimes difficult to administer effectively, they are still critical to the affected markets.

The government's role in remedying information failures goes beyond these simple consumer and investor protections, however. Information is, in many respects, a public good. Giving information to one more individual does not reduce the amount others have. Efficiency requires that information be freely disseminated or, more accurately, that the only charge be for the actual cost of transmitting the information. The private market will often provide an inadequate supply of information, just as it supplies an inadequate amount of other public goods. The most notable example of government activity in this area is the National Weather Service. Another example is the information provided to ships by the U.S. Coast Guard.

Various other market failures are associated with imperfect information. One of the assumptions that went into the proof of the fundamental theorems of welfare economics was that there was perfect information, or, more precisely, that nothing firms or households did had any effect on beliefs or information. In fact, much economic activity is directed at obtaining information, from employers trying to find out who are good employees, to lenders trying to find out who are good borrowers, investors trying to find out what are good investments, and insurers trying to find out who are good risks. Later, we shall see that information problems lie behind several government programs. For instance, many of the problems in the health sector in general and health insurance markets in particular can be traced to problems of information.

Resources devoted to producing new knowledge—**research and development (R&D)** expenditures—can be thought of as a particularly important category of expenditures on information. Again, the fundamental theorems of welfare economics, which form the basis of our belief in the efficiency of market economies, simply assume that there is a given state of information about technology, begging the question of how the economy allocates resources to research and development. Chapter 12 will explain why the market, on its own, may engage in an insufficient amount of at least certain types of R&D.[4]

---

[4] For an extended discussion of the market failures associated with incomplete markets and imperfect information, see B. Greenwald and J. E. Stiglitz, "Externalities in Economies with Imperfect Information and Incomplete Markets," *Quarterly Journal of Economics* 105 (May 1986): 229–264.

6. UNEMPLOYMENT, INFLATION, AND DISEQUILIBRIUM

Perhaps the most widely recognized symptoms of market failure are the periodic episodes of high unemployment, both of workers and machines, that have plagued capitalist economies during the past two centuries. Even though these recessions and depressions were greatly moderated in the period between World War II and 2008, perhaps partly because of government policies, the unemployment rate still climbed over 10 percent in 1982 and reached that level again in 2009. Although this is low compared to the rate during the Great Depression, when unemployment reached 24 percent in the United States, high unemployment rates during the current Great Recession persisted despite government interventions to kickstart the economy. The national unemployment rate also masks significantly higher rates in especially hard-hit parts of the country and among highly vulnerable populations. At one time, more than one of six Americans who wanted a full-time job could not get one. The global economic crisis has hit several countries in Europe particularly hard; by 2013, average unemployment had reached record levels, and several countries were in a depression, with average unemployment in excess of 25 percent and youth unemployment in excess of 50 percent.

Most economists take the high levels of unemployment as prima facie evidence that *something* is not working well in the market. To some economists, high unemployment is the most dramatic and most convincing evidence of market failure.

The issues raised by unemployment and inflation are sufficiently important, and sufficiently complicated, that they warrant a separate course in macroeconomics. Some aspects of these issues are touched on in Chapter 28, which is concerned with the consequences of government deficits and attempts to survey some of the important ways that these macroeconomic considerations affect the design of tax policy.

SIX BASIC MARKET FAILURES

1. Imperfect competition
2. Public goods
3. Externalities
4. Incomplete markets
5. Imperfect information
6. Unemployment and other macroeconomic disturbances

INTERRELATIONSHIPS OF MARKET FAILURES

The market failures we have discussed are not mutually exclusive. Information problems often provide part of the explanation of

missing markets. In turn, externalities are often thought to arise from missing markets: if fishermen could be charged for using fishing grounds—if there were a market for fishing rights—there would not be overfishing. Public goods are sometimes viewed as an extreme case of externalities, where others benefit from my production of the good as much as I do. Much of the recent research on unemployment has attempted to relate it to one of the other market failures.

MARKET FAILURES: EXPLANATIONS OR EXCUSES?

The agricultural price support program provides an illustration of an instance in which the appeal to market failures is more of an excuse for a program than a rationale. There are important market failures in agriculture. Prices and output are highly variable. Farmers typically cannot buy insurance to protect them against either price or output fluctuations. Even though they could reduce their exposure to price risk somewhat by trading in futures and forward markets, these markets are highly speculative, and farmers worry that they are at a marked disadvantage in trading in them. For example, there are five very large traders in wheat that have access to more information; as a result, farmers view trading on futures markets with these informed traders as playing on an unlevel playing field.

What farmers really care about, of course, is not price variability, but income variability. Programs to stabilize prices do not fully stabilize income, as income depends both on the price received and the quantity produced. Indeed, in some cases, stabilizing prices may actually increase the variability of income. Normally, prices rise when, on average, quantities fall. If prices rise proportionately, then income may vary very little, with price increases just offsetting quantity decreases. In such a situation, stabilizing prices will increase income variability.

Price support programs are also justified as helping poor farmers—reflecting the failure of markets to provide an appropriate distribution of income. Critics ask, though, why are poor farmers particularly deserving of aid, rather than poor people in general? Moreover, the price support programs give aid on the basis of how much a farmer produces. Thus, large farmers gain far more than small farmers do.

If the objective of the farm programs were to address these market failures, then the farm program would be designed in a markedly different manner. In fact, a major objective of the farm program is to transfer resources—to subsidize farmers (and not just poor farmers)—not to correct a market failure. The program is designed to keep a large part of its cost hidden: only a part of the cost is reflected in the federal budget; the rest is paid for by consumers in the form of higher prices. The market failure approach has provided some of the rhetoric for the program, but not the rationale. For that, we have to look into politics and the role of special interest groups.

REDISTRIBUTION AND MERIT GOODS

The sources of market failure discussed thus far result in economic inefficiency in the absence of government intervention. Even if the economy were Pareto efficient, though, there are two further arguments for government intervention. The first is income distribution: the fact that the economy is Pareto efficient says nothing about the distribution of income; competitive markets may give rise to a very unequal distribution, which may leave some individuals with insufficient resources on which to live. One of the most important activities of the government is to redistribute income. This is the express purpose of welfare activities, such as food stamps and Medicaid. How we think systematically about issues of distribution is the subject of Chapter 7.

The second argument for government intervention in a Pareto efficient economy arises from concern that individuals may not act in their own best interests. It is often argued that an individual's perception of his or her own welfare may be an unreliable criterion for making welfare judgments. Even fully informed consumers may make "bad" decisions. Individuals continue to smoke, for instance, even though it is bad for them, and even though they know it is bad for them. Individuals fail to wear seat belts, even though wearing seat belts increases the chances of survival in an accident, and even though individuals know the benefits of seat belts. The same holds true for motorcycle helmets. There are those who believe that the government should intervene in such cases, in which individuals seemingly do not do what is in their own best interest; the kind of intervention that is required must be stronger than simply providing information. Goods that the government compels individuals to consume, such as seat belts and elementary education, are called **merit goods**.

The view that the government should intervene because it knows what is in the best interest of individuals better than they do themselves is referred to as **paternalism**. The paternalistic argument for government activities is quite distinct from the externalities argument discussed earlier. One might argue that smoking causes cancer, and that because individuals who get cancer may be treated in public hospitals or financed by public funds, smokers impose a cost on nonsmokers. This, however, can be dealt with by making smokers pay their full costs—for instance, by imposing a tax on cigarettes. Alternatively, smoking in a crowded room does indeed impose a cost on nonsmokers in that room. But this, too, can be dealt with directly. Those who take a paternalistic view might argue that individuals should not be allowed to smoke, even in the privacy of their

own homes, and even if a tax that makes the smokers take account of the external costs imposed on others is levied. Although few have taken such an extreme paternalistic position with respect to smoking, this paternalistic role undoubtedly has been important in a number of areas, such as government policies toward drugs (illegalization of marijuana) and liquor (prohibition in the 1930s).

In contrast to the paternalistic view, many economists and social philosophers believe that the government should respect consumers' preferences. Though there may occasionally be cases that merit a paternalistic role for the government, these economists argue that it is virtually impossible to distinguish such cases from those that do not. And they worry that once the government assumes a paternalistic role, special interest groups will attempt to use government to further their own views about how individuals should act or what they should consume. The view that government should not interfere with the choices of individuals is sometimes referred to as **libertarianism**.

There are two important caveats to economists' general presumption against government paternalism. The first concerns children. Someone— either the parents or the state—must make paternalistic decisions on behalf of children, and there is an ongoing debate concerning the proper division of responsibility between the two. Some treat children as if they were the property of their parents, arguing that parents alone should have responsibility for taking care of their children. Most argue, however, that the state has certain basic responsibilities, such as, for instance, ensuring that every child gets an education and that parents do not deprive their children of needed medical care or endanger them physically or emotionally.

The second caveat concerns situations in which the government cannot, at least without difficulty, commit itself to refrain from helping individuals who make poor decisions. For instance, individuals who do not save for their retirement become a burden on the government; this provides part of the rationale for Social Security. In other instances, individuals who fail to take appropriate precautions become a burden to society, and a sense of compassion makes it difficult, in the face of a crisis, to simply say, "You should have taken appropriate precautions." Government accordingly responds by *forcing*, or at least *encouraging*, precautionary behavior. Individuals, for example, who neither buy earthquake insurance nor build homes that can withstand the effects of an earthquake become a burden on the government when an earthquake strikes. The government finds itself compelled to act compassionately, even if the victims' dire situation is partly of their own making. Recognizing this, the government may compel individuals to take adequate precautions against

the event of an earthquake by, for instance, enforcing high standards for earthquake-resistant construction and making the purchase of earthquake insurance mandatory.

TWO PERSPECTIVES ON THE ROLE OF GOVERNMENT

We saw in Chapter 1 that there are two aspects of the analysis of public sector activities: the normative approach, which focuses on what the government should do, and the positive approach, which focuses on describing and explaining both what the government actually does and what its consequences are. We can now relate our discussion of market failures, redistribution, and merit goods to these two alternative approaches.

NORMATIVE ANALYSIS

The fundamental theorems of welfare economics are useful because they clearly delineate a role for the government. In the absence of market failures and merit goods, all the government needs to do is worry about the distribution of income (resources). The private enterprise system ensures that resources will be used efficiently.

If there are important market failures—imperfect competition, imperfect information, incomplete markets, externalities, public goods, and unemployment—there is a presumption that the market will not be Pareto efficient. This suggests a role for the government, but there are two important qualifications.

First, it must be shown that there is, at least in principle, some way of intervening in the market to make someone better off without making anyone worse off; that is, of making a *Pareto improvement*. Second, it must be shown that the actual political processes and bureaucratic structures of a democratic society are capable of correcting the market failure and achieving a Pareto improvement.

When information is imperfect and costly, the analysis of whether the market is Pareto efficient must take into account these information costs; information is costly to the government, just as it is to private firms. Markets may be incomplete because of transactions costs; the government, too, would face costs in establishing and running a public

insurance program. These costs must be considered in the decision to set up such a program.

Recent research has established a variety of circumstances under which, although the government has no advantage in information or transactions costs over the private market, the government could, in principle, bring about a Pareto improvement. The fact that there may exist government policies that would be Pareto improvements does not, however, necessarily create a presumption that government intervention is desirable. We also have to consider the consequences of government intervention, in the form it is likely to take, given the nature of our political process. We have to understand how real governments function if we are to assess whether government action is likely to remedy market failures.

In the 1960s, it was common to take a market failure, show that a government program could lead to a Pareto improvement (someone could be made better off without making anyone worse off), and conclude that government intervention was called for. When programs were enacted and failed to achieve what they were supposed to, the blame was placed on petty bureaucrats or political tampering. But, as we shall see in Chapters 8 and 9, even if bureaucrats and politicians behave honorably, the nature of government itself still may help explain government's failures.

Public programs—even those allegedly directed at alleviating some market failure—are instituted in democracies not by ideal governments or benevolent despots, but by complicated political processes.

POSITIVE ANALYSIS

The market failure approach to understanding the role of the government is largely a normative approach. The market failure approach provides a basis for identifying situations in which the government *ought* to do something, tempered by considerations of government failure.

The popularity of the market failure approach has caused many programs to be justified in terms of market failures. However, this may simply be rhetoric. There is often a significant difference between a program's stated objective (to remedy some market failure) and its design. Political rhetoric may focus on the failure of markets to provide insurance against volatile prices and the consequences that this has for small farmers, but government agricultural programs may, in practice, transfer income to large farmers. Insight into the political forces at work and the true objectives of the programs may be gained more easily by looking at how the programs are designed and implemented than by looking at the stated objectives of the legislation.

Some economists believe that economists should focus their attention on positive analysis, on describing the consequences of government programs and the nature of the political processes, rather than on normative analysis, what the government should do. However, discussions by economists (and others) of the role that government *should* play constitute an important part of the political process in modern democracies. Beyond that, an analysis of institutional arrangements by which public decisions get made may lead to designs that enhance the likelihood that the public decisions will reflect a broader set of public interests, not just special interests. These matters will be taken up in further detail in later chapters.

REVIEW AND PRACTICE

SUMMARY

1. Under certain conditions, the competitive market results in a Pareto efficient resource allocation. When the conditions required for this are not satisfied, a rationale for government intervention in the market is provided.

2. Government is required to establish and enforce property rights and enforce contracts. Without this, markets by themselves cannot function.

3. There are six reasons why the market mechanism may not result in a Pareto efficient resource allocation: failure of competition, public goods, externalities, incomplete markets, information failures, and unemployment. These are known as market failures.

4. Even if the market is Pareto efficient, there may be two further grounds for government action. First, the competitive market may give rise to a socially undesirable distribution of income. And second, some believe that individuals, even when well informed, do not make good judgments concerning the goods they consume, thus providing a rationale for regulations restricting the consumption of some goods, and for the public provision of other goods, called merit goods.

5. Even though the presence of market failures implies that there may be scope for government activity, it does not imply that a particular government program aimed at correcting the market failure is necessarily desirable. To evaluate government programs, one must take into account not only their objectives, but also how they are implemented.

6. The normative approach to the role of government asks: How can government address market failures and other perceived inadequacies in the market's resource allocation? The positive approach asks: What is it that the government does, what are its effects, and how does the nature of the political process, including the incentives it provides bureaucrats and politicians, help explain what the government does and how it does it?

KEY CONCEPTS

Adverse selection

Incomplete markets

Libertarianism

Marginal revenue

Market failures

Merit goods
Monopolistic competition
Monopoly
Natural monopoly
Negative externalities
Oligopoly
Paternalism
Perfect competition
Positive externalities
Pure public goods
Research and development (R&D)
Tragedy of the anticommons
Tragedy of the commons

QUESTIONS AND PROBLEMS

1. For each program listed below, discuss what market failures might be (or are) used as a partial rationale:
   a. Automobile safety belt requirements
   b. Regulations on automobile pollution
   c. National defense
   d. Unemployment compensation
   e. Medicare (medical care for the aged)
   f. Medicaid (medical care for the indigent)
   g. Federal Deposit Insurance Corporation
   h. Federally insured mortgages
   i. Law requiring lenders to disclose the true rate of interest they are charging on loans (truth-in-lending law)
   j. National Weather Service
   k. Urban renewal
   l. Post office
   m. Government prohibition of the use of narcotics
   n. Rent control

2. If the primary objective of government programs in each of these areas is the alleviation of some market failure, how might they be better designed?
   a. Farm price supports
   b. Oil import quotas (in the 1950s)
   c. Special tax provisions for energy industries

3. Many government programs both redistribute income and correct a market failure. What are the market failures associated with each of these programs, and how else might they be addressed if there were no distributional objectives?
   a. Student loan programs
   b. Public elementary education
   c. Public support for universities
   d. Social Security

4. Draw the average and marginal cost curves for a natural monopoly. Draw the demand and marginal revenue curves.
   a. Show the efficient level of output, at which price equals marginal cost. Explain why if the firm charged a price equal to marginal cost, it would operate at a loss. Show diagrammatically the necessary subsidy.
   b. Show the monopoly level of output, at which marginal revenue equals marginal cost. Explain why the monopoly level of output is smaller than the efficient level of output.
   c. Show the level of output of a government monopoly that was instructed to just break even. How does this level of output compare with the efficient level of output and the private monopoly level of output? (Show diagrammatically.)

5

PUBLIC GOODS AND PUBLICLY PROVIDED PRIVATE GOODS

FOCUS QUESTIONS

1. What distinguishes public goods—those that are typically provided by governments—from privately provided goods? What do economists mean by *pure public goods*?

2. Why will private markets undersupply pure public goods? What is the free rider problem?

3. What are the special characterics and challenges of global public goods?

4. Why do governments provide goods that are not pure public goods?

5. What determines an efficient supply of pure public goods? How is the efficient supply affected by concerns about income distribution? How is it affected by the fact that taxes required to pay for the public good typically introduce distortions in the economy?

6. In what sense is efficient government a public good?

The government supplies a wide variety of goods, from national defense to education to police and fire protection. Some of these goods, like education, are also provided privately; others, like national defense, are the exclusive province of government. What are the economic properties of such goods? How do they differ from goods such as ice cream, automobiles, and the myriad of other goods that are provided principally through private markets?

Earlier chapters noted the central role played by prices in market economies. Because of the price system, markets result in an efficient allocation of resources. Prices ration private goods. Those consumers who are willing and able to pay the requisite price obtain the good. This chapter asks: What is distinctive about the goods typically provided by government? What prevents them in many cases from being provided privately? And if they are provided privately, why is the private supply likely to be inadequate?

PUBLIC GOODS

To distinguish between private and public goods, economists ask two basic questions. First, does the good have the property of rival consumption? **Rival consumption** means that if a good is used by one person, it cannot be used by another. For instance, if Lynn drinks a bottle of apple juice, Fran cannot drink that same bottle of apple juice. By contrast, **non-rival consumption** refers to cases in which one person's consumption does not detract from or prevent another person's consumption.

The classic example of non-rival consumption is national defense. If the government creates a military establishment that protects the country from attack, all citizens are protected. National defense costs are essentially unaffected when an additional baby is born or an additional individual immigrates to the United States. This stands in sharp contrast to private goods. It costs additional resources to provide another bottle of apple juice so Lynn and Fran can each have one. This is the only way for Lynn and Fran each to enjoy a bottle of apple juice. For a non-rival good, such as a lighthouse, although it would indeed cost more to build more lighthouses, there is essentially no additional cost for an additional ship to make use of an existing lighthouse.

The second question we ask to distinguish between private and public goods relates to the property of **exclusion**. Is it possible to exclude any individual from the benefits of the public good (without incurring great costs)? A ship going past a lighthouse, for instance, cannot be excluded from the benefits the lighthouse provides. Likewise, if the country is defended against attack by foreigners, then all citizens are protected; it is difficult to exclude anyone from the benefits. Clearly, if exclusion is impossible, then use of the price system is impossible, because consumers have no incentive to pay. By contrast, private goods always have the property of excludability: individuals can be excluded from enjoying a good unless they pay for it.

Generally, private goods have the properties of rival consumption and excludability; public goods are characterized by non-rival consumption and non-excludability. Goods for which there is *no* rivalry in consumption and for which exclusion is *impossible* are **pure public goods**. To develop a more complete picture of public goods (and pure public goods), we now examine the properties of non-rivalry and non-exclusion in greater detail. We will see how these properties may lead to market failures, creating a rationale for the public provision of public goods.

PUBLIC GOODS AND MARKET FAILURES

To isolate the role of excludability and rivalrousness in consumption, we consider instances in which a good has one property but not the other. For some goods, consumption is non-rival but exclusion is possible. For instance, the marginal cost of an additional individual turning on a television and watching a show is zero; the number of times I watch *American Idol* does not detract from the number of times you can watch it. Exclusion is possible, however, as illustrated by the proliferation of channels available only through subscription cable and satellite TV.

Even if exclusion were possible, when a good is non-rival, there is no impetus for exclusion from the standpoint of economic efficiency. Charging a price for a non-rival good prevents some people from enjoying the good, even though their consumption of the good would have no marginal cost. Thus, charging for a non-rival good is inefficient because it results in underconsumption. The marginal benefit is positive; the marginal cost (of the extra person watching the show) is zero. The underconsumption is a form of inefficiency.

But if there is no charge for a non-rival good, there will be no incentive for supplying the good. In this case, inefficiency takes the form of under-supply.

Thus, there are two basic forms of market failure associated with public goods: **underconsumption** and **undersupply**. In the case of non-rival goods, exclusion is undesirable because it results in underconsumption. Without exclusion, however, there is the problem of undersupply.

PAYING FOR PUBLIC GOODS

If exclusion is possible, even if consumption is non-rival, governments often charge **user fees** to those who benefit from a publicly provided good or service. Toll roads are financed by user fees. The airline ticket tax can be thought of as a user fee; revenues from the ticket tax go to finance airports and the air traffic control system. User fees are often thought of as an equitable way of raising revenues, as those who use the public facility the most (and therefore presumably benefit from it the most) pay the most. However, when consumption is non-rival, user fees introduce an inefficiency. We can use the sort of analysis that will be introduced in Chapter 7 to measure the inefficiency.

This is illustrated in Figure 5.1 for the case of a bridge. We have drawn the demand curve for the bridge, describing the number of trips taken

BRIDGES: HOW A USER FEE CAN RESULT IN UNDERCONSUMPTION

If the capacity is large enough, the bridge is a non-rival good. Although it is possible to exclude people from using the bridge by charging a toll, p, this results in an underconsumption of the good, Q_e, below the non-toll level of consumption, Q_m.
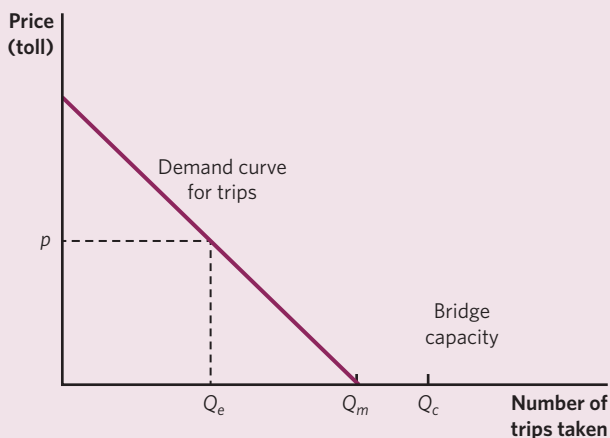
as a function of the toll charged. Lowering the toll results in increased demand for the bridge. The capacity of the bridge is Q_c; for any demand below Q_c, there is no congestion and no marginal cost associated with use of the bridge. As long as the bridge is operating below capacity, consumption is non-rival; additional consumption by one individual does not detract from what others can enjoy. Because the marginal cost of usage is zero, efficiency requires that the price for usage be zero. Clearly, though, the revenue raised by the bridge will then be zero.

This is where the difference between *public provision* and private is clearest: with a single bridge, the monopoly owner would choose a toll to maximize its revenue, and would build the bridge only if those revenues equaled or exceeded the cost of the bridge. The government would face a more complicated set of calculations. It might charge the toll required to just cover the costs of construction, to break even. In doing so, it would recognize that with any toll, the usage of the bridge would be reduced, and some trips whose benefits exceed the social cost (here, zero) would not be undertaken. Thus, it might charge a toll less than that required to break even, raising the revenue required to finance the bridge in some other way. It might not even charge any toll. In making these decisions, it would weigh equity considerations—the principle that those who benefit from the bridge should bear its costs—with efficiency considerations. The distortions arising from the underutilization of the bridge would need to be compared with the distortions associated with alternative ways of raising revenues (e.g., taxes) to finance the bridge. Finally, the government might build the bridge even if the maximum revenue it could obtain from the tolls was less than the cost of the

bridge, as it recognizes that there is some consumer surplus from the bridge: the amount that at least some individuals might be willing to pay for the bridge may be considerably greater than even the amount raised by the revenue-maximizing toll.

THE FREE RIDER PROBLEM

Many of the most important publicly provided goods—such as public health programs and national defense—have the property of non-excludability, making rationing by the price system unfeasible. For instance, the international vaccine program against smallpox virtually wiped out the disease, to the benefit of all, whether they contributed to supporting the program or not. National defense has the property of non-excludability and zero marginal cost, but there are a few goods that have the property of at least high costs of exclusion, even though the marginal cost of using the good is positive. Congested urban streets are an example: under current technology, it is expensive to charge for the use of the street (someone could collect tolls at each corner, but the cost would be extremely high) but the throughput of the street may be limited, so if one more person uses it, another is displaced—indeed, in some cases, as more people attempt to use the street, the total throughput of the street may even be decreased, as gridlock sets in.

The infeasibility of rationing by the price system implies that the competitive market will not generate a Pareto efficient amount of the public good. Assume that everyone valued national defense, but the government did not provide for it. Could a private firm enter to fill this gap? To do so, it would have to charge for the services it provided. However, because all individuals would believe that they would benefit from the services provided regardless of whether they contributed to the service, they would have no incentive to pay for the services *voluntarily*. That is why individuals must be forced to support these goods through taxation. The reluctance of individuals to contribute voluntarily to the support of public goods is referred to as the **free rider problem**.

An example will help to illustrate the nature of this problem. In many communities, fire departments are supported voluntarily. Some individuals refuse to contribute to the fire department, yet, in an area where buildings are close together, the fire department will usually put out a fire in a noncontributor's building because of the threat it poses to adjacent contributors' structures. Knowing that they will be protected even if they do not pay induces some people to be free riders.

Clearly, if it is not possible to use price to ration a particular good, the good is not likely to be provided privately. If it is to be provided at all, government will have to take responsibility.

In a few cases, non-excludable public goods are provided privately. Usually this is because there is a single, large consumer whose direct benefits are so large that it pays for the consumer to provide it for himself or herself. This consumer knows that there are free riders benefiting from these actions, but in deciding how much to supply, the consumer looks only at his or her own direct benefit, not at the benefits that accrue to others. For instance, a large shipowner might find it worthwhile to install a lighthouse and light buoys, even if others cannot be excluded from enjoying the benefits. But in deciding how many lighthouses and buoys to construct, the shipowner looks only at the benefits that accrue to his or her own ships. The total benefit of an additional buoy—including the benefits both to the shipowner's ships and to others, for instance—might be considerable, even though the direct benefit to its own ships might not warrant the additional cost. In that case, the shipowner would not put the additional buoy into place. Thus, even if there is *some* private provision of public goods, there will be an undersupply.

ECONOMISTS AND THE FREE RIDER PROBLEM

The free rider problem is just a reflection of an important incentive problem that arises in the case of public goods: If the good is going to be provided anyway, why should I pay? What I would contribute would be negligible, and would hardly alter the aggregate supply. To be sure, if everyone reasoned the same way, the good would not be supplied. That is one of the arguments for government's providing these goods, because government has the power to compel people to contribute (through taxes).

In many instances, though, public goods—volunteer fire departments, local charities, hospitals, public TV, and many others—are being supported voluntarily. How do we explain these? Is it possible that economists have overemphasized the "selfish" nature of humanity? Several recent studies in experimental economics have suggested that this might be the case. These experimental situations are designed to make individuals face situations in which they could free ride if they wanted to; alternatively, they could cooperate in providing for a public good. Systematically, more cooperative behavior and less free riding are observed than economists' analysis of selfish incentives would suggest. Interestingly, when economists participate in these experiments, their behavior systematically seems much closer in accord with the predictions of their theory.

PURE AND IMPURE PUBLIC GOODS

A pure public good is a public good for which the marginal costs of providing it to an additional person are strictly zero and for which it is impossible to exclude people from receiving the good. National defense is one of the few examples of a pure public good. Many public goods that government provides are not *pure* public goods in this sense. The cost of an additional person using an uncrowded interstate highway is very, very small, but not zero, and it is possible, though relatively expensive, to exclude people from (or charge people for) using the highway.

Figure 5.2 compares examples of goods that are often publicly provided with the strict definition of a *pure* public good. It shows the ease of exclusion along the horizontal axis and the (marginal) cost of an additional individual's using the good along the vertical axis. The lower left-hand corner represents a pure public good. Of the major public expenditures, only national defense comes close to being a pure public good. The upper right-hand corner represents a good (ordinary health services) for which the cost of exclusion is low and the marginal cost of an additional individual using the good is high. It is easy to charge each patient for health services, and it costs a doctor twice as much to see two patients as to see one—there are significant marginal costs of providing health services to additional individuals.

Many goods are not pure public goods but have one or the other property (non-rivalrousness or non-excludability) to some degree. Fire protection is like a private good in that exclusion is relatively easy—individuals

Marginal cost of use

Congested highway

Pure private good: Health services, education

Pure public good: National defense

Fire protection

Ease of exclusion

FIGURE 5.2

PUBLICLY PROVIDED GOODS

Pure public goods are characterized by non-rival consumption (the marginal cost of an additional individual's enjoying the good is zero) and non-excludability (the cost of excluding an individual from enjoying the good is prohibitively high). Goods provided by the public sector differ in the extent to which they have these two properties.

PUBLIC GOODS

- Pure public goods have the properties of perfectly non-rival consumption and non-excludability.
- With non-rival consumption, it is not desirable to exclude anyone from the benefits. With private provision, there will be underconsumption and/or undersupply.
- With non-excludability, it is not feasible to exclude anyone from the benefits of the good. There will be a free rider problem. Such goods typically cannot be provided by the market, and when they are privately provided, they will be undersupplied.

who refuse to contribute to the fire department could simply not be helped in the event of a fire. However, fire protection is like a public good in that the marginal cost of covering an additional person is low. Most of the time, firefighters are not engaged in fighting fires but are waiting for calls. Protecting an additional individual has little extra cost. Only in that rare event when two fires break out simultaneously will there be a significant cost to extending fire protection to an additional person. Even here, though, matters are more complicated: if we want to protect the building next door, which has paid for fire protection, it may be necessary to put out the fire in the building that has not paid for protection—exclusion may not really be feasible. Similarly, although the main beneficiary of a vaccination may be the individual protected, and there is a significant marginal cost of vaccinating an additional individual, the public health benefits from universal vaccination—the reduced incidence of the disease, possibly its eradication—are benefits from which no one can be excluded.

Sometimes the marginal cost of using a good to which access is easy (a good that possesses the property of non-excludability) will be high. When an uncongested highway turns congested, the costs of using it rise dramatically, not only in terms of wear and tear on the road but also in terms of the time lost by drivers using the road. It is costly to exclude by charging for road use—as a practical matter, this can be done only on toll roads, and, ironically, the tollbooths often contribute to the congestion. New technologies, which automatically bill regular users of the road, have radically reduced these costs.

COSTS OF EXCLUSION   For many goods, the issue is not so much the feasibility of rationing, but the cost. Thus, TV and radio provided over the airwaves has one of the two properties of a public good: consumption is non-rival. However, it is now feasible to exclude some consumers, as in the use of cable TV and satellite radio, although there is a cost to exclusion without any benefit to society from doing so. In other cases, such as a slightly crowded highway, there is a cost to exclusion (the cost of collecting tolls), and some benefit (less congestion).

There are, of course, costs associated with exclusion for private goods as well as for public goods. Economists call these **transactions costs**. For example, the salaries of checkout clerks at grocery stores and collectors of tolls along toll highways and at toll bridges are transactions costs,

part of the administrative costs associated with operating a price mechanism. Although the costs of exclusion are relatively small for most private goods, they may be large (prohibitive) for some publicly provided goods.

EXTERNALITIES AS IMPURE PUBLIC GOODS　Pure public goods have the property that if one individual purchases more of a good, all individuals' consumption of that good increases by the same amount. (Individuals may, of course, differ in how they value the increased consumption.) Pure private goods have the property that if one individual purchases more of a good, others are (at least directly) unaffected. Goods for which there are externalities in consumption have the property that others are affected, but not necessarily in the same amount. Externalities can thus be viewed as a form of impure public goods (or, perhaps better stated, public goods can be viewed as an extreme form of externalities).[1]

GLOBAL AND LOCAL PUBLIC GOODS　The benefits of some public goods are enjoyed only locally—by those living in a particular community. These are called *local public goods*, and include, for instance, local police security. However, more and more public goods are **global public goods**, the benefits of which accrue to anyone in the world. The most important global public goods include the global environment (everyone is affected by global warming, though in different ways), global health (everyone can be affected by the spread of a pandemic like the bird flu or other infectious diseases that can move easily across borders), global knowledge (everyone can benefit from developments in modern science), and global security.

Steep declines in transportation and communication costs have created a world in which everyone is more interdependent. Everyone can be affected by threats to global economic security (we worry about the contagion of financial crises) and physical security through global terrorism, cyber warfare, and increased nuclear proliferation.

Whenever there is a public good, there is a need for **collective action**; that is, individuals and firms acting privately will result in an undersupply of the public good. Within a country, the national government provides for national public goods, and at the local level, local governments provide for local public goods; however, there is no global government that provides for global public goods. There are, in particular, no taxes that are levied globally to finance global public goods. The locus of decision making remains largely within the nation-state.

---

[1] This is sometimes expressed by saying that for private goods, J's individual utility depends only on his own purchases, X^J. For a pure public good, J's utility depends on the *sum* of the purchases of all individuals: $X^1 + X^2 + \cdots + X^J + \cdots$. When there is an externality, J's utility may depend more heavily on his own purchases, but it may also depend on others' more weakly; for instance, it might depend on $aX^1 + aX^2 + \cdots + X^J + \cdots aX^n$, where a is a small number.

PROPERTY RIGHTS, EXCLUDABILITY, AND EXTERNALITIES

Some problems of excludability arise not from the *feasibility* of exclusion, but from imperfections in the legal structure that make exclusion difficult. Some economists, such as Nobel Prize–winning economist Ronald Coase, have argued that many public-good and externality problems would be resolved if property rights, which allow exclusion, were established.

Consider a crowded room. The air in the room is a public good: all persons in the room breathe essentially the same air. When any individual smokes, he or she creates an externality on others. In the absence of property rights, the smoker would fail to take that into account in his or her decision to smoke. If property rights were assigned, though, the problem would (so the argument goes) be resolved. Assume the "rights" to the air were given to a particular individual. (Coase argues that, apart from distribution, to whom it is given makes no difference.) That individual would then conduct an auction: he or she would ask the smokers how much they would be willing to pay to allow smoking, would ask the nonsmokers how much they would be willing to pay to prohibit smoking, and would then offer the air for sale to those for whom the value was highest. This would be efficient.*

This "solution," however, ignores the free rider problem: individual smokers have an incentive not to reveal the full value of the right to smoke (if they might actually have to pay that amount); and, similarly, individual nonsmokers have an incentive not to reveal the full value of clean air.

Although in some important cases, assigning property rights would reduce or eliminate externalities or public good problems, in some of the most important cases, assigning property rights is either impracticable or would not resolve the underlying problems. Alternative policy responses are required, such as government regulation or creative solutions that utilize local social capital and traditional community norms (see Chapter 4).

---

*The resource allocation would be Pareto efficient, *given the assignment of the property right of the air to a single individual*. Of course, however, if the smokers bid more than the nonsmokers, the smoker who winds up having to pay for the right to smoke is worse off relative to the initial situation where he or she could smoke without paying anything. Although in this example the smokers could have compensated the nonsmokers, such compensation may well not be made.

Still, the international community has come together to address many of these global public goods. International agreements and treaties, for instance, limit nuclear proliferation and attempt to protect some aspects of the global environment. In some cases, there are even enforcement mechanisms, although they tend to be relatively weak. Countries that did not comply with the agreement to reduce their utilization of gases that destroyed the vital layer of ozone in the atmosphere were threatened with trade sanctions. Typically, though, it is pressure from other countries that generates compliance and cooperation. In some cases, such as with the World Health Organization, most countries see the overwhelming benefits of cooperation, so they do so voluntarily.

International organizations like the World Health Organization and the broader United Nations have made important advances in global cooperation for provision of global public goods (and control of global public bads), but there is a long way to go. Developing more effective institutional arrangements, with more enforcement, and with the ability to raise funds to finance global public goods, is one of the main challenges facing the world today.[2]

PUBLICLY PROVIDED PRIVATE GOODS

Publicly provided goods for which there is a large marginal cost associated with supplying additional individuals are referred to as **publicly provided private goods**. Although the costs of running a market provide one of the rationales for the public supply of some of these goods, it is not the only or even the most important rationale. Education is a publicly provided private good in the sense defined previously—if the number of students enrolled doubles, costs will roughly double (assuming that quality, as reflected in class size, expenditures on teachers and textbooks, and so on, are kept roughly the same). One of the usual explanations given for public provision of education is concerned with distributive considerations; many feel that the opportunities of the young should not depend on the wealth of their parents.

Sometimes when the government provides a private good publicly (like water), it simply allows individuals to consume as much as they want

---

[2] The concept of local public goods was developed by J. E. Stiglitz, "Theory of Local Public Goods," in *The Economics of Public Services*, ed. M. S. Feldstein and R. P. Inman (New York: MacMillan, 1977), pp. 274–333; and that of global public goods by J. E. Stiglitz, "The Theory of International Public Goods and the Architecture of International Organizations," Background Paper No. 7, Third Meeting, High Level Group on Development Strategy and Management of the Market Economy, UNU/WIDER, Helsinki, Finland, July 8–10, 1995.

For an extensive discussion of knowledge as a global public good, see J. E. Stiglitz, "Knowledge as a Global Public Good," in *Global Public Goods: International Cooperation in the 21st Century*, ed. I. Kaul, I. Grunberg, and M.A. Stern, United Nations Development Programme (New York: Oxford University Press, 1999), pp. 308–325 . For a discussion of the role of the international financial institutions, such as the World Bank and the IMF, in the provision of international public goods, see J. E. Stiglitz, "IFIs and the Provision of International Public Goods," European Investment Bank, *Cahiers Papers* 3, no. 2 (1998): 116–134.

For a good summary of the challenges of global public goods, see the other chapters in the Kaul et al. and W. Nordhaus, "Some Foundational and Transformative Grand Challenges for the Social and Behavioral Sciences: The Problem of Global Public Goods," paper submitted to the National Science Foundation as part of its SBE 2020 planning activity, September 2010; for application of these general concepts to global environmental public goods, see R. Arriagada and C. Perrings, "Paying for International Environmental Public Goods," Ecosystem Services Economics Working Paper No. 17, United Nations Environment Programme, October 2011.

without charge. Remember: for these goods, there is a marginal cost associated with each unit consumed. It costs money to purify water and to deliver it from the source to the individual's home. If a private good is provided freely, there is likely to be overconsumption of the good. Because individuals do not have to pay for the good, they will demand it until the point at which the marginal benefit they receive from the good is zero, in spite of the fact that there is a positive marginal cost associated with providing it. In some cases, such as water, satiation may be quickly reached, so the distortion from overconsumption may not be too large (Figure 5.3A). In other cases, such as the demand for certain types of medical services, the distortion may be very large (Figure 5.3B). The welfare loss can be measured by the difference between what the individual is willing to pay for the increase in output from Q_e (where price equals marginal cost) to

FIGURE 5.3

DISTORTIONS ASSOCIATED WITH SUPPLYING GOODS FREELY

(A) For some goods, such as water, supplying the good freely rather than at marginal costs results in relatively little additional consumption. (B) For other goods, such as certain medical services, supplying the good freely rather than at marginal costs results in extensive overconsumption.
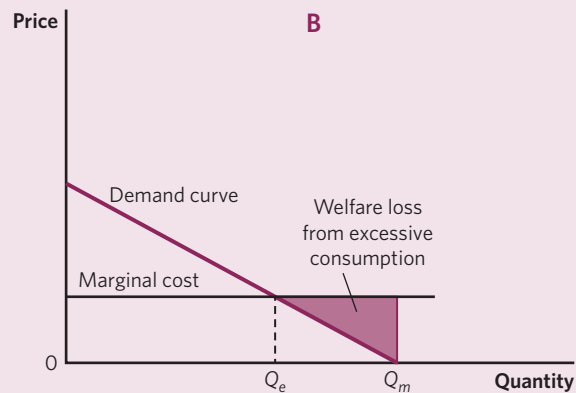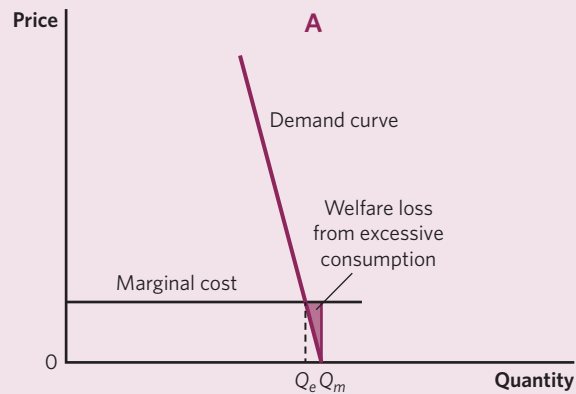
Q_m (where price equals zero) and the costs of increasing production from Q_e to Q_m. This is the area of the shaded triangles in Figure 5.3.

RATIONING DEVICES FOR PUBLICLY PROVIDED PRIVATE GOODS

When there is a marginal cost associated with each individual using a good, if the costs of running the price system are very high, it may be more efficient simply to provide the good publicly and to finance the good through general taxation, even though providing the good publicly causes a distortion. We illustrate this in Figure 5.4, in which we have depicted a good with constant marginal costs of production, c. (It costs the firm c dollars to produce each unit of the good.)[3] However, to sell the good entails transactions costs, which raise the price to p^*. Assume now that the government supplies the good freely. This eliminates the transactions costs, and the entire area in rectangle $ABCD$ is saved. There is a further gain as consumption increases from Q_e to Q_o, as individuals' marginal valuations exceed the marginal costs of production. The shaded area ABE measures the gain. On the other hand, if individuals consume the good until the marginal value is zero, in expanding consumption from Q_o to

---

[3] We assume, moreover, that the demand curve does not shift significantly as we raise taxes.
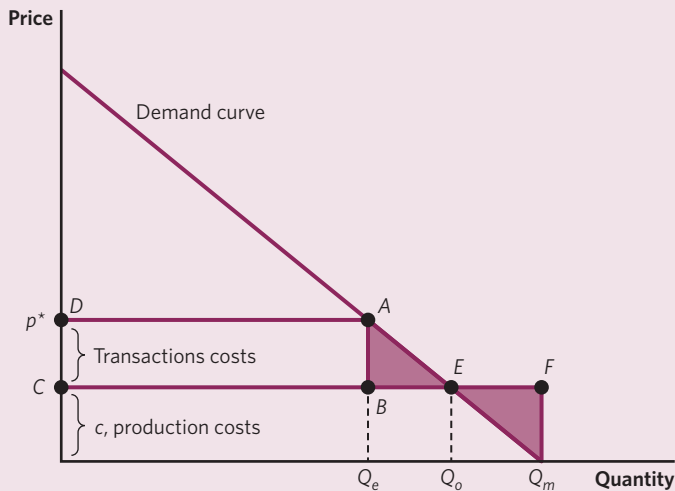
TRANSACTIONS COSTS
When transactions costs are sufficiently high, it may be more efficient to supply the good publicly than to have the good supplied by private markets.

Q_m, the marginal willingness to pay is less than the cost of production. This is obviously inefficient. To decide whether the good should be provided publicly, we must compare the savings in transactions costs plus the gain from increasing consumption from Q_e to Q_o with (1) the loss from the excessive consumption of the good (the shaded area EFQ_m in Figure 5.4) and (2) the loss from the distortions created by any taxes required to finance the provision of the good publicly.

The high costs of private markets' providing insurance has been used as one of the arguments for the public provision of insurance. For many kinds of insurance, the administrative costs (including the selling costs) associated with providing the insurance privately are more than 20 percent of the benefits paid out, in contrast with the administrative costs associated with public insurance, which (ignoring the distortions associated with the taxes required to finance the administrative costs of the social insurance programs) are usually less than 10 percent of the value of the benefits.

Given the inefficiencies arising from overconsumption when no charges are imposed for publicly provided private goods, governments often try to find some way of limiting consumption. Any method restricting consumption of a good is called a **rationing system**. Prices provide one rationing system. We have already discussed how user fees may be employed to limit demand. A second commonly employed way of rationing publicly provided goods is **uniform provision**: supplying the same quantity of the good to everyone. Thus, we typically provide a uniform level of free education to all individuals, even though some individuals would like to have more and some less. (Those who would like to purchase more may be able to purchase supplemental educational services like tutoring on the private market.) This, then, is the major disadvantage of the public provision of goods; it does not allow for adaptation to differences in individuals' needs and desires as does the private market.

This is illustrated in Figure 5.5, in which the demand curves for two different individuals are drawn. If the good was privately provided, Individual 1, the high demander, would consume Q_1, while Individual 2, the low demander, would

THREE METHODS OF RATIONING PUBLICLY PROVIDED GOODS

1. User charges
- Advantage: Those who benefit bear the costs.
- Disadvantages: Results in underconsumption.
  Administering pricing system adds transactions costs.

2. Uniform provision
- Advantage: Saves on transactions costs.
- Disadvantages: Leads some to underconsume, others to overconsume.
  High demanders may supplement public consumption, increasing total transactions costs.

3. Queuing
- Advantage: Goods (like health care) allocated not necessarily on basis of who is wealthiest.
- Disadvantages: Alternative basis of allocation (who has time to spare) may be undesirable.
  Time is wasted.
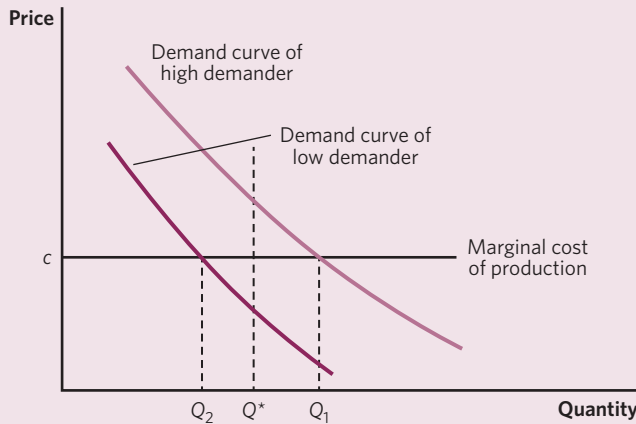
consume the much smaller quantity Q_2. The government chooses to supply each individual with a quantity that is somewhere in between, Q^*. At this level of consumption, the high demander is consuming less than he or she would like; the high demander's marginal willingness to pay exceeds the marginal cost of production. On the other hand, the low demander is consuming more than the efficient level; his or her marginal willingness to pay is less than the marginal cost. (Because the low demander does not have to pay anything for it, and still values the good positively, he or she, of course, consumes up to Q^*.)

For certain types of insurance, say, Social Security for retirement, the government provides a basic, uniform level. Again, those who wish to purchase more can do so, but those who wish to purchase less cannot. The distortion here may not be very great, however; if the uniform level provided is sufficiently low, then relatively few individuals will be induced to consume more than they would otherwise, and the savings in administrative costs that we referred to earlier may more than offset the slight distortion associated with the uniform provision of the basic level of insurance. On the other hand, the system of combining public and private provision may increase total transactions (administrative) costs over what they would be if only the public sector or private sector took responsibility.

A third method of rationing that is commonly employed by the government is **queuing**: rather than charging individuals money for access to the publicly provided goods or services, the government requires that they pay a cost in waiting time. This allows some adaptability of the level of supply to the needs of the individual. Those whose demand for medical services is stronger are more willing to wait in the doctor's office. It is claimed that money is an undesirable basis upon which to ration medical

services: Why should the wealthy have a greater right to good health than the poor? Queues, it is argued, may be an effective device for discriminating between the truly needy (who are willing to wait in line) and those who are less needy of medical care. But queues are far from a perfect way of determining who is deserving of medical care, as those who are unemployed or retired may not be so needy of medical care but more willing to wait than either the busy corporate executive or the low-paid worker holding down two jobs. In effect, willingness to pay has been replaced as a criterion for allocating medical services by willingness to wait in the doctor's office. There is, in addition, a real social cost to using queuing as a rationing device—the waste of time spent queuing is a cost that could be avoided if prices were used as a rationing device.

EFFICIENCY CONDITIONS FOR PUBLIC GOODS

A central concern is how large the supply of public goods should be. What does it mean to say that the government is supplying too little or too much of a public good? Chapter 3 provided a criterion that enables us to answer this question: a resource allocation is Pareto efficient if no one can be made better off without someone else becoming worse off. There we established that Pareto efficiency in private markets requires, among other criteria, that the individual's marginal rate of substitution is equal to the marginal rate of transformation.

In this section of the chapter, we characterize what is required for Pareto efficiency in the supply of pure public goods, and in particular, goods for which the marginal cost of an additional individual enjoying them is zero. *Pure public goods are efficiently supplied when the sum of the marginal rates of substitution (over all individuals) is equal to the marginal rate of transformation.* The *marginal rate of substitution* of private goods for public goods tells how much of the private good each individual is willing to give up to get one more unit of the public good. The sum of the marginal rates of substitution thus tells us how much of the private good all the members of society together are willing to give up to get one more unit of the public good, which will be consumed jointly by all. The *marginal rate of transformation* tells us how much of the private good must be given up to get one more unit of the public good. Efficiency requires, then, that the total amount individuals are willing to give up—the sum of the marginal rates of substitution—must equal the amount that they have to give up—the marginal rate of transformation.

Let's apply this efficiency condition to national defense. Assume that when we increase production of guns (national defense) by one, we must reduce production of butter (basic consumer products) by one pound (the marginal rate of transformation is unity). Guns used for national defense are a public good. We consider a simple economy with two individuals: Crusoe and Friday. Crusoe is willing to give up one-third of a pound of butter for an extra gun, but his one-third pound alone does not buy the gun. Friday is willing to give up two-thirds of a pound of butter for an extra gun. The total amount of butter that this small society would be willing to give up, were the government to buy one more gun, is $1/3 + 2/3 = 1$.

The total amount they would *have to* give up to get one more gun is also one. Thus, the sum of the marginal rates of substitution equals the marginal rate of transformation; their government has provided an efficient level of national defense. If the sum of the marginal rates of substitution exceeded unity, then, collectively, individuals would be willing to give up more than they had to; we could ask each of them to give up an amount slightly less than the amount that would make them indifferent, and it would still be possible to increase the production of guns by one unit. Thus, they could all be made better off by increasing the production of the public good (guns) by one.

DEMAND CURVES FOR PUBLIC GOODS

Individuals do not buy public goods. We can, however, ask how much individuals would demand if they had to pay a given amount for each extra unit of the public good. This is not a completely hypothetical question, for as expenditures on public goods increase, so do individuals' taxes. We call the extra payment that an individual has to make for each extra unit of the public good the **tax price**. In the following discussion, we shall assume that the government has the discretion to charge different individuals different tax prices.

Assume that the individual's tax price is p; that is, for each unit of the public good, the individual must pay p. Then, the total amount the individual can spend, the *budget constraint*, is:

$$C + pG = Y,$$

where C is the individual's consumption of private goods, G is the total amount of public goods provided, and Y is the individual's income. The budget constraint shows the combinations of goods (here, public and private goods) that the individual can purchase, given his or her income and tax price. The budget constraint is represented in Figure 5.6A by the line BB. Along the budget constraint, if government expenditures are
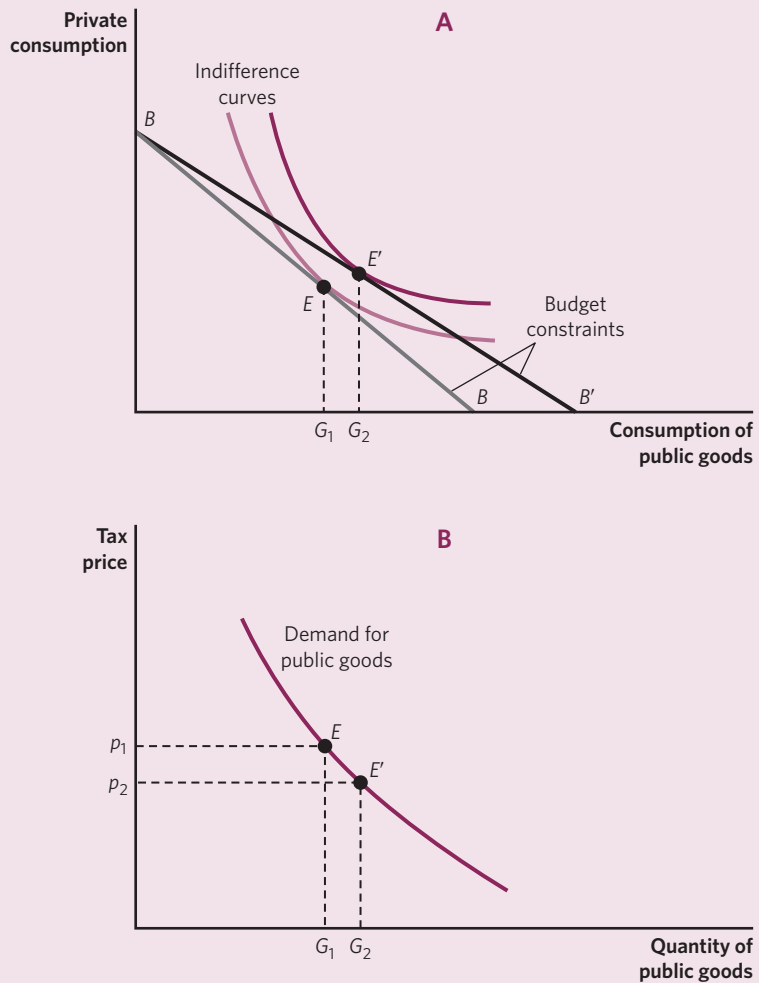
lower, consumption of private goods is obviously higher. The individual
wishes to obtain the highest level of utility he or she can, consistent with
the budget constraint. Figure 5.6A also shows the individual's indiffer-
ence curves between public and private goods. The individual is willing to
give up some private goods to get more public goods. The quantity of pri-
vate goods the individual is willing to give up to get one more unit of pub-
lic goods is his or her marginal rate of substitution. As the individual gets
more public goods (and has fewer private goods), the amount of private
goods he or she is willing to give up to get an extra unit of public goods
becomes smaller; that is, the individual has a diminishing marginal rate
of substitution. Graphically, the marginal rate of substitution is the slope

of the indifference curve. Thus, as the individual consumes more public goods and fewer private goods, the indifference curve becomes flatter.

The individual's highest level of utility is attained at the point of tangency between the indifference curve and the budget constraint, point *E* in Figure 5.6A. At this point, the slope of the budget constraint and the slope of the indifference curve are identical. The slope of the budget constraint tells us how much in private goods the individual must give up to get one more unit of public goods; it is equal to the individual's tax price. The slope of the indifference curve tells us how much in private goods the individual is willing to give up to get one more unit of public goods. Thus, at the individual's most preferred point, the amount that he or she is *willing* to give up to get an additional unit of public goods is just equal to the amount he or she *must* give up to get one more unit of the public good. As we lower the tax price, the budget constraint shifts out (from *BB* to *BB'*), and the individual's most preferred point moves to point *E'*. The individual's demand for public goods will normally increase.

By raising and lowering the tax price, we can trace out a demand curve for public goods, in the same way that we trace out demand curves for private goods. Figure 5.6B plots the demand curve corresponding to Figure 5.6A. Points *E* and *E'*, from panel A, show the quantity of public goods demanded at tax prices p_1 and p_2. We could trace more points for panel B by shifting the budget constraint further in panel A.

We can first use this approach to trace out the demand curves for public goods of two individuals, Crusoe and Friday, and then add them *vertically* to derive the **collective demand curve** in Figure 5.7. Vertical summation is appropriate because a pure public good is necessarily provided in the same amount to all individuals. Rationing is infeasible and is also undesirable, as one individual's usage of the public good does not detract from any other individual's enjoyment of it. Therefore, for a given quantity we add up everyone's willingness to pay to calculate the total willingness to pay; by calculating this amount at every quantity, we trace out the collective demand curve.[4]

The demand curve can be thought of as a "marginal willingness to pay" curve. That is, at each level of output of the public good, it says how much the individual would be willing to pay for an extra unit of the public good. (Remember, the tax price for the public good faced by the individual is set equal to his or her marginal rate of substitution, which simply gives how much of the private good the individual is willing to give up for one more unit of the public good.) Thus, the vertical sum of the demand curves is just the sum of their marginal willingness to pay; that is, it is the

[4] The collective demand curve is also sometimes referred to as the *aggregate demand curve* (not to be confused with the same term used in macroeconomics), and is the analog to the market demand curve for private goods. When constructing the market demand curve for private goods, we add up the quantities demanded for a given price, because all individuals face the same price but may consume different amounts. The *market demand curve* is thus derived by adding up the individual demand curves horizontally.

FIGURE 5.7

COLLECTIVE DEMAND FOR PUBLIC GOODS

Because the price is equal to the marginal rate of substitution at each point on the demand curve, by adding the demand curves vertically we obtain the sum of the marginal rates of substitution, the total amount of private goods that the individuals in society are willing to give up to get one more unit of the public good. The vertical sum thus can be thought of as generating the collective demand curve for the public good.
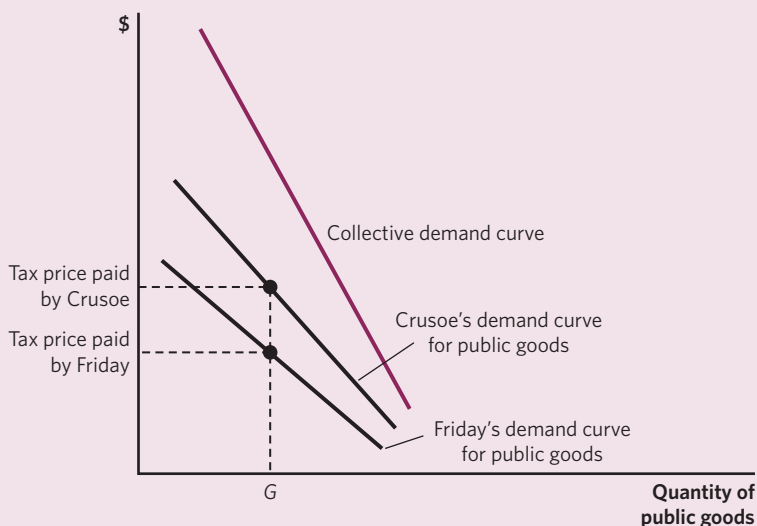
total amount that all individuals together are willing to pay for an extra unit the public good. Equivalently, because each point on the demand curve of an individual represents the individual's marginal rate of substitution at that level of government expenditure, by adding the demand curves vertically we simply obtain the sum of the marginal rates of substitution (the total marginal benefit from producing an extra unit). The result is the collective demand curve shown in Figure 5.7.

We can draw a supply curve just as we did for private goods; for each level of output, the price represents how much of the other goods must be forgone to produce one more unit of public goods; this is the marginal cost, or the marginal rate of transformation. At the output level at which the collective demand is equal to the supply (Figure 5.8), the sum of the marginal willingness to pay (the sum of the marginal rates of substitution) is just equal to the marginal cost of production or the marginal rate of transformation. Because, at this point, the marginal benefit from producing an extra unit of the public good equals the marginal cost, or the sum of the marginal rates of substitution equals the marginal rate of transformation, the output level described by the intersection of the collective demand curve and the supply curve for public goods is Pareto efficient.

Although we constructed each individual's demand curve for public goods in a manner analogous to the manner in which we could construct his or her demand curve for private goods, there are some important distinctions between the two. In particular, although market *equilibrium* occurs at the intersection of the demand and supply curves, we have not provided any explanation for why the equilibrium supply of public goods

Price
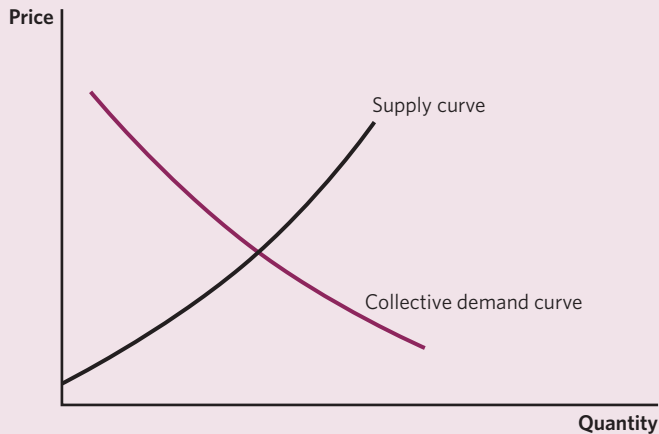
EFFICIENT PRODUCTION OF PUBLIC GOODS
An efficient supply of public goods occurs at the point of intersection of the demand curve and the supply curve. The collective demand curve gives the sum of what all individuals are willing to give up, at the margin, to have one more unit of public goods (one more gun), whereas the supply curve gives the amount of other goods that must be given up to obtain one more unit of the public good.

should occur at the intersection of the demand curve we have constructed and the supply curve. We have only established that if it did, the level of production of the public good would be Pareto efficient. Decisions about the level of public goods are made publicly, by governments, and not by individuals; hence, whether production occurs at this point depends on the nature of the political process, a subject discussed at length in Chapter 9.

Moreover, whereas in a competitive market for private goods all individuals face the same prices but consume different quantities (reflecting differences in tastes), a pure public good must be provided in the same amount to all affected individuals, and we have hypothesized that the government could charge different tax prices for the public good. One way of thinking about these prices is to suppose that each individual is told beforehand the *share* of public expenditures that he or she will have to bear. If some individual must bear 1 percent of the cost of public expenditures, then an item that costs the government $1 costs this individual 1 cent, whereas if an individual has to bear 3 percent of the cost of public expenditures, then an increase in public expenditures by $1 costs that individual 3 cents.

Finally, we should emphasize that we have characterized the Pareto efficient level of expenditure on public goods corresponding to a particular distribution of income. As we see in the next section, the efficient level of expenditure on public goods generally depends on the distribution of income.

EFFICIENCY CONDITION FOR PURE PUBLIC GOODS

- Efficient production requires that the sum of the marginal rates of substitution equal the marginal rate of transformation.
- Efficient production occurs at the intersection of the collective demand curve, formed by vertically adding the demand curves for each individual, with the supply curve.

PARETO EFFICIENCY AND INCOME DISTRIBUTION

As discussed in Chapters 3 and 7, there are many Pareto efficient resource allocations; any point on the utility possibilities schedule is Pareto efficient. The market equilibrium in the absence of market failures corresponds to just one of those points. By the same token, there is not a unique Pareto optimal supply of public goods. The intersection of the demand and supply curves in Figure 5.8 is one of these Pareto efficient levels of supply, but there are others as well, with different distributional implications.

To see how the efficient level of public goods depends on the distribution of income, assume the government transferred a dollar of income from Crusoe to Friday. This would normally shift Crusoe's demand for public goods (at any price) down and Friday's up. In general, there is no reason why these changes should exactly offset each other, so the aggregate level of demand will normally change. With this new distribution of income, there is a new efficient level of public goods. However, efficiency is still characterized by the sum of the marginal rates of substitution equaling the marginal rate of transformation. To put it another way, each point on the utility possibilities schedule may be characterized by a different level of public goods, but at each point the sum of the marginal rates of substitution equals the marginal rate of transformation.

The fact that the efficient level of public goods depends, in general, on the distribution of income has one important implication: one cannot separate efficiency considerations in the supply of public goods from distributional considerations. Any change in the distribution of income, say, brought about by a change in the income tax structure, will thus be accompanied by corresponding changes in the efficient levels of public-goods production.[5]

LIMITATIONS ON INCOME REDISTRIBUTION AND THE EFFICIENT SUPPLY OF PUBLIC GOODS

Governments, in evaluating the benefits of a public program, often seem to be particularly concerned with the question of *who* benefits from the program. They seem to weigh benefits that accrue to the poor more heavily than benefits that accrue to the rich. However, the previous analysis

---

[5] Some economists have suggested that decisions concerning the efficient level of public-goods production and distribution of income can be separated; for instance, there is a view that decisions about the distribution of income should be reflected in tax schedules and welfare programs, but decisions about the supply of public goods can and should be made independently of such considerations. In some cases, the decisions can be separated, but these are indeed special. See Atkinson and Stiglitz, *Lectures in Public Economics* (New York: McGraw-Hill, 1980) or L. J. Lau, E. Sheshinski, and J. E. Stiglitz, "Efficiency in the Optimum Supply of Public Goods," *Econometrica* 46 (1978): 269–284.

suggested that one should simply add up the marginal rates of substitution, the amounts that each individual is willing to pay at the margin for an increase in the public good, treating the rich and the poor equally. How can these approaches be reconciled?

Recall our example of the Robinson Crusoe economy, in which, in the process of transferring oranges from Crusoe to Friday, some of the oranges are lost. In the U.S. economy, we use primarily the tax system and welfare system to redistribute resources. Not only are the administrative costs of running these systems large, but they also have important incentive effects, for instance, on individuals' savings and work decisions. The fact that redistributing resources through the tax and welfare systems is costly implies that the government may look for alternative ways to achieve its redistributive goals; one way is to incorporate redistributive considerations into its evaluation of public projects.

DISTORTIONARY TAXATION AND THE EFFICIENT SUPPLY OF PUBLIC GOODS

The fact that the revenue raised to finance public goods is raised through distortionary taxes, such as the income tax, has some important implications for the efficient supply of public goods. The amount of private goods that individuals must give up to get one more unit of public goods is greater than it would be if the government could raise revenue in a way that did not entail distortionary incentive effects and that was not costly to administer.

We can define a **feasibility curve**, giving the maximum level of private-goods consumption consistent with each level of public goods, for our given tax system. The tax system introduces inefficiencies, so this feasibility curve lies inside the production possibilities schedule, as in Figure 5.9.

The amount of private goods we must give up to obtain one more unit of public goods, taking into account these extra costs, is called the **marginal economic rate of transformation**, as opposed to the **marginal physical rate of transformation** we employed in our earlier analysis. The latter is completely determined by *technology*, whereas the marginal economic rate of transformation takes into account the costs associated with the taxes required to finance increased public expenditure. Thus, we replace the earlier condition, that the marginal physical rate of transformation must equal the sum of the marginal rates of substitution, with a new condition: that the marginal economic rate of transformation must equal the sum of the marginal rates of substitution.

Because it becomes more costly to obtain public goods when taxation imposes distortions, normally this will imply that the efficient level of public goods is smaller than it would have been with nondistortionary taxation.

FIGURE 5.9

THE FEASIBILITY CURVE
The feasibility curve gives the
maximum output (consumption)
of private goods for any level
of public goods, taking into
account the inefficiencies
that arise from the taxes that
must be imposed to raise the
requisite revenue. The feasibility
curve lies below the production
possibilities schedule.
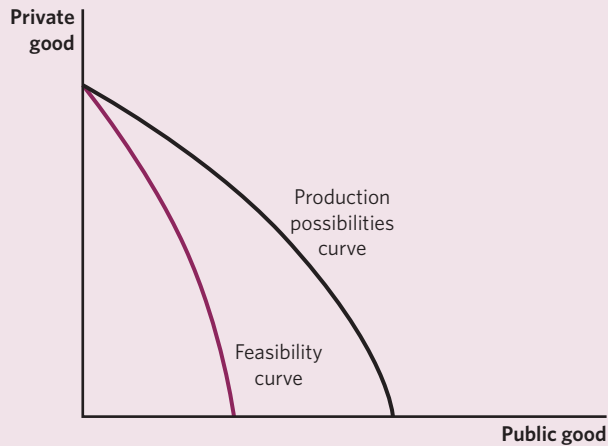
Indeed, it appears that much of the debate about the desirable level of public goods provision centers around this issue. Some believe that the distortions associated with the tax system are not very great, whereas others contend that the cost of attempting to raise additional revenues for public goods is great. They *may* agree on the magnitude of the social benefits that may accrue from additional government expenditures, but disagree on the costs.

EFFICIENT GOVERNMENT AS A PUBLIC GOOD

One of the most important public goods is the management of the government: we all benefit from a better, more efficient, more responsive government. Indeed, "good government" possesses both of the properties of public goods we noted earlier: it is difficult and undesirable to exclude any individual from the benefits of a better government.

If the government is able to become more efficient and reduce taxes without reducing the level of government services, everyone benefits. The politician who succeeds in doing this may get some return, but this return is only a fraction of the benefits that accrue to others. In particular, those who voted against the politician who succeeds in doing this gain as much as those who worked for the politician's election, and the individual who did not vote, who attempted to free ride on the political activities of others, benefits as much as either.

REVIEW AND PRACTICE

SUMMARY

1. This chapter has defined an important class of goods, pure public goods. They have two critical properties:

   a. It is impossible to exclude individuals from enjoying the benefits of the goods (non-excludability).

   b. The marginal cost of an additional individual enjoying the good is zero (non-rival consumption). It is undesirable to exclude individuals from enjoying the benefits of the goods, since their enjoyment of these goods does not detract from the enjoyment of others.

2. Although there are a few examples of pure public goods, such as national defense, for many publicly provided goods exclusion is possible, although frequently costly. Charging for use may result in the underutilization of public facilities. For many publicly provided goods, there is *some* marginal cost of an individual enjoying the good. Although, for example, the marginal cost of an individual using a completely uncongested road may be negligible, if there is some congestion, the marginal cost may be more significant.

3. Private markets either will not supply or will provide an inadequate supply of pure public goods.

4. The problem with voluntary arrangements for providing public goods arises from individuals trying to be *free riders* and enjoying the benefits of the public goods paid for by others.

5. Developing efficient and effective mechanisms for the provision of global public goods is one of the greatest challenges now facing the international community.

6. For many publicly provided goods, consumption is rivalrous; consumption by one individual reduces that of another; or the marginal cost of supplying an extra individual may be significant, equal to, or even greater than, the average cost. These are called publicly provided private goods. If they are supplied freely, there will be overconsumption.

7. For publicly provided private goods, some method of rationing other than the price system may be used; sometimes queuing is used, whereas at other times the good is simply provided in fixed quantities to all individuals. Both of these entail inefficiencies.

8. Pareto efficiency requires that a public good be supplied up to the point at which the sum of the marginal rates of substitution equals the marginal rate of transformation. Different Pareto efficient levels of consumption of the public good will be associated with different distributions of income.

9. The basic rule for the efficient level of supply of public goods must be modified when there are costs (distortions) associated with raising revenue and redistributing income.

10. Efficient management of the government is a public good in itself.

KEY CONCEPTS

Collective action

Collective demand curve

Exclusion

Feasibility curve

Free rider problem

Global public goods

Leftover curve

Marginal economic rate of transformation

Marginal physical rate of transformation

Non-rival consumption

Publicly provided private goods

Pure public goods

Queuing

Rationing system

Rival consumption

Tax price

Transactions costs

Underconsumption

Undersupply

Uniform provision

User fees

QUESTIONS AND PROBLEMS

1. Where should each of the following goods lie in Figure 5.2? Explain why each is or is not a pure public good. Where applicable, note instances where the good is both publicly and privately provided:

    a. College education

    b. A local park

    c. Yosemite National Park

    d. Sewage collection

    e. Water

    f. Electricity

    g. Telephone service

    h. Retirement insurance

    i. Medicine

    j. Police protection

    k. TV

    l. Basic research

    m. Applied research

2. What happens to the efficient allocation between public and private goods as an economy becomes wealthier? Can you think of examples of public goods, the consumption of which would increase more than proportionately to the increase in income? Less than proportionately to the increase in income?

3. The government rations a variety of publicly provided private goods and impure public goods in which there is congestion in a variety of ways. Discuss how each of these is rationed, and consider the effect of alternative rationing systems:

    a. Public higher education

    b. Health services in the United Kingdom

    c. Yellowstone National Park

    What happens to a publicly provided good in which congestion can occur (e.g., a highway or swimming pool on a hot, sunny day), but no direct rationing system is employed?

4. To what extent do you think differences in views between advocates of less spending on public goods and advocates of more spending can be attributed to different assessments of the marginal cost of public goods, including the increased distortions associated with the additional taxes required to finance public goods? What are other sources of disagreement?

5. What implications might the fact that efficient government is a public good have for the efficiency with which governments function?

6. Discuss the issue of vaccination from the perspective of public goods/externalities. Why might individuals not consent voluntarily to be vaccinated?

7. There has been increasing concern about increased atmospheric concentrations of greenhouse gases, such as carbon dioxide, which are likely to lead to global warming. Discuss the world's atmosphere as an "international" public good. What are some of the problems of ensuring that individuals and countries take actions to reduce emissions of greenhouse gases?

8. Discuss how changes in income, technology, or other changes in the economic environment may lead to changes in the balance between public and private provision. Illustrate, for instance, by a discussion of the role of public parks.

APPENDIX: THE LEFTOVER CURVE

In this appendix we provide an alternative, diagrammatic exposition for the basic efficiency condition for public goods discussed in the chapter:

Sum of marginal rates of substitution = Marginal rate of transformation.

In Figure 5.10A we have superimposed Crusoe's indifference curve on the production possibilities schedule. If the government provides a level of public goods G, and wishes, at the same time, to ensure that Crusoe attains the level of utility associated with the indifference curve U_1 drawn in the figure, then the amount of private good that is "left over" for Friday is the vertical distance between the production possibilities schedule and the indifference curve. Accordingly, we call the (vertical) difference between the two the **leftover curve**. (This curve is plotted in Figure 5.10B.) We now superimpose Friday's indifference curves on Figure 5.10B. The highest level of utility he can attain, consistent with the production possibilities schedule, and consistent with the prespecified level of utility of Crusoe, is at the point of tangency between his indifference curve and the leftover curve.

There is a simple way to express this tangency condition. Because the leftover curve represents the difference between the production possibilities schedule for the economy and the first individual's indifference curve, the slope of the leftover curve is the difference between the slope of the production possibilities schedule and the slope of the first individual's indifference curve. The slope of the production possibilities schedule is, as we just saw, the marginal rate of transformation, whereas the slope of the first individual's indifference curve is his or her marginal rate of substitution. If G is the optimal level of public goods, the leftover curve must be tangent to the second individual's indifference curve. Hence, Pareto efficiency of the economy requires that the slope of the leftover curve be equal to the slope of the second individual's indifference curve—that is,

$$MRT - MRS^1 = MRS^2$$

or

$$MRT = MRS^1 + MRS^2,$$

where MRT stands for the marginal rate of transformation, MRS^1 for the first individual's marginal rate of substitution, and MRS^2 for that of the second individual. The marginal rate of transformation must equal the sum of the marginal rates of substitution.

FIGURE 5.10

DETERMINATION OF
THE EFFICIENT LEVEL
OF PRODUCTION OF
PUBLIC GOODS

(A) If the level of public goods is
G, and Crusoe is to get level of
utility U_1, then the distance AB
represents the amount of pri-
vate goods left over for Friday.
(B) Friday's welfare is maximized
at the point of tangency of his
indifference curve and the
"leftover" curve.
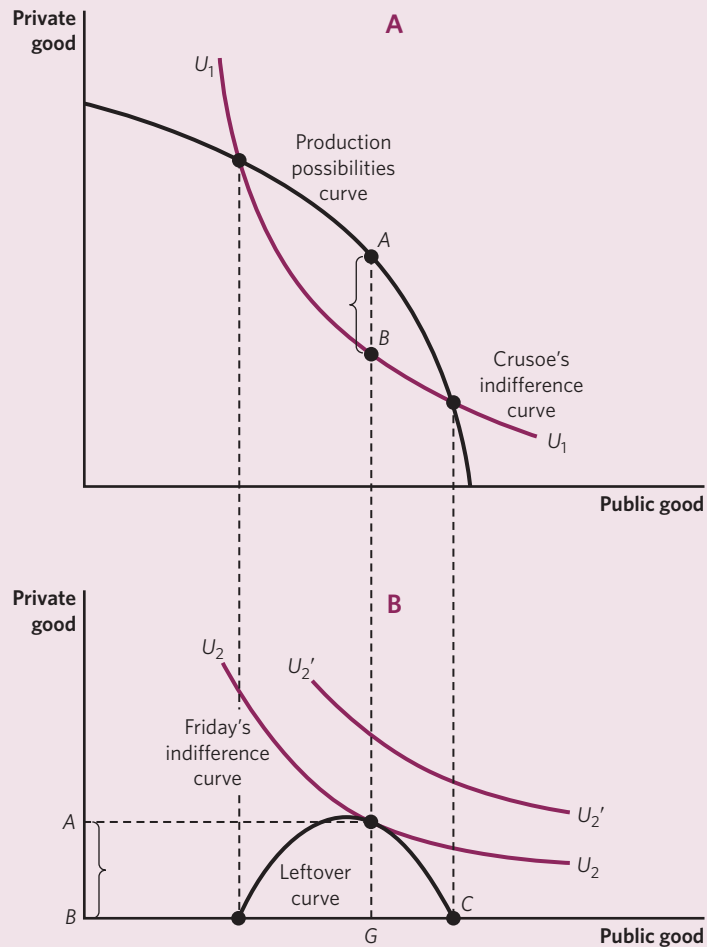
It should be clear that if we chose a different (say, higher) initial level of utility for Crusoe, the leftover curve would be shifted (down), and there would be a new point of tangency of Friday's indifference curve with the new leftover curve. At the new point of tangency, the level of expenditure on public goods may be higher, lower, or the same as in the initial situation. This illustrates the point made in the text: there is not necessarily a single "efficient" level of expenditure on public goods; there are many Pareto efficient levels of expenditures, depending on the distribution of income (welfare). Issues of distribution and allocation cannot, in general, be separated.

6

EXTERNALITIES AND THE ENVIRONMENT

FOCUS QUESTIONS

1. What are externalities?

2. How do private markets respond to externalities? What are the limitations of these private remedies?

3. What are the principal ways by which the public sector attempts to deal with externalities? What are the advantages and disadvantages of these alternative approaches?

4. What currently are the major environmental public policy issues? What policies regarding these issues have succeeded and what policies have failed? What are some of the current controversies in environmental public policy, and what insights does economic analysis provide into these controversies?

The federal government has long had an interest in environmental policy—the earliest federal action appears to have been the Refuse Act of 1899, designed to rid navigable waters of debris. The modern era of federal environmental regulation began with the Water Pollution Control Act of 1948, the first of a series of laws to protect the water we drink and the lakes and rivers in which we swim and fish.

Government activity on behalf of the environment has clearly had some beneficial effect. The quality of air in major industrial cities such as Gary, Indiana and Pittsburgh has improved noticeably since passage of the Clean Air Act of 1963. Lakes such as Lake Erie, which once faced the prospect of becoming so polluted that much marine life would be extinguished, have been saved.

Although it is increasingly agreed that government actions are required to preserve our environment, the extent and form of those actions remain a subject of debate. This chapter describes the economic rationale for government intervention in the environment and reviews the major government programs and policy issues related to environmental intervention.

THE PROBLEM OF EXTERNALITIES

Air and water pollution are two examples of a much broader range of phenomena that economists refer to as **externalities**, one of the market failures discussed in Chapter 4. Whenever an individual or firm undertakes an action that has an effect on another individual or firm for which the latter does not pay or is not paid, we say there is an externality. Markets affected by externalities result in inefficient resource allocations. Levels of production, as well as expenditures directed at controlling the externality, will be incorrect. For instance, consider a firm that could, by expending resources, reduce its level of pollution. Although there would be a large social benefit, there is no private incentive driving the firm to spend the money.

In some cases, the actions of an individual or firm confer (uncompensated) benefits on others; these are called *positive* externalities. A homeowner who maintains his or her property, including planting attractive flowers in front, provides a positive externality. Actions that adversely affect others are called *negative* externalities.

The level of production of negative externality-generating commodities will be excessive. Figure 6.1 shows conventional demand and supply curves. We argued earlier that, in the absence of externalities, the resulting market equilibrium, Q_m, was efficient. The demand curve reflected the individual's marginal benefits from the production of an extra unit of the commodity, and the supply curve reflected the marginal costs of

FIGURE 6.1

EXCESSIVE PRODUCTION OF GOODS YIELDING NEGATIVE EXTERNALITIES

The presence of a negative externality means that marginal social costs exceed marginal private costs, and the market equilibrium will entail an excessive production of the commodity. Q_m is market equilibrium, Q_e is the efficient level of output.
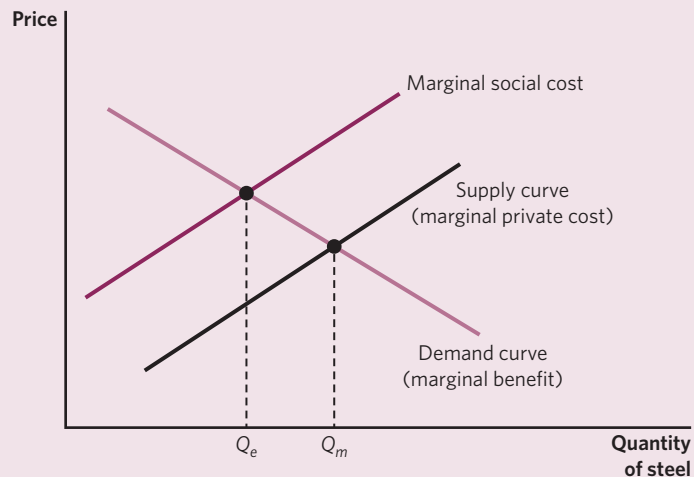
producing an extra unit of the commodity. At the intersection of the two curves, the marginal benefits just equal the marginal costs. Now, with externalities, the industry's supply curve will not reflect marginal *social* costs, only marginal private costs—those borne directly by the producers. If the expansion of steel production increases the level of pollution, there is a real cost to that expansion in addition to the costs of the iron ore, labor, coke, and limestone that go into the production of steel. However, the steel industry fails to take the cost of pollution into account. Thus, Figure 6.1 also shows the **marginal social cost curve**, giving the total extra costs (private and social) of producing an extra unit of steel. This cost curve lies above the industry supply curve. Efficiency requires that marginal social cost equal the marginal benefit of increasing output: production should occur at Q_e, the intersection of the marginal social cost curve and the demand curve. The efficient level of production is lower than the market equilibrium level.

An important class of externalities arises from what is referred to as **common resource problems**. Their central characteristic is that they pertain to a pool of scarce resources to which access is not restricted. Consider a lake in which the total number of fish caught increases with the number of fishing boats, but less than proportionately, so that the number of fish caught per boat decreases as the number of boats increases. Each additional boat reduces the catch of other boats. This is the externality. The marginal social benefit of an additional boat is thus less than the average catch of each boat, as shown in Figure 6.2; some of the fish that the additional boat catches would have been caught by some other boat.
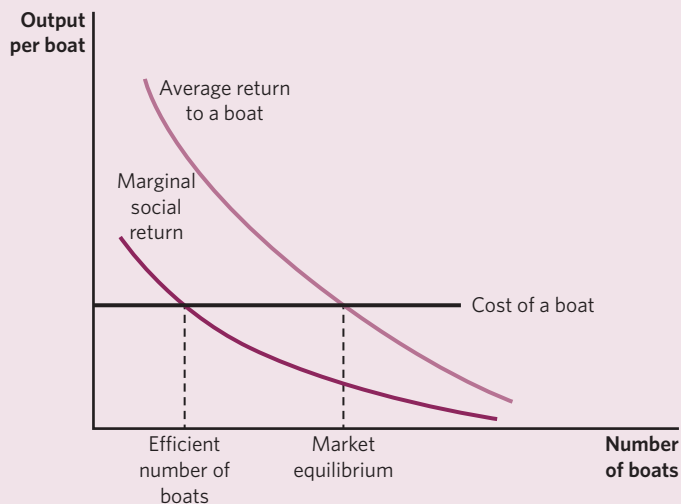
FIGURE 6.2

COMMON RESOURCE PROBLEM LEADS TO EXCESSIVE FISHING

The extra output of an additional boat is less than the average output. There will be an excessive number of boats.

<div style="border: box">

EXTERNALITIES

Externalities arise whenever an individual or firm undertakes an action that has an effect on another individual or firm for which the latter does not pay or is not paid. The consequences are:

1. Overproduction of goods generating negative externalities
2. Undersupply of goods generating positive externalities

</div>

The private return to an additional individual deciding whether to purchase a boat is simply the average return (once they are on the lake, all boats catch the same number of fish), which is much more than the marginal social return. Thus, whereas the private market equilibrium entails average returns equal to the cost of a boat (assumed to be constant), social efficiency requires that the marginal social return be equal to the cost of a boat.

In general, when there are externalities, the market equilibrium will not be efficient.

PRIVATE SOLUTIONS TO EXTERNALITIES

Under some circumstances, private markets can deal with externalities without government assistance.

INTERNALIZING EXTERNALITIES

The simplest way this can be done is by **internalizing externalities**—forming economic units of sufficient size so that most of the consequences of any action occur within the unit. Consider, for instance, any community, whether a group of neighboring houses or a set of apartments in the same or neighboring buildings. The quality of life in the neighborhood is affected by how each household maintains its property. If people plant flowers, they confer a positive externality; if they let their houses run down, they confer a negative externality. Even when each family owns its own apartment, the households may *collectively* decide that maintenance of the facilities that affect them all—including the external appearance—should be undertaken collectively. They form a cooperative or a condominium association.

There must be, of course, some way of enforcing the collective agreement that those who purchase a condominium or an apartment in a cooperative sign. A member of the condominium association might prefer to be a free rider, not paying his or her share of the cost of the maintenance of the common facilities; or the member might refuse to maintain his or her apartment in ways that are collectively agreed upon, and that may

adversely affect neighboring apartments. There must be recourse to the legal system, which ensures that the terms of the agreement—by which those living near each other attempt to deal with some of the externalities they impose on each other and to provide what are "public goods" to the group—are adhered to.

THE COASE THEOREM

As we have noted, externalities arise when individuals do not have to pay for the full consequences of their actions. There is excessive fishing in a common pool because individuals do not have to pay for the right to fish. Frequently, externalities can be dealt with by the appropriate assignment of **property rights**. Property rights assign to a particular individual the right to control some assets and to receive fees for the property's use.

Consider the problem of oil pools. Oil is usually found in large pools beneath the ground. To obtain access to a pool, all one needs to do is to buy enough land to drill a well and equipment for the drilling. The more oil that one well takes out of the pool, the less there is for others to take.[1] The total extra oil obtained as a result of drilling an extra well—the marginal social benefit—is thus less than the amount obtained by the additional well. Too many wells will be drilled.

The reason for this is that no one has the property right to the entire pool of oil. When the oil pool is controlled by a single individual, that individual has an incentive to make sure that the correct number of wells is drilled. Because economic efficiency is enhanced by having a single firm control the entire pool, any firm could buy the land over the pool from its present owners (at what they would have received from selling the oil) and wind up with a profit. In this view, no outside intervention would be required to ensure that an efficient pattern of property rights emerged.

Even when property rights for a common resource are not assigned to a single individual, the market may find an efficient way of dealing with the externality. Owners of oil wells frequently get together to **unitize** their production, thus making it less likely that too many wells will be drilled.[2] Fishermen using the same grounds may get together to devise mutually agreed-upon restrictions to prevent excess fishing.

The assertion that whenever there are externalities the parties involved can get together and make some set of arrangements by which

---

[1] There is another externality: as oil is removed, the costs of pumping out additional oil rise because underground pressure is reduced. Additional wells may actually reduce the total amount that will be extracted.
[2] Under unitization, the development of an oil or gas reservoir is put under a single management, with proceeds distributed according to a formula specified in the unitization agreement. This unitization is not done to reduce competition (it occurs even among small oil companies that take the price of oil as given, unaffected by their actions), but to increase efficiency.

the externality is internalized and efficiency is ensured is referred to as the **Coase theorem**.[3]

For instance, when there are smokers and nonsmokers in the same room, if the loss to the nonsmokers exceeds the gains to the smokers, the nonsmokers might get together and "bribe" (or, as economists like to say, "compensate") the smokers not to smoke. Or, say the smokers are in a nonsmoking compartment of a train, and the restriction on smoking (which can be viewed as an externality imposed on the smokers by the nonsmokers) takes away more from their welfare than the nonsmokers gain. Then the smokers might get together and "compensate" the nonsmokers in order to allow themselves to smoke.

Of course, the determination of who compensates whom makes a great deal of difference to the distributive implications of the externality. Smokers are clearly better off in the regime in which smoking is allowed unless they are paid not to smoke, compared to the regime in which smoking is banned unless they compensate nonsmokers.

USING THE LEGAL SYSTEM

Even when property rights are not perfectly defined, the legal system can provide protections against externalities. Our system of common law does not allow one party to injure another, and "injury" has been interpreted to include a variety of economic costs imposed on others. Implicitly, courts have given individuals some property rights—say, in the waters that they rely on for fishing. Those who have been injured have increasingly turned to courts to enforce those property rights.

When the Exxon tanker *Valdez* spilled oil into Alaska's Prince William Sound in 1989, those damaged by the spill—the fishermen whose catch was diminished, as well as those in the tourist industry who depend on sports fishermen—successfully sued Exxon. Many Americans believed that by spoiling one of the relatively pristine environments in the country, the spill hurt them too. They valued the *existence* of these natural resources, even if they did not immediately enjoy the benefits by visiting Alaska; to that extent, the *Valdez* oil spill had an externality effect upon them too. Courts have recognized these **existence values**—in the *Valdez* case, the state of Alaska, acting as trustee, collected more than a billion dollars in compensation.

Similar issues are now being litigated after the April 2010 explosion and fire on the BP-licensed Transocean drilling rig Deepwater Horizon

---

[3] R. H. Coase, "The Problem of Social Cost," *Journal of Law and Economics* 3 (1960): 1–44.

in the Gulf of Mexico and the subsequent worst oil spill in U.S. history. Although BP set up an independently administered $20 billion escrow fund while oil was still gushing into the gulf, the U.S. government saw this as a down payment toward compensation for victims of the oil spill. It filed suit in December 2010 against BP and several of its partners

THE EXXON *VALDEZ* OIL SPILL

Oil tankers have long been a major source of ocean pollution. The potential for damage was forcefully brought home with the grounding of the Exxon *Valdez* in Alaska's Prince William Sound in March 1989. Nightly pictures on TV depicted graphically the massive death of wildlife, including sea otters, salmon, birds, and seals. How long the devastation would last—or whether nature would ever fully recover—was not clear.

Exxon was made to pay more than $1 billion, most of which was to be spent on correcting the environmental damage; and the company claimed to have spent more than $2 billion beyond that in the months immediately after the spill, trying to limit the extent of damage. Even so, there was debate over whether the amount paid by Exxon was adequate: How much should Americans be compensated for the damage of the spill?

To answer this question, a study was done in which individuals were asked questions about how much they would be willing to pay to preserve a natural habitat, such as that which was harmed by the *Valdez* spill. Just as opinion polls, by sampling a thousand individuals, can provide an accurate forecast about how the entire population will vote, so, too, a sample of individuals can provide an accurate estimate for the value that would be assigned by the entire population. Some individuals will assign a high value, others a relatively low value, but these differences will be reflected in the sample. By projecting the distribution of values in the sample to the entire population, one can calculate the total value for the nation. In the case of the *Valdez* oil spill, the value estimated in this way

was about $3 billion. This methodology for assessing existence values is called *contingent valuation*.

The more fundamental question was how to prevent such disasters or, more accurately, how to make their occurrence less likely and the consequences less severe. As long as oil is transported, there is some risk of a spill, and no one has contemplated a complete ban on shipping oil. Shippers may not have the appropriate incentives to avoid a spill, though, because they do not bear the full consequences. This is a particularly severe problem, as many shipping companies are poorly capitalized, and in the event of an accident they would simply go bankrupt. Only a company as large and strong as Exxon could pay out $3 billion, yet almost any large oil tanker could do comparable damage.

To rectify this problem, Congress passed the Oil Pollution Act of 1990. This combined a system of incentives with regulations. Vessels had to be double-hulled, for example, thus reducing the likelihood of spillage.

One criticism that economists have raised is that the funds paid in compensation for damage, in general, must be used for cleanups. This constraint induces an inefficiency. The amount that the owners of vessels that have spills are required to pay should be designed to provide the corrective incentive to avoid spills. This may be more than the amount that is appropriately spent on cleanup. For instance, suppose that the consequences of a spill would be rectified by nature on its own in a year, and that it would cost an enormous amount to speed up the restoration process. We still would want to penalize firms that spill, but we might not want to spend the money to speed the restoration, as there would be little benefit to doing so.

over violation of safety regulations, seeking unlimited damages to cover the cost of the oil cleanup, long-term environmental damage, and local business losses.

To reduce the uncertainty about these often imperfectly defined property rights, government has tried to clarify them and to specify more precisely the nature and amount of damages that can be collected. Thus, more recent legislation and regulation have recognized the importance of existence values; the government—as "trustee" for the country's natural resources—has the right to sue for damages, although under current legislation the amount recovered must be used for restoration.

FAILURES OF PRIVATE SOLUTIONS

If the arguments asserting that private markets can internalize externalities are correct, is there any need for government intervention, other than to establish clear property rights? Furthermore, if these arguments are correct, why have cooperative agreements failed to take care of so many externalities?

There are several reasons why government intervention is required. The first has to do with the public goods problem discussed in Chapter 5. Many (but far from all) externalities involve the provision of a public good, such as clean air or clean water; in particular, it may be very costly to exclude anyone from enjoying the benefits of these goods. If nonsmokers get together to compensate smokers for not smoking, it pays any individual nonsmoker to claim that he or she is almost indifferent to letting others smoke. This individual will attempt to be a free rider on the efforts of other nonsmokers to induce the smokers not to smoke.

The problems of arriving voluntarily at an efficient solution are exacerbated by the presence of imperfect information. Smokers will try to persuade nonsmokers that they require a lot of compensation to induce them not to smoke. In any such bargaining situation, one party may risk the possibility of not arriving at a mutually advantageous agreement in order to get more out of any bargain that might be made.

Problems may arise even when markets are well established. Consider the problem of an oil pool, the land above which is owned by several individuals. Efficiency can be obtained by bringing all the land covering an oil pool under a single unitized management and control—called unitization.

However, if all but one of the landowners unitize, it may not pay for the last owner to join. This landowner knows that production on the unitized portion will be reduced, thus enabling him or her to increase production. The last owner will join only if he or she receives more than a proportionate share of the revenues. Then, all small owners may believe they can gain by holding out to be the last to join the unitization agreement (or selling to a large firm attempting to purchase all the small owners). States have therefore found it necessary to pass legislation requiring unitization.

> ## FAILURES OF PRIVATE REMEDIES FOR EXTERNALITIES
>
> - Public good (free rider) problems
> - Compounded by imperfect information problems
>   - How much does the individual need to be compensated for externality?
>   - Incentive not to reveal truth
> - Transactions costs
> - Additional problems with litigation
>   - Uncertainty about outcomes
>   - Differential access

Another reason for government intervention concerns transactions costs. The costs of getting individuals together to internalize externalities voluntarily are significant. The provision of those organizational services itself is a public good. Indeed, the government may be looked upon as precisely the mechanism that individuals have set up to reduce the welfare losses from externalities.

Transactions costs are a major disadvantage of dealing with externalities through judicial processes. For many externalities, the losses involved may simply be too small to justify undertaking litigation. Because those generating externalities know that litigation is expensive, they may be inclined to generate their externality just up to the point at which it pays the injured party to sue—giving rise to considerable inefficiencies. One way of partially dealing with this is to charge anyone shown to have imposed an externality on another a multiple of the estimated value of the damages. However, this gives rise to a countervailing danger of unwarranted lawsuits, with defendants settling claims simply to avoid the enormous litigation costs.

Uncertainty about the extent of the injury frequently compounds the problem of transactions costs, and there is also some ambiguity about the outcome of most suits. If litigation costs are large, the uncertainty acts as a further deterrent to individuals contemplating using the court system to deal with externalities.

The high litigation costs and uncertain outcome of the litigation process imply that there is, in effect, differential access to legal remedies—poor people may not be able or willing to bear the risks of litigation—a situation that conflicts with our usual notion of justice in a democracy. Because the legal system and the other private methods of addressing externalities so often work so poorly and so inequitably, there has been increasing reliance on public remedies.

PUBLIC SECTOR SOLUTIONS TO EXTERNALITIES

Public sector solutions to environmental externalities fall into two broad categories: market-based solutions and direct regulation. Market-based solutions attempt to influence incentives to ensure economically efficient outcomes. For instance, fines for polluting can be used to present firms with the true social costs of their actions, thereby diminishing their incentive to pollute. By contrast, government has used direct regulations to limit externalities, as in the case of mandatory emissions standards for automobiles.

Before comparing the merits of these different approaches, we should first dispel the common fallacy that asserts that an individual or firm should never be allowed to impose a negative externality on others. For example, it is sometimes asserted that a firm should never be allowed to pollute the air and water. In the view of most economists, such absolutist positions make no sense. There is indeed a social cost associated with pollution (or any other negative externality), but the cost is not infinite; it is finite. There is some amount of money that people would be willing to receive in compensation for having to live in a community with dirtier air or dirtier water. Thus, we need to weigh the costs and benefits associated with pollution control, just as we need to weigh the costs and benefits associated with any other economic activity. The problem with the market is *not* that it results in pollution; there is, indeed, a socially efficient level

DOUBLE DIVIDEND

Some have argued that there is a double dividend from imposing taxes (or fines) on pollution. Not only does it discourage pollution, but it also raises revenue, so the government has to rely less on distorting taxes. Those who believe that the tax system is distorting—with taxes on capital discouraging savings and taxes on labor discouraging work—emphasize this double dividend. Not only will national output *as conventionally measured* go up, but *true* output—which takes account of the pollution and degradation of the environment—goes up even more than conventionally measured output, as the tax or fine discourages pollution.

of pollution. The problem rather is that firms fail to take into account the social costs associated with the externalities they impose—in this case, pollution—and as a result, the level of pollution is likely to be excessively high. The task of the government is to help the private sector achieve the socially efficient level of pollution, to make individuals and firms act in such a way that they are induced to take into account the effects of their actions on others.

In the ensuing discussion the focus will be on pollution externalities. The arguments, however, extend in a straightforward way to other categories of externalities.

MARKET-BASED SOLUTIONS

Even when markets themselves do not lead to efficient resource allocations—as when there are externalities—market-like mechanisms can often be used to ensure efficient behavior. Market-based solutions to environmental externalities take three forms: fines and taxes, subsidies for pollution abatement, and marketable permits. We now consider each of these solutions in turn.

FINES AND TAXES   The simplest form of market-based solution involves levying fees or taxes in proportion to the amount of pollution emitted. In general, whenever there is an externality, there is a difference between the social cost and the private cost, and between the social benefit and private benefit. A properly calculated fine or tax presents the individual or firm with the true social costs and benefits of its actions. Fines of this sort—designed to make marginal private costs equal marginal social costs, and marginal private benefits equal to marginal social benefits—are called **corrective taxes**, or sometimes **Pigouvian taxes**, after A. C. Pigou, a great English economist of the first half of the twentieth century.[4]

Consider the example, discussed earlier, of steel producers polluting the air. We showed that because firms were concerned only with private marginal costs, not the social marginal costs (the two differing by the marginal costs of pollution), the output of steel would be excessive. By charging each firm an amount equal to the marginal cost of pollution, though, the marginal private costs and marginal social costs are equated.

---

[4] Pigou argued persuasively for the use of corrective taxes in his book, *The Economics of Welfare* (London: Macmillan, 1918).

MARKET EQUILIBRIUM WITH AND WITHOUT FINES

In the absence of a tax on pollution, firms will set price equal to marginal private cost. There will be excessive production (Q_m). By setting a tax equal to the marginal pollution cost, efficiency is obtained.
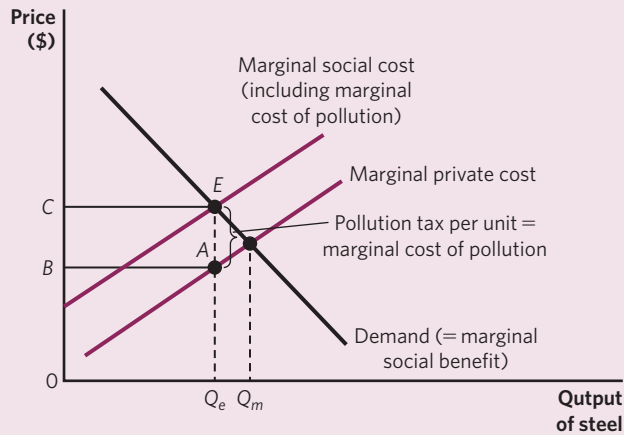
Figure 6.3

In Figure 6.3 we have assumed that the amount of pollution is proportional to the level of output, and the marginal cost of each unit of pollution is fixed; hence, by imposing a fixed charge per unit of output, equal to the marginal social cost of pollution, each firm will be induced to produce the socially efficient level of output. In the figure, the distance *EA* represents the pollution tax per unit output, and the area *EABC* represents the total pollution taxes paid.

Firms can reduce pollution by producing less, or by changing production methods. Changes in production methods may entail direct expenditures for pollution control devices, or changes in the input mixes and other alterations in the production process. Fines related directly to the amount of pollution ensure that firms will undertake the pollution abatement in the least costly—most efficient—manner possible. Assume that there is a given, known marginal social cost imposed on others by each unit of pollution (measured, say, by the number of particles added to the air per unit of time). It is costly to reduce pollution—and we assume that at any given level of production, it costs more to reduce pollution more. In other words, the *marginal* cost of pollution control is rising. This is depicted in Figure 6.4, in which we measure along the horizontal axis the *reduction* in pollution (from what it would be if the firm spent nothing on pollution abatement). Efficiency requires that the marginal social benefits associated with further pollution abatement expenditures just equal the marginal social costs, point P^* in the diagram. If the firm is charged a fine, f^*, equal to the marginal social cost of pollution, the firm will undertake the efficient level of expenditure on pollution abatement.

SUBSIDIZING POLLUTION ABATEMENT   Because a firm is likely to
receive a negligible direct benefit from pollution abatement (most of the
benefits accruing to those who live in the vicinity of the plant), absent
a fine on pollution, it has little incentive to spend money on pollution
abatement. There is, from a social point of view, too little expenditure
on pollution abatement. Rather than taxing pollution, the government
could subsidize pollution abatement expenditures. By providing a sub-
sidy equal to the difference between the marginal social benefit of pollu-
tion abatement and the firm's marginal private benefit, the efficient level
of pollution abatement expenditures can be attained. This is illustrated
in Figure 6.5. The marginal cost of pollution depicted in Figure 6.4 is
directly related to the marginal benefit of pollution abatement depicted
in Figure 6.5. Whereas in Figure 6.4 we assumed a fixed marginal social
cost of pollution, and hence a fixed marginal social benefit from pollution
abatement, in Figure 6.5, as pollution decreases, the marginal social bene-
fit from further pollution abatement decreases. Similarly, in Figure 6.4 we
have assumed rising costs of pollution abatement, whereas in Figure 6.5
marginal costs are constant. Either case may hold in a real situation.

    This remedy, however, does not attain a socially efficient resource allo-
cation. The reason is simple: the total marginal social costs of producing
steel include the costs of the government subsidies for pollution abatement.
Firms fail to take this into account in deciding on the level of production.
Thus, as before, the marginal social cost of steel production exceeds the
marginal private cost. The pollution abatement subsidy reduces the mar-
ginal social cost of output (from the dashed line to the solid black line
in Figure 6.6). But it also reduces the marginal private costs. There is
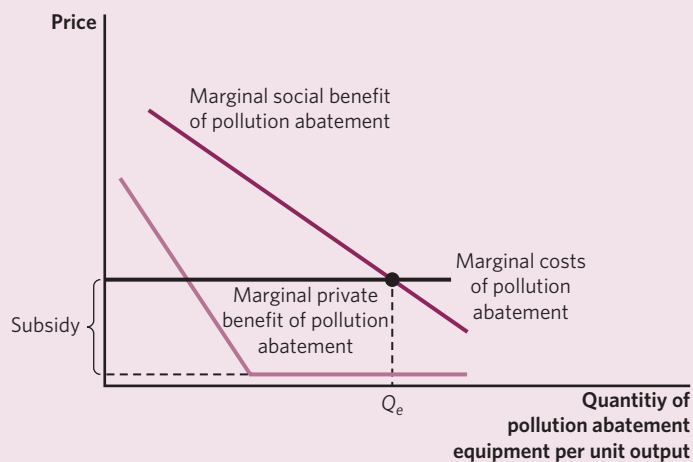
still an excessive level of production of steel, as illustrated by point Q_s in Figure 6.6.[5] Q_m is the output before subsidy, which is markedly greater than Q_e, the efficient level of output with the subsidy. A well-designed subsidy lowers the total marginal social cost of production—there is less pollution. Although there is a cost of the abatement subsidy, including the distortions arising from the taxes required to raise the revenues to finance it, the benefits from the lower pollution exceed these costs; hence, the optimal level of output with the subsidy is greater than the optimal level of output when firms have no incentives to reduce pollution. Thus, Q_o is greater than Q_e. On the other hand, if the pollution abatement equipment confers some ancillary benefits to the firm, it may simultaneously reduce the firm's marginal private costs of production, as indicated by the light gray line. Thus, the firm's output level also increases from Q_m to Q_s. However, because the main benefit of the pollution abatement equipment is to reduce pollution, presumably, the distortion—the magnitude of the excess production—is reduced.

The reason why polluters prefer subsidies over fines for pollution abatement is clear: profits are higher under the former system than under the latter. The distributional consequences are not limited to the polluting firms and their shareholders. Because output will be smaller under the system of fines, prices will be higher, and consumers of the polluting firm's products will be worse off. On the other hand, those who have to

---

[5] If the level of pollution of a firm cannot be directly monitored, a desirable policy would entail a subsidy for expenditures on pollution abatement combined with a tax on output. The tax on output, if set at the appropriate rate, reduces the level of output to the socially efficient level.
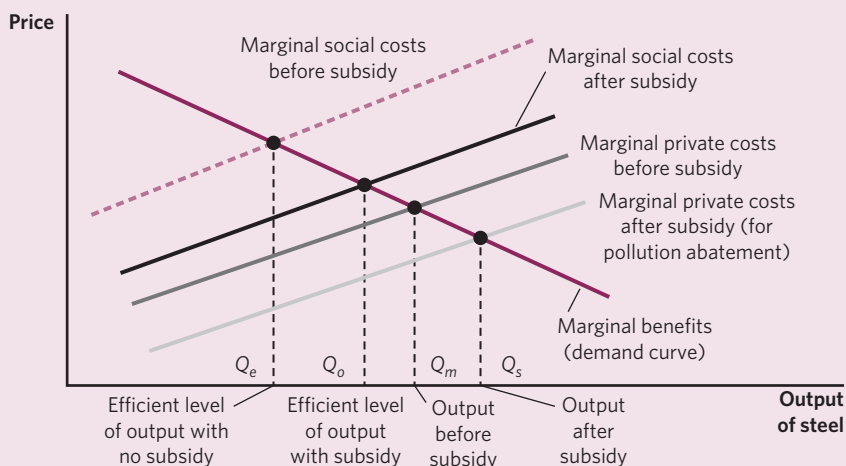
MARKET EQUILIBRIUM WITH POLLUTION ABATEMENT SUBSIDIES
Even after the pollution abatement subsidy, the equilibrium level of output of steel is still inefficient; the firm fails to take into account the extra costs of public subsidies for pollution abatement associated with increased output of steel as well as the marginal social cost of any remaining pollution.

pay the taxes to finance the subsidies for pollution abatement are clearly better off under the system of fines. It should be emphasized, however, that the choice between subsidies and fines is not just a distribution issue. As we have seen, under the pollution abatement subsidy scheme, producers do not face the true social cost of their production; there is an inefficiency. By contrast, with an appropriately designed system of fines, producers do face the true social costs.

MARKETABLE PERMITS   An increasingly popular market-based solution involves **marketable permits**, commonly referred to as tradable permits, which operate under a **cap and trade** system. A limit, or cap, is placed on the total amount of a pollutant that may be emitted, and this limit is either allocated or sold to firms in the form of emissions permits. These limit the amount of pollution that any single firm may emit. For instance, each firm may be allowed to emit 90 percent of the amount it emitted the previous year. Thus, a firm is granted a permit to emit so many units of pollutants. Because what the government cares about is the *total* amount of emission reduction, it allows firms to trade permits. A company that cuts its emissions in half could sell some of its permits to another company that wants to expand production (and hence increase its emission of pollutants).

Under this system, firms will be willing to sell permits as long as the market price of the permit is greater than the marginal cost of reducing pollution, and firms will be willing to buy permits as long as the marginal cost of reducing pollution is greater than the market price of the permit.

Thus, in equilibrium, each firm will reduce pollution to a level such that the marginal cost of pollution reduction is equal to the market price of the permit. Like fines, marketable permits use the market mechanism to ensure economic efficiency in the reduction of pollution: the marginal cost of reducing pollution is the same for every firm.

In the absence of uncertainty, the two systems are essentially equivalent: setting a price (say, for emissions) leads to a particular quantity, and setting a quantity leads to the corresponding price. In the presence of uncertainty, both about the costs of pollution and of pollution abatement, the two may not be fully equivalent. In setting a price, the quantity of pollution that emerges will be variable; in setting a quantity, the price that polluters will have to pay is uncertain.

Consider the problem of global warming, with which we know that there is a high social cost of carbon emissions, but we are uncertain exactly what that cost is. At the same time, we are uncertain about how firms and consumers will respond to any price of emissions imposed. Those who feel confident that we know much more about what we need to do—by how much emissions should be reduced—argue for marketable permits and for forcing firms to bear the risk of the uncertain price for emissions that will emerge. Advocates of a carbon tax (a price system) argue that, in practice, adjustments will have to be made to either prices or quantities over time. What matters is the level of concentration of carbon dioxide in the atmosphere, and we can adjust the flow (the level of emissions) as we see the level of concentration build up and as we observe the consequences, *either* by adjusting the level of marketable permits or by adjusting the price of carbon.

There are three problems with tradable permits. The first is making the initial assignments. Even though the system of assigning firms a fraction of their current levels of pollution might at first glance seem reasonable, it causes a major equity problem: marketable permits are an asset that can be traded, so giving pollution permits is equivalent to giving away money. Why should firms that have been polluters in the past be entitled to receive a bigger gift from the government? There is even a perversity in doing so: "good" firms that have spent large amounts on pollution control are given fewer permits and, if they have already installed state-of-the-art technology, will have a harder time reducing pollution. An alternative system bases permits simply on the level of production. This system is basically the one used when tradable permits were introduced to control acid rain. When Los Angeles introduced tradable permits in 1994, its assignments took into account both the levels of output and pollution and the state of the firm's current technology. There is a third alternative, particularly attractive in an era of budget stringency: auction off the emission permits. Firms that have been polluting will be worse off than before, but

the question is: Why should they receive money for not harming others? (Remember the reason that pollution imposes costs on others.) As we have noted, with such auctions there is little difference between marketable permits and taxes/fines.

This leads to the second problem, which traditionally was viewed as an advantage of tradeable permits: firms could be "bribed" to support the legislation curbing emissions, because they have less (or little) to lose. Indeed, the firms that believe that they can reduce emissions considerably become supporters, because selling their emission permits can be a new source of revenue. In Europe, though, we have seen the downside: governments come under pressure to issue more emission permits (it is like giving away money, without a budget constraint). The result is less reduction in pollution.

The third problem is more subtle. Tradable permits work well only when the location of the pollutant makes no difference. In many situations this is not the case, thus air pollution is much more serious near large cities. Moreover, with prevailing winds blowing from west to east, pollution along the East Coast may not be much of a problem, as most of it gets blown immediately out to sea; but pollution in the Midwest may have adverse effects on all the eastern states. The marketable permits that were introduced to control acid rain did not fully take this problem into account.

REGULATION

Economists have usually argued that market-based solutions provide the most promise for curbing environmental externalities, but government traditionally has relied on direct regulation. It has set emission standards for automobiles; put forth detailed regulations relating to the disposal of toxic chemicals; outlawed smoking on domestic airline flights; imposed laws requiring oil companies with wells in the same oil pool to unitize their production; imposed restrictions on fishing and hunting to reduce the inefficiencies associated with excessive utilization of these common resources. These examples illustrate the myriad forms that regulation may take.

Advocates of regulations argue that they provide greater certainty: if firms are prohibited from emitting more than a given level of pollution into the water, then one knows the maximum level of pollution; with fines, the level of pollution depends on the costs of reducing the pollution level. However, advocates of fines argue that one can easily adjust fines to induce firms to lower pollution to the desired level. Moreover, marketable permits provide a market-based way to attain efficient pollution reduction

and certainty of outcome. Indeed, a major criticism of regulations is that they do not reduce pollution in the most efficient way: different firms may face different marginal costs of further pollution abatement. Furthermore, regulations typically provide little or no incentive for firms to reduce pollution below the standard that has been set, regardless of how low the cost of doing so.

In the case of pollution, we should distinguish between two important classes of regulations. Recall that the market-based mechanisms discussed earlier focus on the *amount* of pollution; to pollute more, a firm must pay more in fines or buy more permits. This is a performance-based system, since the government only cares about the final outcome—how much pollution is produced. There are many **performance-based regulations**, such as regulations on automobile emissions, that also focus on the final outcome. However, much of pollution regulation has focused on standards, practices, and inputs, rather than performance. For instance, the government may prohibit the use of certain grades of coal, or it may require firms to employ scrubbers and other pollution abatement devices, or to construct smokestacks to specific heights. These are called **input regulations**. Market-based mechanisms may also focus on inputs and practices in this way; for example, a tax may be levied on high-sulfur coal, rather than on the pollution emitted.

When feasible, it is preferable to focus on performance, either for regulations or for market-based mechanisms. The one argument for focusing on inputs and practices is that they may be more easily monitored. Thus, it may be difficult to measure the amount of pollution coming out of a smokestack, but it is certain that if scrubbers (devices that reduce the amount of sulfur being emitted by a coal-burning electric power plant) are used, the amount being emitted will be less than if scrubbers are not used.

Although there may be good reasons for these policies, in some cases, politics rather than policy has dominated the decision. In the case of coal, had a performance-based standard been used, eastern coal producers would have been disadvantaged relative to western coal producers, because eastern coal contains more sulfur. To attain the same level of sulfur, firms using eastern coal would have had to use scrubbers, whereas those using western coal would not. Eastern coal producers successfully lobbied for the universal imposition of the requirement to use scrubbers.

INNOVATION

One of the reasons for performance-based regulations (as opposed to input standards) and pollution-based taxes (as opposed, say, to subsidies

for particular forms of abatement equipment) is that they directly address what is of concern—the level of pollution—and they may induce innovations, such as new ways of producing that generate less pollution or new techniques for abating pollution at lower costs. Advances in technology have improved the ability to monitor some kinds of pollution; ongoing innovation should enhance this further.

There has been considerable controversy over the best way to stimulate innovation and about the scope for innovation. Some environmentalists are less convinced than economists of the power of normal economic forces. Many believe that industry must be *forced* to innovate. Thus, by imposing extremely stringent standards—for instance, that cars get at least 40 miles per gallon—they will force industry to develop a product meeting these standards. Implicitly, they believe that the benefits of the innovation would outweigh the costs, but that the incentives that could be provided by the price system—charging car companies taxes in proportion to the amount of pollution—simply do not suffice to warrant their attention on this area.

In practice, the success of this strategy has been mixed. In some cases, rather than inducing innovation, stringent regulations have induced litigation: it may appear cheaper to a firm to try to persuade a court that the regulation is unreasonable than to spend the money to meet the standards imposed by the regulation. In some cases, firms have played a game of chicken, gambling that if they fail to meet the standards, the government will not shut them down, for fear of a political backlash from workers who are put out of a job. In some cases, however, industry unity has been broken by an innovative firm that showed that the standards are indeed attainable, or can even be surpassed. For instance, the great commercial success of the Toyota Prius has spurred its American competitors to develop hybrid cars of their own to compete in a rapidly growing market.

Environmentalists who doubt the effectiveness of market incentives by themselves in inducing innovation often point to the large gap between best practices—which often seem to be the most cost-effective practices—and what actually occurs. They point out, for instance, that there are energy-efficient light bulbs that more than pay for their higher costs in terms of reduced usage costs. Sometimes there is a coordination failure that government action can help remedy: no one wants to install fixtures that use energy-efficient light bulbs if it is going to be difficult to replace the bulbs when they burn out, and stores will not carry such light bulbs if there is no demand; and there won't be a demand if builders do not put them in houses they construct and lamp manufacturers do not put them in the lamps they make. Sometimes there are information barriers and other types of barriers to the adoption of cost-efficient energy-saving

technologies, and government can help promote the diffusion of these technologies by disseminating information.

Behavioral economics has explained how sometimes there can be inertia in changing customs. Government can help change norms, though. There have been large changes, for instance, in norms concerning smoking or the use of plastic bags, induced in part by government regulations. Of course, prices too can help in changing practices. One of the reasons that energy-efficient light bulbs are more prevalent in Japan than in the United States is that the price of electricity is higher there. Between these two camps are those who argue that simply providing information and small price signals will not lead to the large changes in behavior that are needed; changes are likely to be modest and slow. Today, labels on many electric products identify energy usage and costs, helping buyers make more intelligent decisions about lifetime costs of different products.

Thus, there may be win–win regulations with which efficiency is enhanced at the same time that environmental costs, especially those associated with the use of energy, are reduced; slight modifications in construction practices—the color of roof shingles or the planting of trees, for example—can have a noticeable effect on energy consumption.

Critics of approaches focusing on inputs rather than performance argue that such approaches are not only inefficient, but also stifle innovation and push it in the wrong direction. For instance, rather than seeking the most effective way of reducing emissions from coal-burning power plants, research is focused on making cheaper scrubbers. Moreover, research directed at improving the ability to monitor outputs accurately—thus reducing the necessity of relying on input regulations—is not encouraged.

INFORMATION DISCLOSURE

In some areas, governments have been experimenting with another approach, focusing on public pressure rather than the heavy hand of government. Government's role would be limited to requiring firms to disclose, for instance, the potentially cancer-inducing chemicals that they discharge into the water or emit into the air. Government would not even comment on the extent of scientific evidence concerning the impact of the chemicals on humans. Critics of this approach often argue that the costs of such information disclosures can be high, but their real concern is that government would be encouraging a scare campaign. Most people would simply assume that if a chemical is listed as dangerous, it must be dangerous—or, in any case, why risk it? People in the neighborhood would

put enormous pressure on the firm to eliminate the chemical, without any assessment of the costs or benefits of doing so. There could be enormous adverse economic effects. Particularly troubling is the evidence of one study by the Environmental Protection Agency (EPA) that showed the magnitude of popular misconceptions about environmental risks. Scientists and nonscientists were asked to rank a number of different potential environmental health hazards, and there was little correlation between the two. Among the risks rated most highly by the nonscientists were several that were ranked at the bottom by the scientists, and vice versa.

COMPENSATION AND DISTRIBUTION

So far most of our discussion has focused on the *efficiency* of alternative ways of controlling pollution (externalities). But much of the debate is about *distribution*—who bears the costs. Different systems of controlling pollution may have markedly different distributive consequences. Subsidies for pollution abatement equipment may result in a less efficient resource allocation than a system of fines for polluting; but firms will clearly prefer subsidies. Greater efficiency means that, in principle, the overall gains to society from using the more efficient system are such that the gainers could compensate the losers.

Why then do governments so often resort to inefficient systems like abatement subsidies? The reason is that the compensation is typically not paid. Partly this is because it is often difficult to measure the gains and losses to each individual—the information required to implement the desired compensations is simply not available—and partly it is because those who benefit from the inefficient system are more politically organized. The losers from a system of fines are clear—both the owners of the polluting firms,

ASSESSING ALTERNATIVE APPROACHES TO CONTROLLING EXTERNALITIES

Performance-based versus input-based
- Performance-based is more efficient when performance can be measured.
- Costs of monitoring inputs may be lower.

Regulation versus fines
- Fines related to costs of pollution provide appropriate incentives.
- However, there may be greater uncertainty about the actual level of pollution.
- Regulations provide greater certainty, strong incentives to meet the regulatory standards, but no incentive to reduce pollution below the standard, regardless of how low the cost is to do so.

Marketable permits versus fines
- Both can result in efficient reductions in pollution levels.
- Marketable permits provide greater certainty about the level of pollution.
- Marketable permit systems face difficult problems in allocating initial permit rights.
- There are further problems if the costs of pollution depend on *where* the pollution occurs.

Subsidies for pollution reductions versus fines for pollution
- Both can induce reductions in pollution levels and even achieve efficient levels of pollution abatement. The level of production of the pollution-inducing industry will be too high, however, because the firm will not take into account full costs—including costs of pollution abatement. Firms prefer pollution abatement subsidies.

who see their profits decreased, and the consumers of their products, who see their prices increased. The gainers are more dispersed—all the taxpayers who bear the burden of the taxes used to pay the subsidies, and all the consumers of all the other products who might see their prices rise slightly as production shifts slightly toward the subsidized industry. As is so often the case, the losers are much easier to identify than the gainers; thus, it is much easier for the losers to get together and use the political process to argue for a system that, though inefficient, makes them bear less of the cost of reducing pollution.

One of the reasons that governments may resort to regulation rather than the market-based systems described earlier is that the distributive consequences may be less; for instance, individuals and firms may respond only to a limited extent to a small price of emission permits. To induce the large changes in behavior that are required to avoid global warming (say, to avoid an increase in the world's temperature by more than 2 degrees Celsius, a goal of the international environmental community) might require a very large increase in the price of emissions, which would translate into large increases in energy prices, with significant distributive consequences. It may be possible to "force" the adjustments the economy needs with less adverse distributional effects by imposing regulations on electricity producers and car manufacturers.

Recent discussions of environmental policy have also focused on the distributional consequences of environmental hazards. For instance, the poor are more likely to live near toxic waste sites, and exposure to these hazards imposes large costs not only on parents (on their health and life expectancy), but also on their children. There is a greater likelihood that their children will be born with low birth weight, which, in turn, is associated with lifelong consequences, including lower average lifetime incomes.

PROTECTING THE ENVIRONMENT: THE ROLE OF GOVERNMENT IN PRACTICE

We now look more closely at the actual policies that the government undertakes to protect the environment. For convenience, we divide them into three categories: those directed primarily at air, water, and land. There are, of course, important interactions among these pieces of our environment, so several of the policies affect two or more of these categories.

AIR

The air we breathe has been taken for granted since the beginning of time, but by the middle of the twentieth century this was no longer possible in many major cities. London became famous for its pea-soup fog generated by pollution; Los Angeles for its life-threatening smog; Gary and Pittsburgh for their brilliant red overcast skies, a product of the steel mills on which those cities' economy depended. People with weak respiratory systems knew the dangers of living in these cities, but the fact that *all* individuals faced greater health risks was recognized only slowly.

There are several aspects of the nation's attempt to control air pollution. Two have been marked by considerable success; there is heated controversy over a third one; and in the fourth, progress remains slight.

The most marked success is associated with ozone depletion and chlorofluorocarbons.

OZONE DEPLETION   The Earth's atmosphere has a thin layer of ozone, which shields us from harmful solar radiation. In the late 1980s it became clear that a hole was appearing in the ozone layer over Antarctica, and that the cause of the hole was chlorofluorocarbons (CFCs) and other ozone-depleting substances (ODSs). The nations of the world responded in 1987 with a treaty signed in Montreal, Canada, called, appropriately, the Montreal Protocol. It initially required the production and consumption levels of CFCs to be cut in half by 1999, but was strengthened in 1990 with a commitment to phase out CFCs and halons entirely by 2000 (by 2010 for less developed countries). Since then, ODS atmospheric concentration has steadily declined, although complete recovery of the Antarctic ozone layer is not expected until 2050 at the earliest.

ACID RAIN   The control of sulfur dioxide (SO_2), which gives rise to acid rain and is emitted especially by coal-burning power plants, is another success story. In the 1970s, we became aware that the leaves on the trees in many of our forests were turning yellow and many of our lakes seemed to be devoid of fish. The Acid Precipitation Act of 1980 and the Clean Air Act Amendments of 1990 began a program to control these emissions, and a national program of tradable permits was introduced. These are estimated to have significantly reduced the overall cost of bringing down the level of pollution. Regional cap and trade markets for nitrogen oxide (NOx)—another pollutant—have been established as well.

An especially problematic aspect of acid rain is that SO_2 and NOx emissions cross state lines. In 2011, the U.S. Environmental Protection Agency issued the Cross-State Air Pollution Rule (CSAPR) to strengthen

its court-challenged legal authority to regulate interstate acid rain emissions. The CSAPR focuses on pollution in the eastern United States, with the goal of reducing power plant SO_2 emissions by 73 percent and NOx emissions by 54 percent from 2005 levels by 2014 in affected regions.[6]

PARTICULATE AIR POLLUTION   Small particulates in the air can pose serious health problems for certain individuals. The federal government's involvement in addressing the challenges of particulate air pollution has evolved over the past half-century from purely research (Air Pollution Control Act of 1955), to setting of standards (Clean Air Act of 1970), to enhanced authority for pollution control (1990 Clean Air Act Amendments). Although the EPA estimates that the direct benefits from the 1990 Clean Air Act Amendments will reach an annual economic value of almost $2 trillion by 2020, much less that the estimated $65 billion annual cost of public and private efforts to comply with the requirements of this law, there is still considerable debate about the assumptions underlying this cost–benefit analysis.[7] The controversy concerns the cold calculations of the costs of reducing the level of particulates versus the health benefits. As hard as it is to quantify these benefits, there are standard procedures by which this is done; such calculations are made routinely in evaluating how much to spend to make a safer highway, a safer car, or a safer airplane. These changes can result in a slightly smaller probability of an accident from which one can calculate, *on average*, how many lives will be saved. The government must have a systematic way of deciding whether the benefits exceed the costs, by doing so, it places a value on life—in this case, the value of reducing the risk of fine particle-related premature mortality.

GLOBAL WARMING   The one area in which little progress has been made is global warming and greenhouse gases. The Swedish chemist Svante Arrhenius explained as early as 1896 that carbon dioxide emissions from the burning of coal would enhance Earth's natural greenhouse effect (retention of solar radiation by Earth's atmosphere) and thus lead to global warming, but it was not until the 1980s that the world's attention turned to the issue. The scientific community was able to show that the current level of carbon dioxide concentration in the atmosphere due to the burning of carbon—from coal, gas, and oil—is, indeed, substantially greater than it was at the beginning of the Industrial Revolution, and

---

[6] U.S. Environmental Protection Agency, *Cross-State Air Pollution Rule (CSAPR)*, accessed July 20, 2011, http://www.epa.gov/crossstaterule.
[7] For the complete cost–benefit analysis, see U.S. Environmental Protection Agency, Office of Air and Radiation, *The Benefits and Costs of the Clear Air Act from 1990 to 2020* (Washington, DC: U.S. EPA, March 2011).

continues to grow. Scientists have also established that there was over-whelming evidence that these substantial increases were leading to significant increases in Earth's temperature. Since 1990, the Intergovern-mental Panel on Climate Change (IPCC) has issued periodic assessments on the state of global warming, and, together with former vice president Al Gore, was awarded the Nobel Peace Prize for contributions to interna-tional environmentalism.

To many, the world seems embarked on a risky experiment with our planet as we continue to add carbon dioxide and other greenhouse gases to the atmosphere; it is leading not just to a warming of Earth, but also a rise in the sea level. There will be severe adverse effects from this warm-ing, especially on the tropics, and the increase in sea level will obviously have adverse effects on low-lying islands and countries, such as Bangla-desh, Vietnam, and the Netherlands. Other predicted effects include an increase in the variability of weather. Even though there is clear evi-dence for the increase in greenhouse gases, and there is a general (but not universal) consensus on the long-run effects, there is more contro-versy over whether the effects are already being felt. There is evidence, for instance, of a marked increase in losses from weather—far greater than the increased losses from nonweather events such as earthquakes. Although there is some disagreement among economists about the mag-nitude of the overall costs—with some countries in cold climates actually benefiting—the consensus, reflected in an agreement made in Copenha-gen, is that the costs will be large.

As the scientific evidence has mounted, the issue has taken on greater urgency over the past two decades. In 1997 collective global concern resulted in adoption of the Kyoto Protocol, under which industrialized countries pledged to reduce their emissions of greenhouse gases. Most countries have ratified the Kyoto Protocol, with the notable exception of the United States. A key U.S. objection is the protocol's exclusion of devel-oping countries such as China and India, who will be the largest future sources of CO_2 emissions.

In response, developing countries have argued that most of the prob-lem has been caused by the profligacy of the advanced countries, their own contributions to the greenhouse gases thus far have been relatively small, their projected future contributions to greenhouse gases will still be much lower than high-income countries on a per capita basis, and they are too poor to devote much of their resources to reducing their emissions below the level that economics dictates.

To reduce overall costs, the countries agreed at Kyoto to explore more market-based mechanisms—tradable permits and a variant called "the clean development mechanism" or **joint implementation**—because

one country would "buy" the greenhouse gas reduction from another; in effect, by paying for it, they could be thought of as "jointly" implementing the greenhouse gas reduction. Joint implementation can be thought of as a limited form of marketable permits. Some critics suggested that the United States was advocating joint implementation not out of a commitment to economic efficiency, but because it could not or would not take measures that would reduce greenhouse gases within its own borders.

Today, there are many active cap and trade systems for greenhouse gases. The European Union Emission Trading Scheme (EU ETS) was established by the EU to help meet its Kyoto Protocol commitments by the time the protocol expired at the end of 2012, and is now the world's largest CO_2 emissions trading system.

Negotiations for a new global agreement on greenhouse gases are continuing, with very slow progress. In Copenhagen, countries agreed to draw up their own action plans to reduce emissions. There was peer pressure to make significant reductions, but no commitments were made. Subsequently, parties to the United Nation's Framework Convention on Climate Change concluded their December 2012 climate conference in Doha, Qatar, with agreement to continue negotiations on a successor to the Kyoto Protocol, which they hope to approve during their 2015 conference in France.

The easiest and most cost-effective way of reducing emissions in the United States would be a carbon or energy tax, but this option has been adamantly opposed by the powerful gas, coal, and oil interests. Currently, gasoline taxes in the United States are markedly lower than in Europe and most other industrialized countries, as shown in Table 6.1.

TABLE 6.1 GASOLINE TAXES AROUND THE WORLD

COUNTRY	TAX PER LITER IN U.S. DOLLARS
United Kingdom	1.32
Germany	1.26
France	1.19
Italy	1.14
Spain	0.90
Japan	0.77
Canada	0.38
United States	0.11

SOURCE: International Energy Agency, *End-Use Petroleum Product Prices and Average Crude Oil Import Costs*, Table 3 (Average end-use prices) and Table 5 (Average end-use taxes), March 2011.

A central problem in any international agreement is *enforcement.* Global warming is a global public good, and there is an incentive for each country to be a free rider; that is, each would like to continue to pollute, but also to enjoy the benefits of a world that is not subject to global warming as a result of the efforts of others. In the Montreal Protocol, discussed earlier, countries that did not comply were threatened with trade sanctions. Some have suggested that similar sanctions might be desirable, or even necessary, to deal with global warming. Indeed, many European countries that were making large efforts to reduce their emissions thought that the failure of the United States to do anything was giving U.S. firms an unfair competitive advantage. European steel producers, for instance, had to pay for their carbon emissions, while U.S. steel producers did not. It was as if the U.S. producers were being subsidized, because they did not bear the full costs of their production. Some have suggested, accordingly, that it is only appropriate that those who do impose a carbon price (either through a carbon tax or tradable permits) impose a tax on imports from the United States or any other country refusing to do similarly.

WATER

The debate over clean air today centers not around whether pollution should be controlled, but rather how and at what levels. Likewise, in the case of water, there is consensus that the controls that have been put in place for drinking water make sense, but controversy remains over the benefits relative to costs of stringent regulations attempting to reduce pollution in streams and rivers. Much of today's water pollution comes not from factories, which can be more easily controlled, but from difficult-to-control sources such as runoff from farms. Controlling such pollution would require controlling the use of fertilizers and pesticides. Although price mechanisms (taxes) might discourage the use of fertilizers and pesticides, it would be virtually impossible to differentiate between usages that contribute to pollution and those that do not. Moreover, there is controversy over some of the benefits. How worried should we be about the pollution of groundwater that will almost surely never be used for drinking and that is unlikely to seep into wells or springs? Some environmentalists believe that we should never spoil a part of our natural heritage; keeping groundwater clean has nothing to do with the use to which groundwater might be put. Others take a very risk-averse stance: How can we be sure that the groundwater will never be used for drinking? Controversy over these issues has prevented reauthorization of the Clean Water Act, although many limited, specifically targeted clean water bills have been enacted in recent years.

LAND

TOXIC WASTES   Newspapers have presented graphic stories of rivers, canals, and land that chemical companies have turned into toxic waste sites, subjecting those who come into contact with them on a regular basis to increased risk of cancer. Americans had nightmares about discovering that their homes had been built over toxic waste sites. In response, in 1980 Congress passed the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), commonly known as **Superfund**, after the trust fund it established with a tax on chemical and petroleum companies, to help clean up these sites.

The law was badly designed, however. It was based on the principle that those who contributed to the pollution should pay for its cleanup, but it provided that anyone who had contributed at all to the particular site was liable for the entire cleanup costs. This enabled the government to go after large corporations—those with "deep pockets"—forcing them to pay for the cleanup and letting them sue the other polluters to recover their shares. (This is called the **system of joint and several liability**.) Furthermore, it provided for perpetual liability: even after the site was seemingly cleaned up, it was always possible that on the site another chemical with adverse effects would be discovered, so the guilty party is never completely free from liability. To make matters worse, the polluters' insurance companies argued that their general liability policies did not cover pollution other than that which originated as a result of an accident. In several states, the courts supported the insurance companies, whereas in other states, the courts said that the insurance companies were liable. The upshot was litigation—between the government and the polluters, among the different polluters, and between the polluters and their insurance companies. The lawyers made out like bandits, but the toxic wastes did not get cleaned up. Over 70 percent of insurance company expenditures went to legal fees, and more than a quarter of what the polluters spent went to lawyers. Worse still, property owners had a new nightmare to worry about: that toxic waste would be discovered on their property that might not hurt their health but would definitely hurt their pocketbooks. They would be responsible for cleaning it up, and the EPA often set standards for which the costs simply could not be justified by the benefits. In some places, the result was "gridlock" in the land market: no one would buy potentially polluted sites, which hampered efforts to redevelop inner-city areas. Banks would not make loans, lest they wind up holding the property (and being responsible for the cleanup) in case of a default. America's landscape was scarred with such "brownfields" (as they came to be called), and America's firms had to get out to find green fields to

build their new factories. A law intended to preserve and protect the land had led to opposite results.

Thirteen years after the bill was passed, only one of seven major sites that had been identified had been cleaned up. Four years later, the administration was claiming that two-thirds of the sites had been, or were on the way to being, cleaned up. The situation was similar in 2010, when more than half the work had not yet been completed for the remedial construction phase of cleanup for over 60 percent of the 239 nonfederal National Priorities List (NPL) sites with unacceptable or unknown human exposure.

Despite this modest progress over the thirty years since enactment of the Superfund legislation, the other problems, such as the brownfields and the inequities associated with joint and several liability, persisted. Even though there was universal agreement that reforms were needed, the differing perspectives of environmentalists, insurance companies, and polluters made resolution of these problems difficult. There were conflicts about both the standards of cleanup and about who should pay. Although forcing those who pollute to pay provides strong incentives not to pollute, there was controversy over whether it was right to force people to pay for actions that were not illegal at the time they occurred, and whose consequences might not even have been apparent. Indeed, the worst effects of the powerful chemicals that dry cleaners had used was probably not on the land onto which some of these chemicals spilled but on the owners and workers who spent their lives working in these cleaning plants, unaware of the health effects. They had already paid a high price—and given the competitive nature of the industry, the benefits of their using the chemicals were received by their customers: had they been required to dispose of the chemicals in another way, they would have done so and passed the costs on to their customers. And in any event, in many cases, it was not the polluters that would actually pay but their insurance companies, thus undermining the moral argument that polluters should pay. However, environmentalists point out that firms are in the best position to judge the risks posed by the chemicals they use, and that insurance firms are in a good position to put pressure on those they insure to look carefully at the risks of their products. The debate continues, although the EPA claims, "Over the past 20+ years, we've located and analyzed tens of thousands of hazardous waste sites, protected people and the environment from contamination at the worst sites, and involved others in cleanup."[8]

[8] U.S. Environmental Protection Agency, *Superfund: Basic Information*, accessed July 20, 2011, http://www.epa.gov/superfund/about.htm.

ENDANGERED SPECIES   There has been concern not only to protect the environment from pollution, but also to preserve it. As populations expand, they crowd out nature. Throughout the world, a multitude of species are threatened with extinction. At the global level, international treaties have been signed to combat these threats; 1992 was marked not only by a treaty on global warming, but also by a treaty intended to preserve the world's biodiversity—plants as well as animals. Particularly powerful in this debate was the recognition that within this diverse biological heritage, there might live cures for a myriad of diseases. Other international treaties are directed at preventing the extinction of whales and eliminating trade in ivory (which might encourage the extinction of elephants) and rhinoceros horn (highly valued in certain parts of the world for its alleged powers in enhancing sexual potency).

In 1973, the United States passed the Endangered Species Act. The legislation has been highly controversial because of its potentially strong economic impact. For instance, logging in large parts of Washington and Oregon was halted because of a concern over the destruction of the habitat of the spotted owl, an endangered species; and in Texas, development of areas near Austin was halted over fear of destroying the habitat of some endangered species of spiders.

Critics argue that the preservation of these species is a public good, but a public good that owners of these particular parcels of land are made to pay for. If the public wants these species to be preserved, it should buy the land. Prohibiting owners from developing the land is almost tantamount to seizing it. Indeed, many argue that any restrictions on usage represent a "taking" of property; just as the government cannot simply take away your property without compensation, it should not be allowed to take away the uses to which you can put your property without compensation. There is a fundamental difference between laws that stop a person from imposing an externality on others, and laws that require a person to provide a public good (the protection of an endangered species) to others.

Supporters of endangered species legislation, even when they recognize these arguments, say that there simply is not enough money available to provide compensation to property owners; the choice is a pragmatic one—allow the species to become extinct, or impose these mandates on property owners. Besides, the longer the law is on the books, the less these arguments on "takings" become relevant: those who buy property know that their use may be encumbered by the Endangered Species Act, and this is reflected in the purchase price they pay. The cost was effectively borne by the owners of the land at the time the law was enacted; if anyone should be compensated, it is the former owners, not the current ones.

Protecting endangered species is only one of several pieces of legislation designed to protect our natural environment. There is also, for

instance, important legislation protecting the wetlands and coastal land. President Theodore Roosevelt protected 230 million acres of some of the country's most treasured areas while serving from 1901 to 1909 by establishing 5 national parks, 150 national forests, 4 national game preserves, and 51 federal bird reservations.

CONCLUDING REMARKS

Although the specifics of environmental legislation—how best to improve the environment and how high to set the standards—are likely to remain contentious, there is a growing awareness that the government has at its disposal a wide range of instruments and a growing consensus on a set of general principles: the environment is of critical importance; markets alone will not provide efficient outcomes because of important externalities; some form of government action is required; when possible, interventions should be performance based and market oriented; the government must be sensitive to the distributional consequences both of environmental degradation and the policies that are implemented to ensure the protections of the environment; and the environment is so important that we cannot make the perfect the enemy of the good.

REVIEW AND PRACTICE

SUMMARY

1. Externalities are actions of an individual or firm that have an effect on another individual or firm for which the latter does not pay or is not paid.

2. Sometimes economic efficiency can be attained without resorting to government intervention

   a. By establishing sufficiently large economic organizations, the externalities can be internalized.

   b. By establishing clear property rights, private parties can bargain toward an efficient solution, as suggested by Coase.

   c. By using the legal system, imposers of externalities can be forced to compensate victims.

3. There are important limitations to each of these private remedies. For instance, public goods problems and transactions costs impede efficient bargaining solutions in the manner suggested by Coase. These failures necessitate a greater role for government in remedying the problems of externalities.

4. There are four methods by which the government has attempted to induce individuals and firms to act in a socially efficient manner: fines and taxes, subsidies, tradable permits, and regulation.

5. When there is good information about the marginal social cost of the externality (as with pollution), and the fines can be adjusted to reflect those

costs, then a fine system can attain a Pareto efficient outcome. Subsidies to pollution abatement, while enabling the efficient level of pollution abatement to be attained, will result in excessive production of the pollution-generating commodity. In principle, the gainers under the fine system could more than compensate the losers, but in practice these compensations are seldom made. Thus, the choice of the system for controlling externalities has important distributional consequences.

6. Tradable permits (cap and trade systems) can also result in efficient pollution abatement.

7. Regulations focusing on inputs or standards are likely to result in inefficiency.

8. The Clean Air Act has greatly reduced the level of pollution in the air. There is increasing concern about greenhouse gas emissions, which may lead to global warming.

9. The Clean Water Act has greatly reduced water pollution. Controversy remains over whether standards are excessively stringent, so that at the margin, costs exceed benefits.

10. The Superfund program, which is intended to clean up toxic waste sites, faces several problems, including excessive litigation costs and slow cleanups. Remedies include reforms in the legal system and the cleanup standards, and must address the problem of who should bear the costs of cleanup.

11. There has been increasing interest in preserving biodiversity and protecting endangered species. There is concern, however, that restrictions on land usage required to protect endangered species constitute an unfair "taking" of property.

12. There are several global environmental problems, the most important of which is global warming. There is broad consensus among the scientific community that increases in the concentration of greenhouse gases in the atmosphere will lead to higher temperatures, an increase in sea level, and increased weather variability, with significant economic consequences. The international community has so far not been able to reach an agreement to curb emissions of carbon dioxide and other greenhouse gases.

KEY CONCEPTS

Cap and trade

Coase theorem

Common resource problems

Corrective taxes (Pigouvian taxes)

Existence values

Externalities

Input regulations

Internalizing externalities

Joint implementation

Marginal social cost curve

Marketable permits

Performance-based regulations

Property rights

Superfund

System of joint and several liability

Unitize

QUESTIONS AND PROBLEMS

1. Make a list of the positive and negative externalities that you generate or that affect you. For each, discuss the advantages and disadvantages of each of the remedies discussed in the text.

2. An important class of externalities to which attention has recently been directed is called *information externalities*. The information produced by one individual or firm generates benefits for others. The success of an oil well on one tract of land increases the likelihood of oil's being found on an adjacent tract, and hence increases the value of that tract. Can you think of other examples of information externalities? What are the likely consequences of information externalities for the efficiency of resource allocations? Discuss the possibilities of private market solutions to these problems.

3. Explain why subsidies for pollution abatement equipment, even if they result in an efficient level of pollution abatement, will not result in an efficient resource allocation.

4. Assume that there is uncertainty about the value of pollution control as a result of, for instance,

uncertainty about the costs of pollution. Draw two different "demand curves" (or benefit curves) for pollution, showing the marginal benefit of reducing pollution by one more unit decreasing as the level of pollution reduction increases. Assume that the marginal cost of pollution abatement increases as the level of pollution abatement increases.

a. Assume the government can regulate the amount of pollution control *after* it knows what the benefits are. Show what the level of pollution control will be in each situation.

b. Assume the government can impose a tax on pollution *after* it knows what the benefits are. Show what the level of tax (or fine) will be in each situation. Is there any difference between regulations and fines in these circumstances?

c. Now assume that the government must set the level of allowable pollution before it knows what the benefits are. How will it set the level of allowable pollution? (Hint: there is a cost to society due to allocative inefficiency: from allowing too much pollution, if it turns out the benefits of pollution reduction are high; from being too restrictive, if it turns out the benefits are low.) How do you minimize the sum of these two social welfare losses?

d. Now assume the government must set the level of fine before it knows what the benefits are. How will it set the fine? Is there a difference between fines and regulations here?

e. Now assume that marginal benefits of pollution abatement are unknown. Assume also that when marginal benefits are high, the marginal costs are also high, and similarly, costs are low when benefits are low. Contrast a system of fines and regulations under these circumstances, where the level of fine or regulation must be set before costs are known.

5. Assume there are two types of communities in the United States, those in which there is a high benefit of pollution control and a high cost of pollution control, and those in which there is a low benefit of pollution control and a low cost of pollution control. Assume that the government must set either uniform regulations (a uniform level of pollution control) or a uniform fine for pollution. Show diagrammatically that a regulatory scheme may be preferable to a system of fines. How does your answer change if communities in which there is a high marginal cost of pollution control happen to be communities in which there is a low marginal benefit; and communities with a low marginal cost of pollution control happen to be communities in which there is a high marginal benefit?

6. The impact of some externalities is very local, such as noise from airplanes landing and taking off at an airport. Such externalities depress the value of the immediately surrounding real estate. We say that the cost of the externality is *capitalized* in the value of the property. Assume that poor individuals are more willing to accept the high level of noise pollution, in return for the much lower rents they have to pay for housing. Describe the *incidence* of a regulation lowering the noise level surrounding the airport; that is, who benefits? (Hint: What will happen to land values? To rents?)

7. Zoning laws, which restrict how individuals can use their land, are sometimes justified as a means of controlling externalities. Explain. Discuss alternative solutions to these externalities.

8. What is the externality associated with an additional individual's driving on a congested road? How do tolls help alleviate this externality? How should the toll be set?

9. Explain why a system of joint implementation for reducing greenhouse gases is more efficient than a system whereby each country must reduce its pollution by a fixed amount.

10. Many economists are worried that unless all countries are required to reduce their levels of greenhouse gas emissions, reductions in emissions in one country may be partially offset by increases in another. Explain how this might occur.

11. Global warming is related to the concentration of greenhouse gases in the atmosphere. Once in the atmosphere, gases remain there for long periods of time (centuries). Greenhouse gases include carbon dioxide and methane.

    a. Assume the effect on global warming of a given amount of carbon dioxide is four times that of methane. What should be the relative fine (tax) on emissions of the two gases?

    b. How should the tax vary over time?

    c. A carbon tax is a tax related to the amount of carbon dioxide that burning gas, oil, or coal adds to the atmosphere. With a carbon tax, coal is taxed very heavily (relative to the amount of energy put out) and natural gas relatively lightly. A BTU (British thermal unit) tax is a tax related to the amount of energy produced, for instance, by burning gas, oil, or coal. If one is concerned about greenhouse gas warming, why is a carbon tax preferable to a BTU tax?

12. Two different strategies are debated for reducing greenhouse gas emissions. One is that all the countries in the world should adopt common measures, such as a carbon tax. The other is that all countries in the world should adopt common goals, such as reducing the level of emissions to the levels of 1990. Explain the distribution and efficiency aspects of these two strategies.

13. Corporate Average Fuel Economy (CAFE) standards stipulate the average fuel efficiency (miles per gallon) of cars produced by each manufacturer. That is, the average fuel efficiency of cars sold by GM, Ford, Chrysler, Toyota, and so forth must be at least equal to the standard set by the government. Explain why such a system introduces both inefficiencies and inequities among different automobile manufacturers. (Hint: Consider a company that specializes in small cars versus one that specializes in large cars.)

14. Discuss some of the problems with CAFE standards. How might a system of "tradable CAFE standards" be designed, and why might such a system improve efficiency?

15. To resolve the controversy over insurance coverage of Superfund sites, it has been proposed that the insurance industry be taxed to create a fund out of which claims would be paid. What difference does it make if the tax is levied on the basis of:

    a. Insurance premiums as of 1980?

    b. Current insurance premiums?

    Which insurance companies might be expected to prefer each way of levying the tax? (You can use a supply and demand diagram to illustrate the answers.) How does each form of tax affect the supply curve of insurance?

16. One proposal to reduce automobile emissions involves "pay at the pump insurance," under which individuals would pay, say, 25 cents per gallon of gasoline, with the proceeds going toward an insurance fund. What might be the environmental effects of such a proposal? Can you think of other grounds on which such a proposal might be attractive?

7

EFFICIENCY AND EQUITY

Chapter 3 took up Pareto efficiency, the condition in which no one can be made better off without making someone else worse off. It showed that in the absence of market failures, a free market would be Pareto efficient. Even if the competitive economy is efficient, however, the distribution of income to which it gives rise may be viewed as undesirable. One of the main consequences, and main objectives, of government activity is to alter the distribution of income.

The evaluation of a public program often entails balancing its consequences for economic efficiency and for the distribution of income. A central objective of welfare economics is to provide a framework within which these evaluations can be performed systematically. This chapter shows how economists conceptualize the trade-offs between efficiency and equity.

EFFICIENCY AND DISTRIBUTION TRADE-OFFS

Consider again a simple economy with two individuals, Robinson Crusoe and Friday. Assume initially that Crusoe has ten oranges, while Friday has only two. This seems inequitable. Assume, therefore, that we play the role of government and attempt to transfer four oranges from Crusoe to Friday, but in the process, one orange gets lost; hence, Crusoe ends up with six oranges and Friday with five. We have eliminated most of the inequity, but in the process, the total number of oranges available has been diminished. There is a **trade-off** between efficiency—the total number of oranges available—and equity.

The trade-off between equity and efficiency is at the heart of many discussions of public policy. Two questions are debated. First, there is disagreement about the nature of the trade-off. To reduce inequality, how much efficiency do we have to give up? Will one orange or two be lost in the process of transferring oranges from Crusoe to Friday?

Second, there is disagreement on the relative value to be assigned to a decrease in inequality compared to a decrease in efficiency. Some people claim that inequality is the central problem of society, and society should simply minimize the extent of inequality, regardless of the consequences to efficiency. Others claim that efficiency is the central issue. They argue that even if one wishes to help the poor, in the long run, the best way to do that is not to worry about how the pie is to be divided but to increase the size of the pie—to make it grow as rapidly as possible—so that there are more goods for everyone.

These disagreements relate to social choices between equity and efficiency. We now take a closer look at these choices.

ANALYZING SOCIAL CHOICES

When economists analyze consumer choice, the opportunity set is defined by the consumer's budget constraint, and the consumer's preferences are described by indifference curves (see Chapter 3). The individual chooses the point on the budget constraint that is tangent to an indifference curve, which puts him or her on the highest indifference curve feasible, given the budget constraint.

Economists have tried to use the same framework for analyzing social choices. The utility possibilities curve, introduced in Chapter 3, describes the opportunity set. It gives the highest level of utility (or welfare)

attainable by one individual, given the levels of utility attained by others. An economy is Pareto efficient if and only if it is operating along the utility possibilities schedule. The first fundamental theorem of welfare economics says that competitive economies are always on the utility possibilities schedule. The second fundamental theorem of welfare economics says that every point on the utility possibilities schedule can be attained through a competitive market process if the government redistributes initial endowments accordingly.

How does society select a point along the utility possibilities curve? Just as indifference curves for individuals describe how they make trade-offs between different goods, **social indifference curves** describe how society might make trade-offs between utility levels of different individuals. A social indifference curve gives the combinations of utility of, say, Crusoe and Friday, between which society is indifferent.

The two central questions of welfare economics can now be restated in terms of this social choice framework. Assume the current competitive market equilibrium is represented by the point A on the utility possibilities schedule depicted in Figure 7.1. Suppose society decides to move, say, from point A to point B along the utility possibilities schedule, representing an increase in Friday's utility and a reduction in Crusoe's utility. The first question is: What is the trade-off? The utility possibilities curve gives
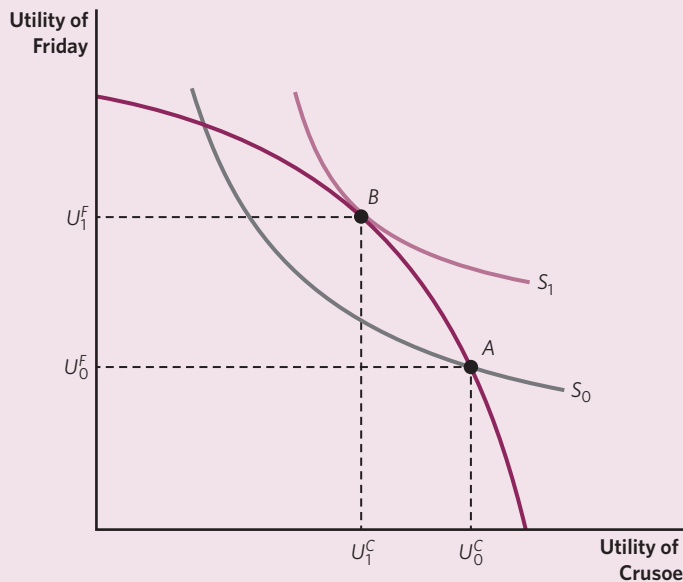
FIGURE 7.1

SOCIAL INDIFFERENCE CURVES

The social indifference curves describe how society evaluates trade-offs between Friday and Crusoe; it gives the combinations of utilities between which society is indifferent. Society is better off on a higher social indifference curve, just as an individual is better off on a higher individual indifference curve. Just as the individual chooses the point on the budget constraint at which the indifference curve is tangent to the budget constraint, society's preferred point on the utility possibilities curve is the point at which the social indifference curve is tangent to the utility possibilities curve.

the answer by showing the increase in Friday's utility from U_0^F to U_1^F and the decrease in Crusoe's utility from U_0^C to U_1^C. The second question concerns social preferences: How does society evaluate the trade-off? The slope of the social indifference curves gives the trade-offs for which society is indifferent. Point B is on the social indifference curve S_1, which is tangent to the utility possibilities curve, and lies on a higher indifference curve than S_0. Point B is therefore preferred by society.

The next two sections take a closer look at each of these questions regarding trade-offs and the economist's framework for analyzing social choice.

DETERMINING THE TRADE-OFFS

As we saw in Figure 7.1, the utility possibilities schedule shows us the trade-offs of transferring utility from Crusoe to Friday. The shape of the utility possibilities schedule tells us something more about those trade-offs. Consider the utility possibilities schedule shown in Figure 7.2. Assume that the economy lies at point A, at which Crusoe enjoys much more utility than Friday. Moving up and to the left along the schedule increases Friday's utility and decreases Crusoe's. Suppose we transfer oranges from Crusoe to Friday by moving in two steps, from point A to B to C. Clearly, this makes Crusoe worse off. As depicted in the figure, the decreases in Crusoe's utility are small in comparison to the increases in Friday's utility.

FIGURE 7.2

CRUSOE'S AND FRIDAY'S UTILITY POSSIBILITIES CURVE

As oranges are transferred from Crusoe to Friday, Crusoe's utility is decreased and Friday's increased. In moving from point A to B, the gain in Friday's utility appears much greater than the loss in Crusoe's utility. That is because Friday is so much worse off than Crusoe. In moving from B to C, the gain in Friday's utility is still larger than the loss in Crusoe's utility, but the trade-off has changed so that Friday's gain is smaller than the gain from A to B.
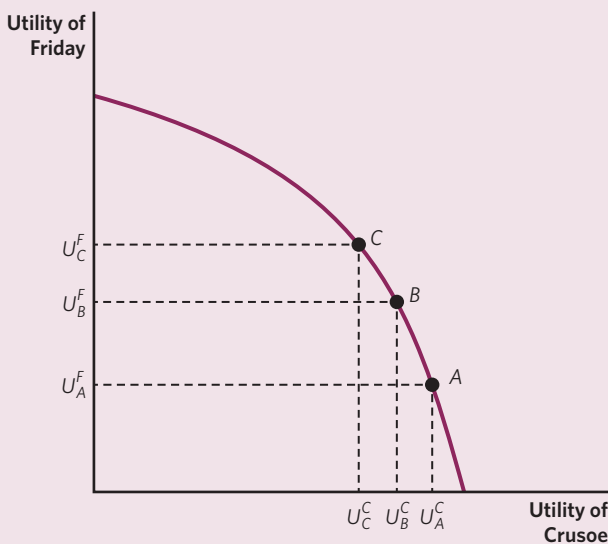
Utility theory helps explain this outcome. Economists use the term **utility function** to describe the relationship between the number of oranges and Friday's level of utility; the extra utility Friday gets from an extra orange is called his **marginal utility**. These are shown in Figure 7.3. At each point, marginal utility is the slope of the utility function—the change in utility from a unit change in orange consumption. Notice that as more oranges are consumed, utility rises more slowly, and marginal utility falls. (Thus, the slope of the utility function at point *C* is less than
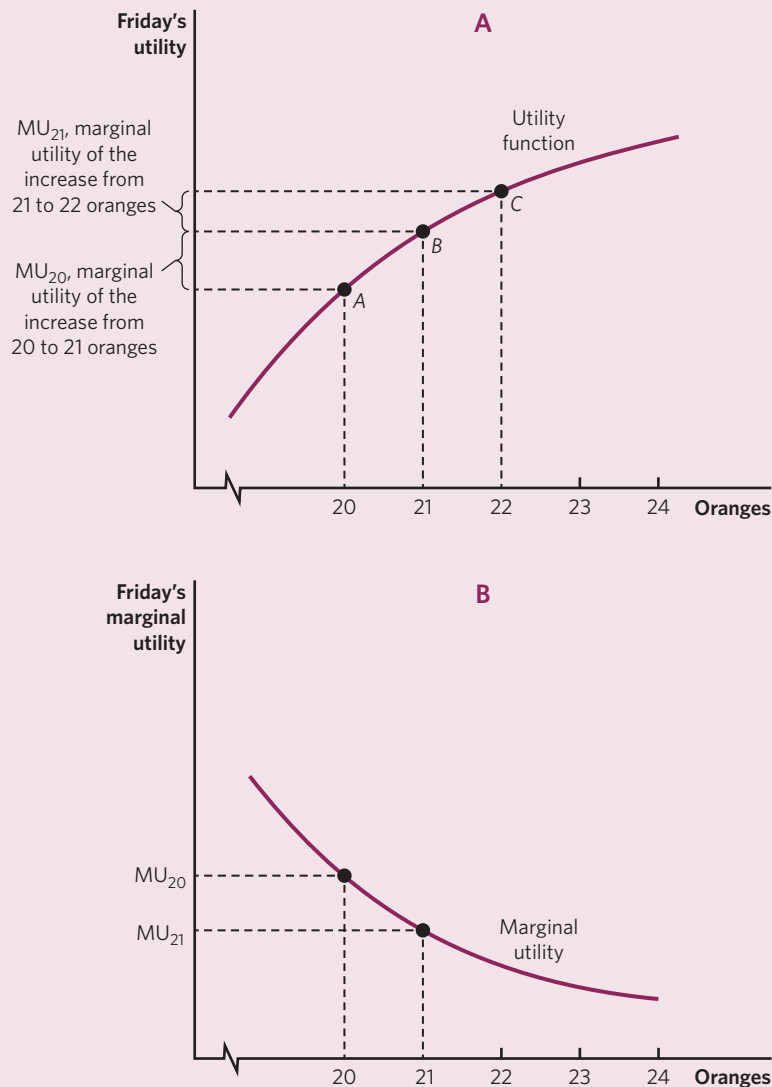
FIGURE 7.3

THE UTILITY FUNCTION AND MARGINAL UTILITY
(A) Shows the utility function: as we give Friday more oranges, his utility increases, but each additional orange gives him less extra utility.
(B) Shows marginal utility: the extra utility Friday gets from an extra orange decreases as the number of oranges increases, corresponding to the decreasing slope of the utility function.

FIGURE 7.4

UTILITY POSSIBILITIES
SCHEDULE WITH
COSTLY TRANSFERS

The set of points we can achieve through redistribution, when transfers are costly, lies within the utility possibilities curve, given costless transfers.
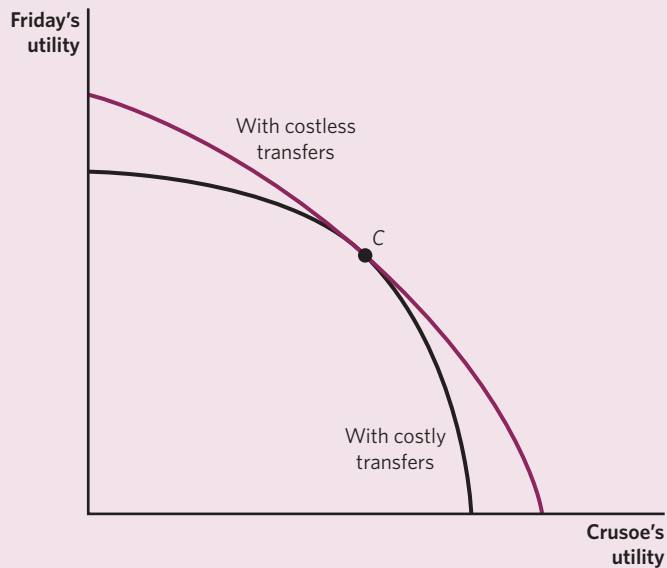
the slope at *A* or *B*.) This is because Friday enjoys the first orange very much, the next one a little less, and additional oranges still less. Finally, he becomes satiated and derives very little additional enjoyment from an additional orange. As an individual consumes more of any good, the extra gain from having one extra unit of that good becomes smaller. This phenomenon is referred to as **diminishing marginal utility**.[1]

By the same token, as we take away oranges from Crusoe, his utility decreases; and as we take away more and more oranges, the extra utility he loses from each additional loss of an orange increases. That is why with diminishing marginal utility, the utility possibilities schedule has the shape depicted in Figures 7.1 and 7.2. This shape says that when Friday has very little income (few oranges), we can increase his utility a great deal with a small decrease in Crusoe's utility, but when Friday is much better off, we can increase his utility only a little with even a large decrease in Crusoe's utility.

A second important determinant of the shape of the utility possibilities schedule is the efficiency with which we can transfer resources from one individual to another. In our society, the way we transfer resources from one group (say, the rich) to another (say, the poor) is by taxing the rich and subsidizing the poor. The way we do that normally interferes with economic efficiency.

---

[1] We write the utility function as $U = U(C_1, C_2, \ldots, C_n)$, where $C_1, C_2, \ldots, C_n$ represent the quantities of consumption of the various goods. Marginal utility of, say, C_1, is then simply the increase in U (utility) from an increase in consumption of C_1. Diminishing marginal utility implies that successive increments in C_1 yield successively smaller increments to U.

The rich may work less hard than they would otherwise because they reap only a fraction of the returns for their effort, whereas the poor may work less hard because by working harder, they may lose eligibility for benefits. The magnitude of these disincentives—a subject of considerable controversy—affects the entire shape of the utility possibilities schedule. In Figure 7.4, the red line represents the utility possibilities schedule assuming that it is costless to transfer resources. The black line, which lies far below the previous locus, except at point *C*—the point that occurs without any redistribution—represents the schedule when transfers are very costly.

EVALUATING THE TRADE-OFFS

The second basic concept used in analyzing social choices is the social indifference curve. As described in Chapter 3, an indifference curve gives those combinations of goods which give the individual the same level of utility. Just as individuals derive utility from the goods they consume, we can think of society as deriving its welfare from the utility received by its members. The **social welfare function** gives the level of social welfare corresponding to a particular set of levels of utility attained by members of society. The *social indifference curve* is defined as the set of combinations of utility of different individuals (or groups of individuals) that yields equal levels of welfare to society—for which, in other words, the social welfare function has the same value.

The social welfare function provides a basis for ranking any allocation of resources: we choose the allocations that yield higher levels of social welfare. The Pareto principle says that we should prefer those allocations in which at least some individuals are better off and no one is worse off. It says that if some individuals' utility is increased and no one else's utility is decreased, social welfare increases. Thus, in Figure 7.5, the combinations to the northeast of *A* make everyone better off, and hence satisfy the Pareto principle.

Unfortunately, most choices involve trade-offs, with some individuals being made better off and others worse off. At point *B* the second group is better off than at *A*, but the first group is worse off. We thus need a stronger criterion, and this is what the social welfare function provides. The social indifference curves provide a convenient diagrammatic way of thinking about the kinds of trade-offs society faces in these situations. Thus, in Figure 7.5, all combinations of the utilities of Groups 1 and 2 that are on the social indifference curve labeled W_2 yield a higher level of social welfare than those combinations on the curve labeled W_1. This shows that *B* is preferred to *A*.

Social welfare functions can be thought of as a tool economists use to summarize assumptions about society's attitudes toward different

SOCIAL INDIFFERENCE
CURVES

Society is willing to trade off
some decrease in the utility
of one group for an increase in
that of another group. A social
indifference curve gives the
combinations of utilities of
Group 1 and Group 2 between
which society is indifferent.
Points on the social indifference
curve labeled W_2 yield a higher
level of social welfare than do
points on the social indifference
curve labeled W_1.
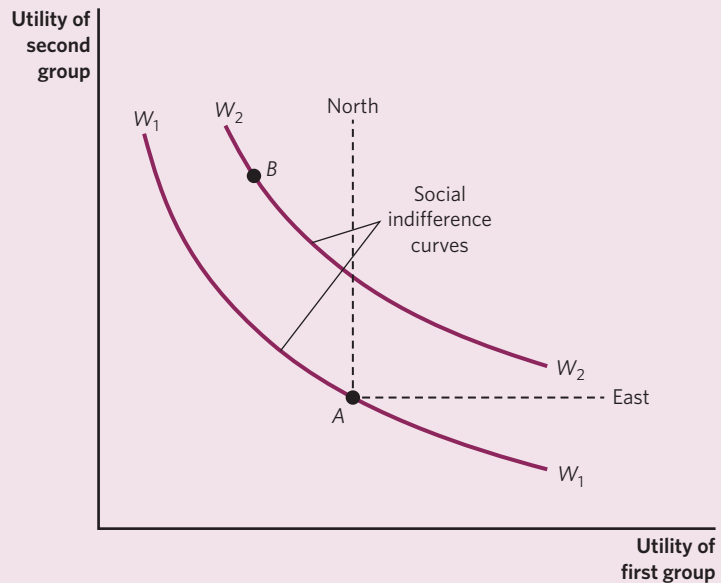
distributions of income and welfare. If society is very concerned about inequality, it might not care that Crusoe has to give up seventy oranges for Friday to get one orange, "since Crusoe has so many to begin with." As long as Friday is poorer than Crusoe, any sacrifice on Crusoe's part that makes Friday better off would be justified. On the other hand, society might not care at all about inequality; it could value an orange in the hands of Friday exactly the same as an orange in the hands of Crusoe, even though Friday is much poorer. In that case, it would focus only on efficiency—the number of oranges available. No redistribution of oranges from Crusoe to Friday would be justified if, in the process, a single orange was lost.

UTILITARIANISM   Social welfare functions—and the associated social indifference curves—can take a variety of shapes; Figure 7.6 illustrates three different cases. In Figure 7.6A, the social indifference curve is a straight line, implying that no matter what the level of utility of Friday and Crusoe, society is willing to trade off one "unit" of Friday's utility against one unit of Crusoe's. The view represented by this social indifference curve has a long historical tradition. Jeremy Bentham was the leader of a group, called **utilitarians**, that argued that society should maximize the sum of the utilities of its members; in our simple example with two individuals, the social welfare function is

$$W = U_1 + U_2.$$

A

Crusoe's utility (U_2)

Friday's utility (U_1)

Utilitarian social indifference curve

B

Crusoe's utility (U_2)

Change in U_2

Friday's utility (U_1)

Change in U_1

C

Crusoe's utility (U_2)

$U_2^{\star}$

Rawlsian social indifference curve

$U_1^{\star}$

Friday's utility (U_1)
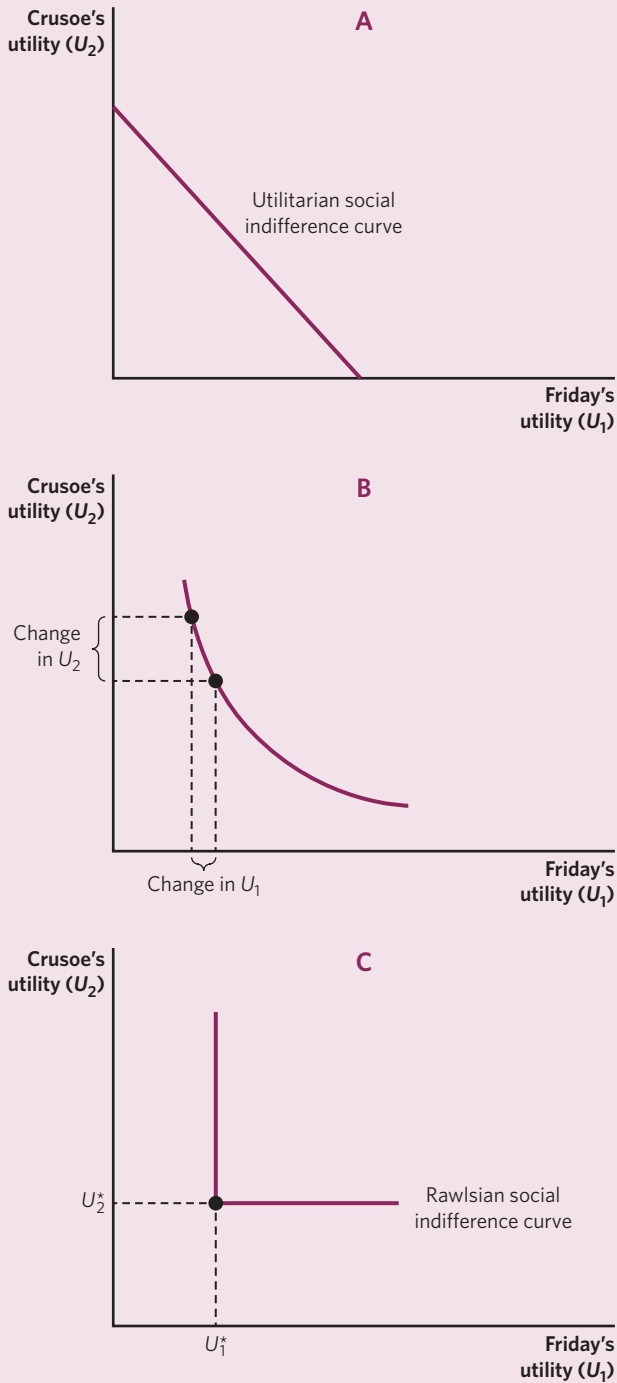
FIGURE 7.6

ALTERNATIVE SHAPES OF SOCIAL INDIFFERENCE CURVES

(A) A utilitarian is willing to give up some utility for Crusoe as long as Friday gains at least an equal amount of utility. The social indifference curves are straight lines. (B) Some argue that society requires more than an equal increase in the utility (U_2) of a rich individual to compensate for a decrease in the utility (U_1) of a poor individual. (C) Rawls maintains that no amount of increase in the welfare of the rich can compensate for a decrease in the welfare of the poor. This implies that the social indifference curves are L-shaped.

It is clear that with this social welfare function, the social indifference curve has the shape depicted in Figure 7.6A.

It is important to emphasize that with a utilitarian social welfare function, society is not indifferent to an increase of one orange (or one dollar of income) for Individual 1 and a decrease of one orange (or one dollar of income) for Individual 2. If Individual 1 has a lower level of income (fewer oranges) than Individual 2, then the increase in utility of Individual 1 from one more orange (one more dollar) will be greater than the decrease in utility for Individual 2. What the utilitarian social welfare function says is that the *utility* of any individual should be weighted equally to the utility of any other individual.

Many would argue that when one individual is worse off than another, society is not indifferent to a decrease in the utility of the poorer (Individual 1) matched by an equal increase in the utility of the richer (Individual 2). Society should be willing to accept a decrease in the utility of the poor only if there is a much larger increase in the utility of the rich. The social indifference curve reflecting these values is drawn in Figure 7.6B, where it appears not as a straight line but as a curved one; as the poorer individual becomes worse and worse off, the increment in utility of the richer individual that makes society indifferent must be larger and larger (i.e., the slope of the social indifference curve becomes steeper and steeper).

RAWLSIANISM   An extreme position of this debate was taken by John Rawls, a former professor of philosophy at Harvard University. Rawls argued that the welfare of society depends only on the welfare of the worst-off individual. So society is better off if you improve that individual's welfare, but gains nothing from improving the welfare of others. In his view, there is no trade-off. If Friday is worse off than Crusoe, then anything that increases Friday's welfare increases social welfare. As oranges are transferred from Crusoe to Friday, it makes no difference how many are lost in the process—how inefficient the transfer process is—as long as Friday gets *something*. To put it another way, no amount of increase in the welfare of the better-off individual could compensate society for a decrease in the welfare of the worst-off individual. Diagrammatically, this is represented by an L-shaped social indifference curve, as in Figure 7.6C.[2]

---

[2] The social welfare function is written:

$$W = \min \{U_1, \ldots , U_n\}.$$

Social welfare reflects only the utility of the worst-off member of society.

TWO CAVEATS

Many public sector economists have made extensive use of the concepts of social welfare functions and the utility possibilities curve, but these concepts have also been extensively criticized, on several grounds.

INTERPERSONAL COMPARISONS   We assume that when an individual consumes more, his or her utility rises. But we cannot measure the level of utility or the change in utility. Social welfare functions seem to assume not only that there is a meaningful way of measuring an individual's utility,[3] but that there is a meaningful way of comparing the utility of different individuals. For example, with the utilitarian social welfare function, we add up the utility of the different members of society. Because we add Crusoe's and Friday's utility together, we are assuming that somehow we can compare, *in a meaningful numerical way*, their levels of utility. But when we transfer an orange from Robinson to Friday, how can we compare in an objective way the value of Friday's gain and Robinson's loss?

The same problem arises with a **Rawlsian social welfare function**, where we are told to maximize the welfare of the worst-off member of society. To judge who is worst off, we must somehow compare utilities.

Many economists believe that these **interpersonal utility comparisons** cannot be made in any meaningful way. I may claim that although I have a much higher income than my brother, I am less happy; not only that, I may claim that I know how to spend income much better, so that the extra increment in my utility from a dollar given to me is much greater than the extra increment in utility that he would get from receiving an extra dollar. How could anyone prove that I was wrong (or right)? Because there is no way of answering this question, economists argue that there can be no scientific basis for making welfare comparisons. And because there is no "scientific" basis for making such welfare comparisons, many economists believe they should limit themselves to describing the consequences of different policies—only pointing out who are the gainers and who are the losers—and that should be the end of their analysis.

> ## SOCIAL CHOICE IN THEORY
>
> **1.** *Construct the opportunity set.* The utility possibilities schedule describes how much one person's utility must be decreased when another's is increased.
> **2.** *Define preferences.* Social indifference curves describe how much society is willing to decrease one person's utility to increase another's by a given amount.
> **3.** *Adopt programs that increase social welfare.* Find the programs that put society on the highest social indifference curve.

---

[3] In some situations, it may be possible to use the amount of money an individual would be willing to pay for an object as a *measure* of the utility of that object. However, this does not resolve the problem of comparing utilities across individuals.

COMPARING INDIVIDUAL AND SOCIAL CHOICES

INDIVIDUAL CHOICES	SOCIAL CHOICES
Step 1: Define Opportunity Set	
Budget constraint	Utility possibilities curve
Step 2: Define Preferences	
Individual indifference curve	Social indifference curve
Step 3: Choose Preferred Point	
Tangency between individual indifference curve and budget constraint	Tangency between social indifference curve and utility possibilities curve

They believe that the only circumstances in which economists should make welfare judgments are those in which the policy change is a Pareto improvement. Unfortunately, as we have said, few policy changes are Pareto improvements; hence, without making interpersonal comparisons of welfare, economists have little to say regarding policy.

WHENCE SOCIAL WELFARE FUNCTIONS? The second set of objections concerns the very nature of social welfare functions. Individuals have preferences; they can decide whether they prefer some combination of apples and oranges to another combination of apples and oranges to another combination. Society consists of many individuals, but society itself does not have preferences. We can describe the preferences of each individual, but whose preferences does the social welfare function represent? If there were a dictator, the answer to that question would be easy: the social welfare function would reflect the preferences of the dictator. But in a democratic society, there is no easy answer to the question. Some individuals (particularly the rich) may care little for redistribution, whereas others (particularly the poor) may argue that greater weight should be placed on redistribution.

As a descriptive matter—as part of a positive analysis—societies seldom exhibit consistency. One of the results to be described in Chapter 9 explains why this is not unexpected. Most economists think of the concepts we have described—as part of a normative analysis—as tools that help us think systematically about the trade-offs society constantly must face. As we noted earlier, the systematic analysis of these trade-offs actually constitutes an important part of the process by which decisions get made.

SOCIAL CHOICES IN PRACTICE

In practice, government officials do not derive utility possibilities schedules, nor do they write down social welfare functions. But their approach to deciding whether, say, to undertake any particular project does reflect the concepts we have introduced.

First, they attempt to identify and measure the net benefits (benefits minus costs) received by different groups. Second, they ascertain whether

the project is a Pareto improvement—that is, whether everyone is better off. If so, clearly the project should be undertaken (this is the Pareto principle).

If the project is not a Pareto improvement, matters are more difficult. Some gain, some lose. The government needs to make an overall judgment. One commonly used approach looks at two summary statistics, describing "efficiency" and "equity" effects. *Efficiency* is measured by simply summing the gains or losses for each individual (which are calculated in a manner to be described shortly). *Equity* is measured by looking at some overall measure of inequality in society. If a project has net positive gains (positive efficiency effects) and reduces measured inequality, it should be undertaken. If a project has net positive losses and increases measured inequality, it should not be undertaken. If the efficiency measure shows gains but the equality measure shows losses (or vice versa), there is a trade-off, which is evaluated using a social welfare function: How much extra inequality is society willing to accept for an increase in efficiency?

There are numerous examples in which choices between equality and efficiency must be made. For instance, in general, the more a tax system redistributes income, the greater the inefficiencies it introduces. There is a trade-off between equality and efficiency. There are, of course, important instances of poorly designed tax systems; such tax systems put the economy below its utility possibilities schedule. In such cases, it may be possible to increase both equality and efficiency.

We now take a closer look at how economists measure efficiency and inequality.

MEASURING BENEFITS

The first problem is how to measure the benefits of some program or project to particular individuals. In the earlier discussion of utility theory, we described how giving Friday more oranges increased his utility. But how do we measure this?

The standard way this is done is in terms of *willingness to pay*. We ask how much an individual would be willing to pay to be in one situation rather than another. For example, if Joe likes chocolate ice cream more than vanilla, it stands to reason that he would be willing to pay more for a scoop of chocolate ice cream than for a scoop of vanilla. Or if Diane would rather live in California than in New Jersey, it stands to reason that she would be willing to pay more for the West Coast location.

Notice that how much a person is willing to pay is different from how much that individual *must* pay. Just because Joe is willing to pay more for

FIGURE 7.7

UTILITY AND
MARGINAL UTILITY

(A) shows that utility increases
continually with consumption
but tends to level off as
consumption climbs higher.
(B) explicitly shows marginal
utility; notice that it declines
as consumption increases.
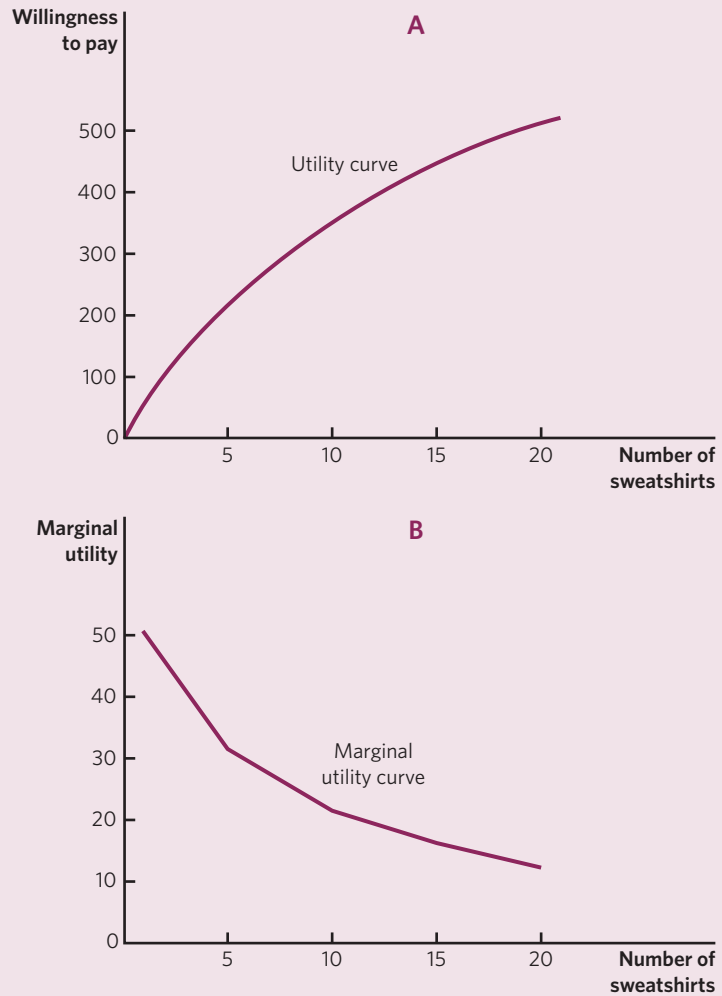
chocolate ice cream than for vanilla does not mean he will have to pay more. What he has to pay depends on market prices; what he is willing to pay reflects his preferences.

Using willingness to pay as our measure of utility, we can construct a diagram like of Figure 7.7A, which shows the level of utility Mary receives from sweatshirts as the number of sweatshirts she buys increases. This information is also given in Table 7.1. Here we assume that Mary is willing to pay $200 for five sweatshirts, $228 for six sweatshirts, $254 for seven sweatshirts, and so on. Thus, five sweatshirts give her a utility

TABLE 7.1 UTILITY AND MARGINAL UTILITY

NUMBER OF SWEATSHIRTS	MARY'S WILLINGNESS TO PAY (UTILITY)	MARGINAL UTILITY
0	0	50
1	50	45
2	95	40
3	135	35
4	170	30
5	200	28
6	228	26
7	254	24
8	278	23
9	301	22
10	323	21
11	344	20
12	364	19
13	383	18
14	401	17
15	418	16
16	434	15
17	449	14
18	463	13
19	476	12
20	488	

of 200, six a utility of 228, and seven sweatshirts a utility of 254. Mary's willingness to pay increases with the number of sweatshirts, reflecting the fact that additional sweatshirts give her additional utility. The extra utility of an additional sweatshirt measured here by the additional amount she is willing to pay, is the marginal utility. The numbers in the third column of Table 7.1 give the marginal, or extra, utility she received from her last sweatshirt. When Mary owns five sweatshirts, an additional sweatshirt yields her an additional, or marginal, utility of 28 (228 – 200); when she owns six sweatshirts, an additional one gives her a marginal utility of only 26 (254 – 228). Figure 7.7B traces the marginal utilities of each of the increments.[4]

---

[4] As marginal utility is the extra utility from an extra unit of consumption, it is measured by the slope of the utility curve in Figure 7.7A.

ORDINARY AND COMPENSATED DEMAND CURVES

We can use the concept of willingness to pay to construct a demand curve. We have already asked how much Mary is willing to pay for each additional sweatshirt. If the price of sweatshirts is $29, then she will buy five sweatshirts. She would have been willing to pay $30 for the fifth sweatshirt, so clearly, the marginal benefit of the fifth sweatshirt exceeds its cost; but she is willing to pay only $28 for the sixth sweatshirt, so the marginal benefit is less than the cost. Thus, the marginal utility curve drawn in Figure 7.7B can also be thought of as the demand curve.

However, this is a special demand curve, called the **compensated demand curve**, which differs slightly from the ordinary demand curve. Recall that we constructed the compensated demand curve by asking how much Mary would be willing to pay for each additional sweatshirt; thus, as we give her more sweatshirts, we are always keeping her at exactly the same level of utility.

To construct the ordinary demand curve, we need to know how many units of the commodity Mary would buy at each price. As the price is lowered, Mary not only demands more, but is made better off. As prices are lowered, individuals substitute the cheaper good for others goods. If the price of sweatshirts is lowered, Mary will substitute sweatshirts for sweaters. This is called the **substitution effect**. Because of the lower price, Mary is better off; if she bought exactly the same amount of goods that she did

FIGURE 7.8

COMPENSATED VERSUS UNCOMPENSATED DEMAND CURVES

The compensated demand curve gives the demand for a good, assuming as price is changed that money is taken away, or given to the individual to leave that individual just as well off as he or she was before the price change. It thus measures only the substitution effect associated with the price changes. Because, as price is lowered individuals are better off, and as a result buy slightly more of (normal) commodities, the ordinary demand curve is slightly flatter than the compensated demand curve.
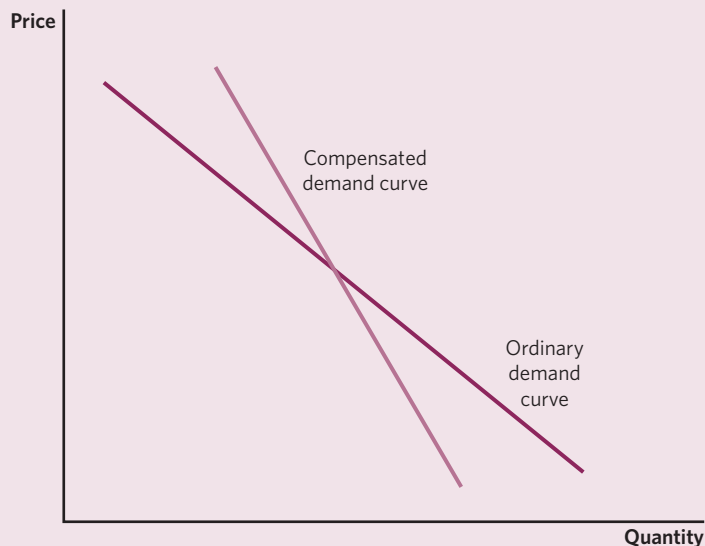
before, she would have money left over. She spreads this money around. Some of it is spent on buying sweatshirts. The increase in demand for sweatshirts as a result of the fact that Mary is better off—it is *as if* she had more income—is called the **income effect**. If we take away this extra money, we have the compensated demand curve; we eliminate the income effect. Thus, the compensated demand curve reflects only the substitution effect. In most cases, the differences between the two are negligible. If Mary spends one-tenth of 1 percent of her income on sweatshirts, taking away the extra income has almost no effect on her demand for sweatshirts, or any other commodity. Thus, Figure 7.8 shows the ordinary and compensated demand curves as being almost the same, with the ordinary demand curve being slightly flatter (lowering the price from its current level results in a slightly greater increase in the quantity demanded, and raising the price from its current level results in a slightly greater decrease in quantity demanded).

CONSUMER SURPLUS

The difference between what an individual is willing to pay and what he or she has to pay is called the **consumer surplus**. Mary would have been willing to pay $50 for the first sweatshirt, $45 for the second, $40 for the third, and so on. If the market price is $29, however, that is all she has to pay for each sweatshirt. Thus, on the first sweatshirt, she gets a surplus of $21 ($50, what she was willing to pay, minus $29, what she actually pays); on the second sweatshirt, she gets a surplus of $16; on the third sweatshirt, she gets a surplus of $11, and so on. The total consumer surplus is thus the sum: $21 + $16 + $11 + $6 + $1 = $55.

Diagrammatically, the consumer surplus is depicted in Figure 7.9 as the shaded area under the compensated demand curve and above the price line. Of course, because the compensated and uncompensated demand curves are almost the same, typically, we calculate the consumer surplus simply by looking at the area under the ordinary demand curve above the price line.

USING CONSUMER SURPLUS TO CALCULATE THE BENEFITS OF A GOVERNMENT PROJECT   The compensated demand curve can be useful for measuring the benefits of government projects. For instance, constructing a bridge on which no toll will be charged can be thought of as lowering the price from "infinity" (one simply cannot buy trips across a nonexistent bridge) to zero. The welfare gain is just the total consumer surplus, the area under the demand curve in Figure 7.10. This measures the maximum individuals could pay and still be as well off with the bridge as they were without it. Clearly, if the consumer surplus is less than the

FIGURE 7.9

GRAPHICAL REPRESENTATION OF CONSUMER SURPLUS

An individual's surplus is the difference between what the individual is willing to pay (represented by the area beneath the demand curve) and what he or she actually pays (the area under the price line). The consumer surplus here is indicated by the shaded region.
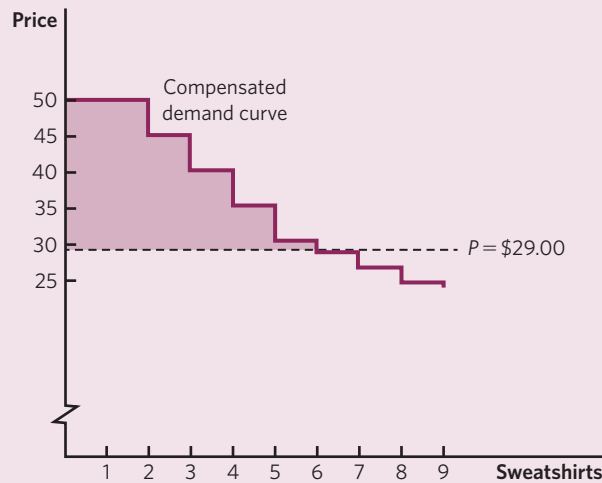
cost of the bridge, it does not pay to construct it, whereas if the consumer surplus is greater than the cost of the bridge, it does pay to build it.

There are several ways that economists go about trying to measure consumer surplus and willingness to pay. For many goods, there are data with which economists can construct the demand curve (the quantity that individuals are willing to purchase at each price) and the compensated

FIGURE 7.10

MEASURING THE BENEFITS OF A GOVERNMENT PROJECT: BUILDING A BRIDGE

The benefits of a bridge for which no tolls will be charged can be measured by the total area under the demand curve—the total consumer surplus.
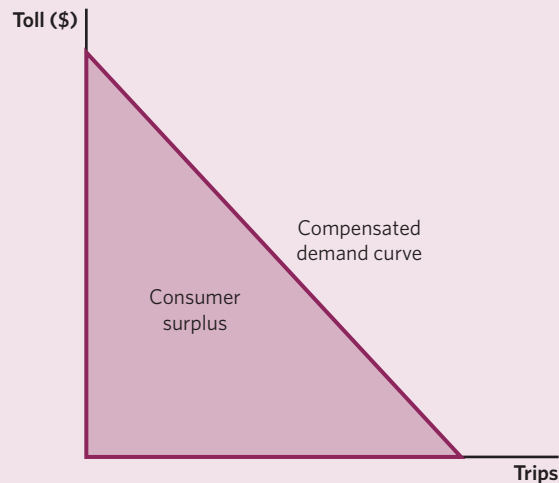
demand curve.[5] In that case, willingness to pay can be calculated simply as the area under the compensated demand curve. For some goods, such as the Grand Canyon, there is no market demand curve, yet the government still might want to know how much citizens are willing to pay to preserve it in its pristine condition. Economists have designed elaborate survey techniques to elicit meaningful answers from individuals concerning their willingness to pay. These methods are discussed at greater length in Chapter 11.

CONSUMER SURPLUS

- Measured by the area under the (compensated) demand curve.
- Used to measure the value of a government project or assess the magnitude of an inefficiency.

MEASURING AGGREGATE SOCIAL BENEFITS

We have now described how we can measure the benefits that an individual receives. Social benefits are typically measured by adding up the benefits received by all individuals. The numbers obtained represent the total willingness to pay of all individuals in society. The difference between the total willingness to pay and the total costs of a project can be thought of as the net "efficiency" effect of the project. It is a dollar value of the net benefits.

MEASURING INEFFICIENCY

In assessing alternative policies, economists have put particular emphasis on economic efficiency. Taxes are criticized for discouraging work effort, monopolies for restricting production and driving up prices. To measure the dollar value of an inefficiency, economists use exactly the same methodology they use to measure the dollar value of a new project. There, we calculated the consumer surplus associated with the project. Here, we calculate the consumer surplus associated with the elimination of the inefficiency. That is, economists ask: How much would an individual be willing to give up to have the inefficiency eliminated? Consider the inefficiency caused by a tax on cigarettes. We ask each individual how much he or she would be willing to pay to have the tax on cigarettes eliminated. Say one answer is $100; thus eliminating the cigarette tax and imposing in its place a $100 **lump-sum tax**—that is, a tax that the individual would

---

[5] As was noted previously, for most goods, the compensated and uncompensated demand curves are very similar. If the *income elasticity* (the percentage increase in the demand for the good when income increases by 1 percent) is known, one can calculate the compensated demand curve from the uncompensated demand curve.

have to pay regardless of what he or she did—leaves this individual's welfare unchanged. The difference between the revenue raised by the cigarette tax (say, $80) and the lump-sum tax that the individual would be willing to pay is called the **deadweight loss** or **excess burden** of the tax. It is the measure of the inefficiency of the tax. Taxes, other than lump-sum taxes, give rise to a deadweight loss because they cause individuals to forgo more-preferred consumption in favor of less-preferred consumption to avoid payments of the tax. Thus, even a tax that raises no government revenue—because individuals completely avoid purchasing the taxed commodity—can have a substantial excess burden.

DRAWING A POVERTY LINE

The official poverty line determines how many people the government counts as poor. But what determines the poverty line itself?

In the early 1960s, Mollie Orshansky, an official at the Social Security Administration, developed a method of measuring poverty from a survey of household expenditures. She found that a typical family spent one-third of its income on food. She then gathered information on minimum food budgets for families of various sizes, and multiplied that number by 3 to get an estimate of the poverty line for the different family sizes. With minor changes, Orshansky's poverty line was officially adopted in 1969 and it has been increased by the overall rate of inflation since then.

One can ask a number of questions about how poverty is measured. Here are three:

First, the survey Orshansky relied on to find that households spent one-third of their income on food was taken in 1955. Since then, household expenditures have shifted. Households now spend a much lower percentage of income on food, perhaps one-fourth or one-fifth. If the minimum food budget were accordingly multiplied by 4 or 5, the poverty line would be much higher.

Second, the poverty line does not take in-kind benefits into account. In-kind benefits include any benefits that are not received in cash form, such as Medicaid, food stamps, and subsidized school lunches. If those benefits are measured as additional income, the number of people below the poverty line falls by about 20 percent.

Finally, some critics have proposed that poverty should be thought of as a relative rather than an absolute concept. They argue that those at the bottom of society—say, the bottom 5 or 10 or 20 percent—are poor relative to everyone else. Poverty is more appropriately viewed as an extreme case of inequality.

For many, this last criticism goes too far. They fear that a relative concept of poverty could reduce the moral urgency of fighting poverty. There is broad social support for efforts to ensure that people have basic levels of food, housing, clothing, and medical care, even if defining those amounts is controversial.

In 1995, a National Academy of Sciences (NAS) study proposed major revisions in how we measure poverty. Although there was agreement about including noncash income, the difficult problems of how best to include health care expenditures were not fully resolved. Should a sick, poor person who receives $150,000 for a kidney transplant have that added to his or her income, in which case he or she now appears to be in an upper income bracket? The study proposed an adjustment in the poverty level that went beyond just taking into account inflation, but it did not propose increasing the poverty level in proportion to increases in average income, which would have made poverty a purely relative phenomenon. However, even this compromise generated

a strong dissent from one of the members of the academy's panel.

The U.S. Census Bureau tested a number of experimental poverty measures based on recommendations of the 1995 NAS report, and, in 2009, the Interagency Technical Working Group on Developing a Supplemental Poverty Measure was formed. In 2011, the Census Bureau, with support from the Bureau of Labor Statistics, released its first report presenting a newly developed supplemental poverty measure (SPM) that incorporated many NAS-based measures. According to the Census Bureau, the SPM will not replace the official poverty measure and will not be used to determine eligibility for government programs. Instead, it is designed to provide an alternative perspective of poverty that better reflects contemporary social and economic conditions, including government policies that significantly alter resources available to families, and hence, their poverty status. When compared with the official poverty measure in 2011, the SPM estimates a slightly higher overall poverty rate (16.1 versus 15.1 percent), but a substantially lower poverty rate for those under 18 years old (18.1 versus 22.3 percent) and almost double the poverty rate for those 65 years and older (15.1 verus 8.7 percent).

SOURCES: Joyce E. Allen and Margaret C. Simms, "Is a New Yardstick Needed to Measure Poverty?" *Focus* (February 1990): 6–8; *Measuring Poverty: A New Approach* (Washington, DC: National Academy of Sciences, 1996); Gordon M. Fisher, "The Development and History of the U.S. Poverty Thresholds—A Brief Overview," *GSS/SSS Newsletter* (Winter 1997): 6–7; and Kathleen Short, "The Research Supplemental Poverty Measure: 2011," *Current Population Reports*, P60-244, United States Census Bureau, November 2012.

We can calculate the deadweight loss using compensated demand curves. Assume the cost of producing a cigarette is c_0, and the tax raises the price from c_0 to $c_0 + t$, where t is the tax per pack. We assume the individual consumes q_0 packs of cigarettes with the tax, and q_1 after the tax has been removed (but replaced by a lump-sum tax that leaves the individual no better or worse off than when there was a cigarette tax). We have drawn the resulting compensated demand curve in Figure 7.11. The deadweight loss is measured by the shaded area ABC, the area under the compensated demand schedule and above c_0, between the output with and without the tax.

The triangle ABC is sometimes called a **Harberger triangle**,[6] in honor of University of Chicago and UCLA economist Arnold Harberger, who used such triangles not only to measure the inefficiencies associated with distortionary taxation but also to measure other inefficiencies, such as those associated with monopoly. Why does the Harberger triangle provide a measure of deadweight loss? The price tells us the value of the last unit consumed; that is, at q_0, the individual is willing to trade off $p_0 = c_0 + t$ units of "income" (with which he or she could have purchased other goods) for one more pack of cigarettes. Of course, when the individual has $q_0 + 1$ packs of cigarettes, he or she will value an additional pack of cigarettes less than when he or she has q_0 packs, and so the price the individual is willing to pay will fall.

---

[6] See, for instance, A. Harberger, "Taxation, Resource Allocation and Welfare," in *The Role of Direct and Indirect Taxes in the Federal Revenue System*, ed. J. Due (Princeton, NJ: Princeton University Press, 1964), reprinted in A. Harberger, *Taxation and Welfare* (Chicago: University of Chicago Press, 1974).
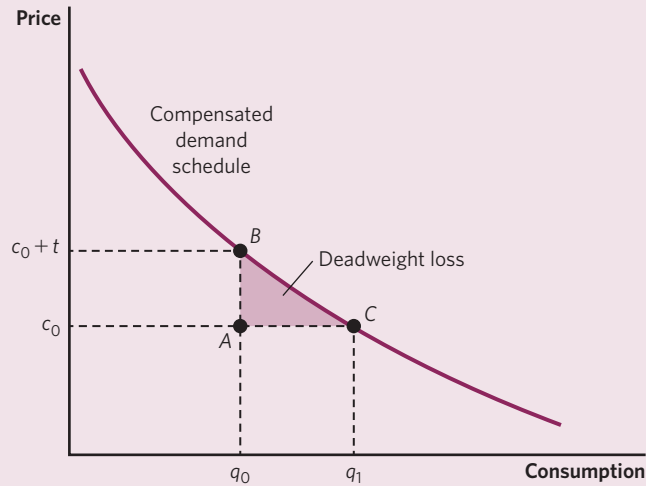
Assume that initially consumption is 100 packs, and consumption increases by 10 packs when the tax is removed; the tax is 10 cents, and the cost of production is $1 per pack. (Tax revenue is 100 packs times 10 cents per pack, or $10.) The individual is willing to pay $1.10 for the first additional pack, $1.09 for the second, $1.08 for the third, and so on. If the tax were eliminated, and the price fell to c_0, the cost of production ($1 a pack), the total amount that the individual would be willing to pay would be 10 cents times 100 packs = $10 (the amount saved on the first 100 packs the individual has purchased, which is equal to the tax revenue) plus 10 cents for the 101st pack (the difference between how much the individual values the 101st pack and what he or she must pay), 9 cents for the 102nd pack, and so on. Remember, we are calculating how much *more* the individual would be willing to pay beyond the $1 that he or she will have to pay for each pack. The total that the individual would be willing to pay is thus $10.50. Because the tax raised revenue of $10, the deadweight loss is 50 cents, which is, of course, just the area under the compensated demand curve and above c_0, between q_0 and q_1.

QUANTIFYING DISTRIBUTIONAL EFFECTS

Assessing the distributional effects of a project or a tax is often far more complex than assessing the efficiency effects. There are many groups in a society, and each may be affected differently. Some poor individuals may be hurt, some helped; some middle-income individuals may be helped, others hurt. In some cases, the rich may be helped the most, the poor helped moderately, and the middle class made only slightly worse off.

THE GREAT GATSBY CURVE

Rising inequality in the United States since the 1970s has been well documented; examples include the dramatic contrast in real after-tax income growth between the top 1 percent of families and everyone else, the steadily shrinking middle class, and the substantial rise in the Gini coefficient.

However, less attention has been given to the relationship between current levels of inequality and intergenerational income mobility. Recent studies have found that your parents' income is a good predictor of your subsequent income.

Alan Krueger, Chairman of the Council of Economic Advisers under President Obama, created what he calls the "Great Gatsby Curve," named after F. Scott Fitzgerald's fictional character known for his lavish lifestyle during the Roaring Twenties. The Great Gatsby Curve plots the **Gini coefficient**—current income distribution—on the *x*-axis and intergenerational income elasticity (IGE)—relation between parents' and children's income—on the *y*-axis. It shows a disconcerting link between IGE and income inequality at a given point in time: countries that have a high degree of inequality also tend to have less economic mobility across generations. For example, OECD countries that had more inequality across households in the 1980s also had more persistence in income from one generation to the next.

Not only did the United States have greater income stagnation than most other OECD countries in this study, but one cannot help but be concerned that already low income mobility across generations has been exacerbated by the rise of inequality in the United States. Using the Great Gatsby Curve, Krueger estimates that the IGE for the United States will increase from 0.47 to 0.56 for the next generation. In other words, lack of equality threatens equality of opportunity in the United States—the fortunes of one's parents are increasingly important in determining the fate of their children.

In practice, governments focus on a few summary measures of inequality. Because the poor are of particular concern, they receive special attention. The **poverty index** measures the fraction of the population whose income lies below a critical threshold; below that threshold, individuals are considered to be in poverty. In 2010, the poverty threshold for a family of four was $22,113.[7]

Another measure is the **poverty gap**. The poverty index only counts the number of individuals who are below the poverty threshold; it does not look at how far below that threshold they are. The poverty gap asks: How much income would we have to give to the poor to bring them all up to the poverty threshold?

Two other measures are briefly discussed in the appendix to this chapter.

---

[7] U.S. Census Bureau, *Poverty Thresholds*, http://www.census.gov.

THREE APPROACHES TO SOCIAL CHOICES

We now have the basic tools for describing social choices in the difficult cases in which the project does not constitute a Pareto improvement. There are three approaches, which we shall refer to as the *compensation principle*, *trade-offs across measures*, and the *weighted benefits approach*.

THE COMPENSATION PRINCIPLE

What happens if the total willingness to pay exceeds the total costs, but the costs borne by some individuals exceed their willingness to pay? Should the project be undertaken? The **compensation principle** says that if the aggregate willingness to pay exceeds the cost, the project should be undertaken. Most economists criticize this principle, for it ignores distributional concerns. Only if the compensation is actually paid to those adversely affected can we be sure that the project is desirable, for then it is a Pareto improvement.

Because the compensation principle does not pay adequate attention to distributional concerns, economists have turned to two other approaches.

TRADE-OFFS ACROSS MEASURES

With a measure of efficiency (net benefits) and a measure of inequality, public decision making—conceptually, at least—should be easy: one simply evaluates whether the increase in efficiency is worth the increase in inequality, or vice versa.

The previous two sections have described how we measure total efficiency and inequality. These are just statistics, numbers that help to summarize the impacts of a project or program. Such summary statistics, though useful, often submerge some of the detailed information that is important in public decision making. Ideally, we would look at the impacts on each individual, and then use the social welfare function to add up the effects. In practice, the government does not attempt to identify impacts on every individual, but it does attempt to ascertain the effects on each major group. For instance, it may look at the impact on individuals in different income categories—say, families with incomes below $10,000, between $10,000 and $20,000, and so forth.

WEIGHTED NET BENEFITS

This may be all the information required for policy makers to make a decision. If the aggregate net benefit (the sum of the willingnesses to pay minus costs) is positive, and if the poor are net beneficiaries and the rich are net losers, then the project increases both efficiency and equity and should be adopted. Often, however, matters are more complicated. For instance, the poor and the rich may be worse off, but middle-income individuals better off. How do we assess such a change? Again, we turn to our social welfare function to add up the effects. The **weighted net benefits approach** assigns weights to the net gains of different groups to summarize the impacts in a single number. The social welfare function tells us how to do that. Because of the concern for equity, effects on higher-income groups are weighted less heavily; how much less heavily may determine whether it is desirable to undertake a project. For instance, a project that helps the middle class but hurts the poor and the rich might not be undertaken if we weight the losses of the poor much more heavily than the gains to the middle class.[8]

The use of weights can be thought of as based on three assumptions: first, that there is diminishing marginal utility; second, that different individuals have the same relation between utility and income; and third, that society is concerned with total utility—the sum of the utilities of all individuals (the utilitarian social welfare function). Although each of these assumptions may be questioned, we can also think of these procedures as simply a convenient way to summarize data that decision makers often find helpful.

SOCIAL CHOICE IN PRACTICE

- Identify Pareto improvements.
- If some individuals are better off while others are worse off, identify groups of individuals who are better off and groups that are worse off (by income, region, age), and gains and losses of each major group:

  Ascertain whether aggregate net benefits are positive (compensation principle).

  Look at change in measure of efficiency and measure of inequality, and evaluate trade-offs.

  Calculate weighted net benefits, weighting gains and losses to the poor more heavily than those to the rich, according to the social welfare function.

---

[8] Given the importance of these weights in social decision making, economists have looked for a rational basis for assigning weights. One way is to think about how rapidly marginal utility diminishes with increased income. Inferences about this can be made from observing individual behavior in risky situations: if marginal utility diminishes very rapidly, individuals will be very averse to undertaking large risks, and will be willing to pay large premiums to divest themselves of risk. On the basis of this evidence, most economists argue that a doubling of income will lower the marginal utility of income by a factor of between 2 and 4, so that a change in the income of a middle-class individual with an income of $30,000 should be weighted half to a quarter of the same change in income of a poor individual with an income of $15,000.

THE TRADE-OFF BETWEEN EFFICIENCY AND FAIRNESS REVISITED

The trade-off between efficiency and fairness is based on two notions: the economy is (Pareto) efficient, so no one can be made better off without making others worse off, and transferring income from one person to another is costly. As we noted in Chapter 3, competitive economies (without other market failures) are Pareto efficient. However, the conditions required are highly restrictive; in general, there is room for improvement. There are many Pareto efficient outcomes, one corresponding to each set of initial endowments (of, say, wealth), but, typically, taxing rich individuals to help those less well-off imposes costs, as such taxes may discourage work or savings.

However, increasingly around the world, and especially in the United States, it has become clear that it is possible to have more equality and greater efficiency at the same time. The obvious example is a monopoly (like Microsoft), which raises prices and impedes economic efficiency, but garners huge wealth for the monopolist. More effective enforcement of antitrust laws would have led a more efficient economy with less inequality. This is an example of a wide range of activities, referred to as *rent seeking*, by which individuals increase their income, not so much by adding to the size of the economic pie, but by increasing their share of the economic pie. Other examples include CEOs who use deficiencies in corporate governance (the rules that govern corporations) to pay themselves outsized bonuses; those in the financial sector who engage in predatory lending and other activities that do not improve economic performance, and likely hinder it, but garner for the bankers enormous incomes; corporations that use their political influence to get government contracts that overpay them for what they sell to the government; or corporations that obtain natural resources from the government, paying only a fraction of what those resources are worth. Tax laws that tax capital gains at a low rate, sought after by the rich—a form of income that overwhelmingly accrues to those at the top—distorts the economy. They encourage more resources to go into speculative activities, which can destabilize the economy, rather than into, say, real research that would enhance growth and efficiency. Reforming these laws would enhance both efficiency and equity.

Efficiency and equity can be simultaneously enhanced in other ways. In the United States, a child's life prospects are more dependent on the

income and education of his or her parents than in almost any of the other advanced countries, which means that a large fraction of the children born to poor families are not living up to their potential. Better public schools would create more equality and more equality of opportunity, and improve overall economic performance.

In short, a combination of market and government failures produces an economy characterized by widespread rent seeking and other distortive behavior that benefits a very small number of extremely wealthy citizens at the expense of everyone else. In such circumstances, we can have both a more efficient and productive economy, as well as increased equality. In fact, it is inequality that is reducing the economy's efficiency and lowering its productivity.[9]

---

[9] For an in-depth examination of the relationship between economic efficiency and social equality in the United States today, see J. E. Stiglitz, *The Price of Inequality: How Today's Divided Society Endangers Our Future* (New York: W. W. Norton & Company, 2012).

REVIEW AND PRACTICE

SUMMARY

1. Welfare economics—or normative economics—is concerned with criteria for evaluating alternative economic policies. In general, it takes into account both efficiency and equity.

2. The social welfare function provides a framework within which the distributional consequences of a policy may be analyzed. It specifies the increase in utility of one individual that is required to compensate for a decrease in utility of another.

3. In the utilitarian social welfare function, social welfare is equal to the sum of the utilities of the individuals in society. In the Rawlsian social welfare function, social welfare is equal to the utility of the worst-off individual in society.

4. The concept of consumer surplus—how much individuals are willing to pay for a project or program in addition to what they have to pay—is used to measure the aggregate benefits of a project

or program. The concept of deadweight loss is used to measure the inefficiency of a tax; it asks how much extra revenue could have been generated by a lump-sum tax that would have left individuals just as well off as the tax that was imposed.

5. As a practical matter, in evaluating alternative proposals we do not detail the impact each proposal has on each individual in society, but rather we summarize its effects by describing its impact on some measure of inequality (or on some well-identified groups) and describing the efficiency gains or losses. Alternative proposals often present trade-offs between efficiency and distribution; to get more equality, one must give up some efficiency. Differences in views arise concerning the nature of the trade-offs (how much efficiency one needs to give up to get some increase in equality), and values (how much efficiency one should be willing to give up, at the margin, to get some increase in equality).

6. Three approaches for making social choices when there is not a Pareto improvement are:
   a. the compensation principle;
   b. trade-offs across measures of efficiency and equality; and
   c. the weighted benefits approach.
7. The poverty index measures the fraction of the population whose income lies below some threshold.

KEY CONCEPTS

Compensated demand curve
Compensation principle
Consumer surplus
Dalton–Atkinson measure of inequality
Deadweight loss (excess burden)
Diminishing marginal utility
Gini coefficient
Harberger triangle
Income effect
Interpersonal utility comparisons
Lorenz curve
Lump-sum tax
Marginal utility
Poverty gap
Poverty index
Rawlsian social welfare function
Social indifference curves
Social welfare function
Substitution effect
Trade-off
Utilitarians
Utility function
Weighted net benefits approach

QUESTIONS AND PROBLEMS

1. Assume that Crusoe and Friday have identical utility functions described by the following table.

Utility Functions for Friday and Crusoe

NUMBER OF ORANGES	UTILITY	MARGINAL UTILITY
1	11	
2	21	
3	30	
4	38	
5	45	
6	48	
7	50	
8	51	

Draw the utility function. Fill in the marginal utility data in the table above, and draw the marginal utility function.

2. Assume that there are eight oranges to be divided between Friday and Crusoe. Take a utilitarian view—assume that social welfare is the sum of the utilities of the two individuals. Using the data from Problem 1, what is the social welfare corresponding to each possible allocation of oranges? What allocation maximizes social welfare? Show that it has the property that the marginal utility of an extra orange given to each individual is the same.

3. Now take a Rawlsian view and assume that the social welfare function is the level of utility of the individual with the lowest utility level. Using the data from Problem 1, and again assuming there are eight oranges, what is the social welfare associated with each allocation of oranges? What allocation maximizes social welfare?

4. Draw the utility possibilities schedule based on the data from Problem 1. Mark the points that maximize social welfare under the two alternative criteria from Problems 2 and 3.

5. Assume that Crusoe's and Friday's utility functions are described in Problem 1. Assume now, though, that initially Crusoe has six oranges and Friday two. Assume that for every two oranges taken away from Crusoe, Friday gets only one, an orange being lost in the process. What does the utility possibilities schedule look like now? Which of the feasible allocations maximizes

social welfare with a utilitarian social welfare function? With a Rawlsian social welfare function?

6. If marginal utility did not decrease at all for both Friday and Crusoe, what would the utility possibilities schedule look like?

7. Consider an accident in which an individual loses a leg. Assume that it lowers the individual's utility at each level of income but increases his or her marginal utility (at each level of income), though only slightly. Show diagrammatically the utility functions before and after the accident. Show that if you were a utilitarian, you would give more income to the individual after the accident, but that even after the transfer, the individual who had the accident is worse off than before. Show the compensation that a Rawlsian would provide. Is it possible for a utilitarian to give more to the individual who had experienced the accident than a Rawlsian? Under what circumstances would a utilitarian give nothing to an individual who had experienced an accident?

8. For each of the following policy changes, explain why the change is or is not likely to be a Pareto improvement:

   a. Building a park, financed by an increase in the local property tax rate.

   b. Building a park, financed by the donation of a rich philanthropist; the city acquires the land by exercising the right of eminent domain.[10]

   c. Increasing medical facilities for lung cancer, financed out of general revenues.

   d. Increasing medical care facilities for lung cancer, financed out of an increase in the cigarette tax.

   e. Replacing the system of agricultural price supports with a system of income supplements for poor farmers.

   f. Protecting the automobile industry from cheap foreign imports by imposing quotas on the importation of foreign cars.

   g. Increasing Social Security benefits, financed by an increase in the payroll tax.

   h. Replacing the primary reliance at the local level on the property tax with state revenues obtained from an income tax.

   i. Eliminating rent control laws.

   In each case, state who the losers (if any) are likely to be. Which of these changes might be approved under the compensation principle? Which might be approved under a Rawlsian social welfare function?

9. Assume you are shipwrecked, and there are ten of you in a lifeboat. You know that it will take ten days to reach shore and that there are rations for only ten person-days. (The ration is the minimum amount needed for survival.) How would a utilitarian allocate the rations? How would a Rawlsian? Some people think that even Rawlsian criteria are not sufficiently egalitarian. What might an extreme egalitarian individual advocate? What does Pareto efficiency require? In each case, state what assumptions you need to make to make the decision.

---

[10] The right of eminent domain gives public authorities the right to take property, with compensation, for public uses.

APPENDIX: ALTERNATIVE MEASURES OF INEQUALITY

In the text, we introduced the two most commonly used measures of inequality. These measures are criticized, however, for focusing exclusively on the impact on the very poor. In this appendix, we discuss two more-inclusive measures.

THE LORENZ CURVE

Economists often represent the degree of inequality in an economy by a diagram called the **Lorenz curve**, shown in Figure 7.12. The Lorenz curve shows the cumulative fraction of the country's total income earned by the poorest 5 percent, the poorest 10 percent, the poorest 15 percent, and so on. If there were complete equality, then 20 percent of the income would accrue to the lowest 20 percent of the population, 40 percent to the lowest 40 percent. The Lorenz curve would be a straight line, as depicted in Figure 7.12A. On the other hand, if incomes were very concentrated, then the lowest 80 percent might receive almost nothing, and the top 5 percent might receive 80 percent of total income; in this case, the Lorenz curve would be bowed, as illustrated in Figure 7.12B. When there is a great deal of inequality, the shaded area between the 45-degree line and the Lorenz curve is large (Figure 7.12B). When there is complete equality, this area is zero (Figure 7.12A).

The ratio of the area that lies between the line of perfect income equality and the Lorenz curve over the total area under the line of perfect income equality is a commonly employed measure of inequality, called the Gini coefficient. The lower the Gini coefficient, the more equal the distribution of income: 0 indicates complete equality (every person has the same income) and 1 indicates complete inquality (one person has all the income).

Figure 7.13 shows Lorenz curves for the United States, both before and after the government tax and transfer programs have had their effect. The after-tax curve is decidedly inside the pre-tax, indicating that the combined effect of government redistribution programs is to make incomes more equal than the market would have made them. Thus, while the efficiency costs are less clear-cut, the redistributive gains are undeniable.
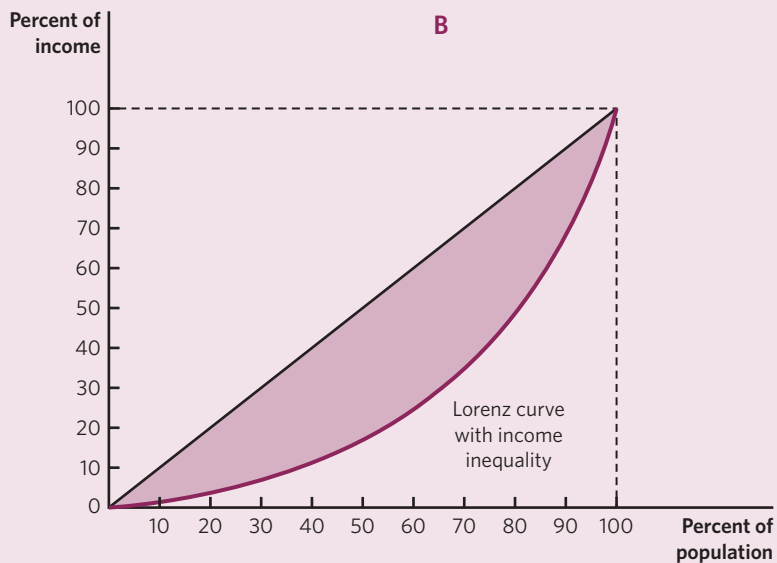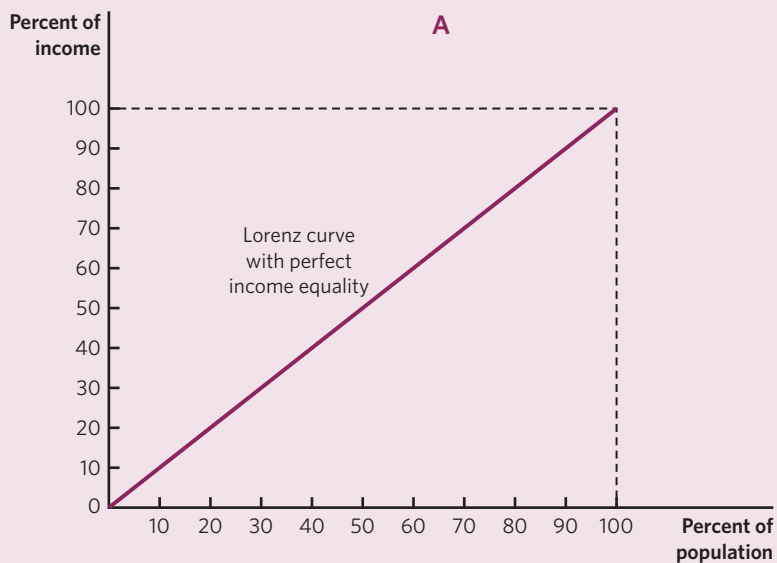
FIGURE 7.12

THE LORENZ CURVE
(A) Shows a Lorenz curve for an economy in which income is evenly distributed. The bottom 20 percent of the economy has 20 percent of income, the bottom 40 percent has 40 percent of income, and so on. (B) Shows a Lorenz curve for an economy in which income is unequally distributed. The curvature of the line indicates that now the bottom 20 percent has less than 20 percent of income, the bottom 40 percent has less than 40 percent of income, and so on.
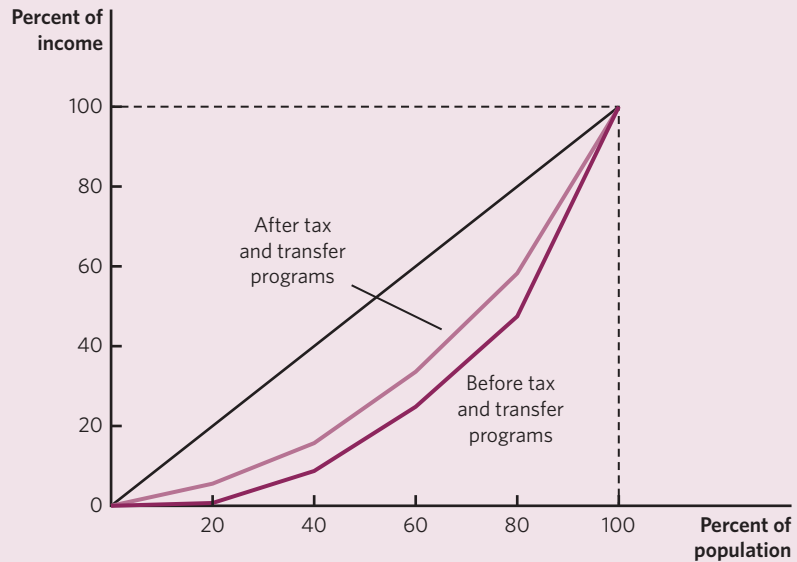
THE DALTON–ATKINSON MEASURE

There is another measure, first introduced by Sir Hugh Dalton, a professor of public finance at the London School of Economics who went on to become the Chancellor of the Exchequer for the United Kingdom. This measure was based on the premise that societies prefer more egalitarian distributions. Figure 7.14 shows two distributions of income. In distribution *B*, more of the income is concentrated at the center, and for societies that value equality, this is the preferred distribution. We can ask: If society could move from its current distribution to a situation in which income was completely equally distributed, what fraction of its total income would it be willing to give up? This fraction is called the **Dalton–Atkinson measure of inequality**. Of course, different individuals might have different views on the amounts that society should be *willing* to give up (this says nothing about how much they would *have* to give up to accomplish the redistribution). The amount society would be willing to give up depends on its social welfare function. With a Rawlsian social welfare function, the amount would be much larger than with a utilitarian social welfare function.
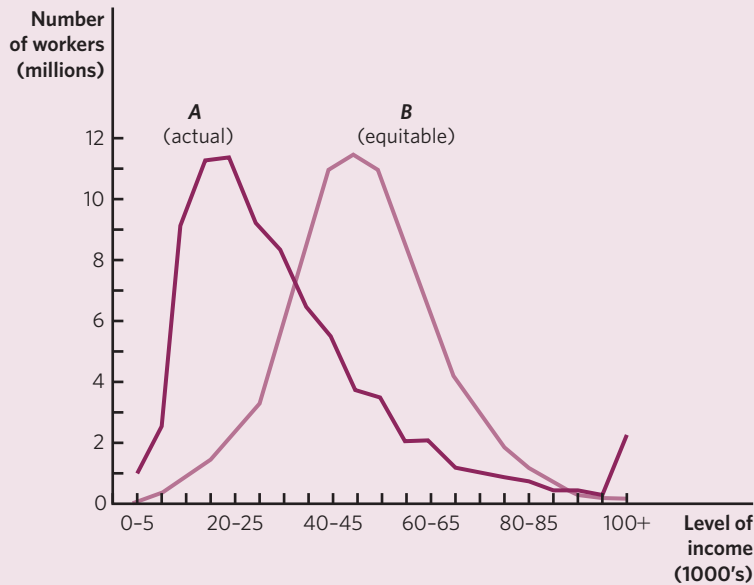
Number of workers (millions)

A (actual)

B (equitable)

12, 10, 8, 6, 4, 2, 0

0–5   20–25   40–45   60–65   80–85   100+

Level of income (1000's)

Anthony Atkinson, of Nuffield College, Oxford, argued that the amount was significant, often between a quarter and a third of total income in more developed economies. Changes in the Dalton–Atkinson measure can be used to assess the impact on inequality of any proposed government program.[11]

---

[11] Formally, the Dalton–Atkinson measure can be defined as follows.

Assume a utilitarian social welfare function

$$W = U(Y_1) + U(Y_2) + U(Y_3) + \cdots$$

and let Y be the average income. Then the Dalton–Atkinson measure D is given by

$$U[(1 - D)Y] = U(Y_1) + U(Y_2) + U(Y_3) + \cdots.$$

The measure clearly depends on the utility function. Atkinson, in his analysis, focused on constant-elasticity utility functions, which have the form

$$U = \frac{Y^{(1-\alpha)}}{1 - \alpha}.$$

He used values of α between 1 and 2.

PUBLIC EXPENDITURE THEORY

Part Three covers the basic theory of public expenditures. Chapter 8 is concerned with the government as a producer. It asks, for instance, whether there are reasons one might expect the government to be less efficient than private firms in the production of goods and services. It also explores the trade-offs entailed in privatization, as well as alternative institutional models for public goods production such as government corporations, performance-based organizations, and public–private partnerships. Chapter 9 focuses on public choice theory in explaining how the level of expenditures on public goods is determined, with particular emphasis on the consequences of majority voting.

8

PUBLIC PRODUCTION OF GOODS AND SERVICES

This chapter is concerned with the role of government in production. Chapter 4 identified several market failures. When there is a market failure, some form of government intervention is required. For instance, Chapter 5 explained why there will be an undersupply of public goods, and described the efficient level of provision of public goods. But government does not have to produce these goods; all it has to do is pay for them. There are many public goods that are privately produced. There are also many private goods that are publicly produced, such as postal services and utilities.

Even though market failures provide a rationale for some form of government intervention, they do not, by themselves, provide a rationale for government production; yet there are certain areas in which government production dominates, and others in which it is very commonly used. For instance, with few exceptions, governments have not relied on mercenary armies. In most countries, governments run the school systems, and, in almost all countries, the postal system. Until a few years ago, in most countries governments also ran public utilities such as telecommunications.

Two common threads run through these examples. First, in many of these cases, competition is not viable. Remember, markets result in

efficiency *when they are competitive*. Historically, only one firm has provided postal services and one company has provided telephone services. Without government intervention one firm, or a few firms, would be able to exercise market power and exploit consumers. Governments have intervened in two ways. In the past, most governments have chosen to take charge of the industry directly, providing telephone services or electricity. As an alternative to producing the goods itself, however, government can regulate private firms—for example, by controlling their prices—to ensure that they do not exercise their monopoly power. In the last several years, there has been a shift away from public production toward private production with regulation. This process of **privatization** has been particularly widespread in the market-based economies of Japan and Western Europe, as well as the transitional economies of Eastern Europe and Asia, especially in utilities (gas, electricity, and telecommunications) and transportation (railroads and airlines). Although some privatizations have been successful, many others have fallen far short of expectations.

The other common thread in many of the examples of government production is that the public interest has many dimensions. Will the actions of profit-maximizing firms reflect these broader public interests? There is often no simple way government can intervene to ensure that they do. This is why the government does not contract out to private firms to run the national defense system. It does contract out specific activities, such as building ships or airplanes, but it does not, for example, say to a private firm, "Run our defense establishment in Europe." Similarly, some believe that schools serve a variety of social functions, which go beyond conveying skills and knowledge. They transmit national values and help form a sense of nationhood. There is concern that a private school system, as effective as it might be in imparting skills, might not work as effectively in advancing this broader set of public objectives. There has also been a resurgence of leaders, particularly in Latin America, who believe that the state can best achieve its social objectives by exerting direct control over key economic sectors. This has resulted in government takeover, or **nationalization**, of a large number of corporations, and in some cases, of entire industries, many which had been previously privatized. For example, in 2006 Bolivian President Evo Morales nationalized firms in oil and gas, telecommunications, and power; and in 2007, Venezuelan President Hugo Chávez nationalized oil, cement, steel, rice processing and packaging, and supermarket businesses.[1]

---

[1] Deliberate nationalization as part of a government's development strategy should not be confused with unintended, de facto nationalization that often takes place amidst an economic crisis—for example, nationalization of some Swedish banks in 1992 and nationalization of several severely distressed financial institutions in the United States in 2008, such as the Federal National Mortgage Association (Fannie Mae) and the Federal Home Mortgage Corporation (Freddie Mac). These government takeovers are meant to be temporary, as indicated by the terms and form of the takeover. There is often a timetable for the public sale of the government's stake, and use of the conservatorship model or purchase of preferred (nonvoting) shares signal the government's desire to refrain from engagement in the business's day-to-day operations.

The global financial crisis of 2008 brought about another wave of nationalizations around the world, especially in finance, as it became clear that without government help, firms would collapse and unemployment would increase still further. In the United States, the government took over the major mortgage financing companies, the country's largest insurance company, the largest auto financing companies, and two of the three major auto companies. In addition, it provided massive funds for banks, although it did not take them over. (Some critics, however, say it should have.) By contrast, in many other countries, including the United Kingdom, the government took over the major banks. There is an important distinction between these nationalizations and earlier ones: for the most part, these were viewed explicitly as temporary measures to be taken in an emergency, until the private economy got back on its feet. However, when these measures were taken, it became clear how dependent the private sector was on the public sector.

Although market impediments (such as limited competition) and concerns about broader objectives provide motivations for public production, there is one compelling argument against public production: often, but not always, governments seem to be inefficient producers. There are examples showing that government enterprises can be efficient: Chile's publicly owned copper mines have been (many believe) just as efficient as the private ones, but provide far higher revenues to the public; Korea's publicly owned steel company was one of the most efficient in the world—far more efficient than America's privately owned steel companies. However, critics of government ownership suggest that they are the exceptions that prove the rule. Thus, identifying when government should engage in production and when it should use private firms involves a balance. To better understand the nature of the balance, we begin by taking a look at government production of private goods, for which issues of "social values" enter in a much more limited way. The primary concern is that the private good—electricity, mail service, or telephone service—be provided in the most efficient way at the lowest possible cost to consumers.

NATURAL MONOPOLY: PUBLIC PRODUCTION OF PRIVATE GOODS

The most important market failure that has led to public production of private goods arises when markets are not competitive. This provides at least part of the explanation for government production in postal

services, telecommunications, water, harbors, and electricity. As we saw in Chapter 4, a common reason why markets may not be competitive is the existence of increasing returns to scale; that is, the average costs of production decline as the level of production increases. In that case, economic efficiency requires that there be a limited number of firms. Industries for which increasing returns are so significant that only one firm should operate in any region are referred to as **natural monopolies**.

Water is a good example. The major cost associated with delivering water is the network of pipes. Once pipes have been installed, the additional costs of supplying water to one extra user are relatively insignificant. It would clearly be inefficient to have two networks of pipes, side by side, one delivering to one home, the next to a neighbor's. The same is true of electricity, cable TV, and natural gas.

THE BASIC ECONOMICS OF NATURAL MONOPOLY

The average cost curve and the demand curve for a natural monopoly are represented in Figure 8.1. Because the average costs of production decline as the level of production increases, it is efficient to have only one firm. In the case depicted, there is a whole range of viable outputs for which the

FIGURE 8.1

NATURAL MONOPOLY

With no sunk costs and potential entry, a natural monopolist would operate at Q_1, the lowest price consistent with at least breaking even. With sunk costs, the price will be higher. The monopolist unconcerned with the threat of entry operates at Q^*, where marginal revenue equals marginal cost.
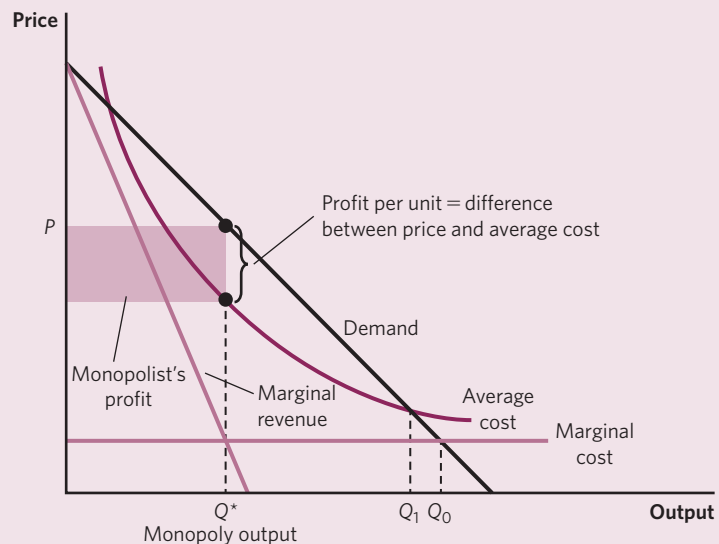
firm makes a profit. The maximum viable output (without subsidies) is Q_1, where the demand curve intersects the average cost curve.

In these situations, we cannot rely on the kinds of competitive forces that we discussed earlier to ensure that the industry operates at the efficient level. Efficiency requires that price equal marginal cost (at quantity Q_0). If the firm charges a price equal to marginal cost, however, it will suffer a loss, as marginal cost is lower than average cost for industries with declining average cost.

One common recommendation in this situation is for the government to provide a subsidy to the industry and insist that the firm charge a price equal to the marginal cost. This policy, however, ignores the question of how the revenues required to pay the subsidy are to be raised; in particular, it assumes that there are no costs associated with raising this revenue. Moreover, it assumes that the government knows the magnitude of the subsidy that will enable the firm to be viable.

In practice, most governments have attempted to make such industries pay for themselves. (They may also be concerned with the equity of making general taxpayers pay to subsidize a private good that is enjoyed by only a portion of the population, or enjoyed by different individuals to different extents.) Thus, they have insisted on government-managed natural monopolies operating at the intersection of their demand curves and their average cost curves (Q_1 in Figure 8.1). This is called the **zero profit point**.

The zero profit point is precisely the point at which natural monopolies may operate, under the assumption that there is effective potential competition. Assume a firm tried to charge a price that exceeded the average cost of production. If it were easy to enter (and exit) from an industry, a firm that tried to capture a profit for itself would instantaneously be threatened with entry by other firms willing to provide the given service or commodity at a lower price. New firms could come in and provide the services or commodities at a profitable price, without worrying unduly about the reactions of the original firm.[2] Thus, the presence of a single firm in an industry does not, in itself, imply that the firm can exercise monopoly power. As long as there are potential entrants, that single firm must charge a price equal to average cost.

---

[2] In the literature on industrial organization, markets with decreasing average costs but no sunk costs, in which price is maintained at a level equal to average costs, are referred to as *contestable*. See W. J. Baumol, J. Panzar, and R. Willig, *Contestable Markets and the Theory of Industrial Organization* (New York: Harcourt Brace Jovanovich, 1982). For a simple exposition of the theory of contestable markets, see W. J. Baumol, "Contestable Markets: An Uprising in the Theory of Industry Structure," *American Economic Review* 72 (1982): 1–15.

EFFECTS OF SUNK COSTS   All this changes when there are **sunk costs**. Sunk costs are costs that are not recoverable upon the exit of the firm. Most research and development expenditures represent sunk costs—but a building that can be converted costlessly for another use does not represent a sunk cost. An airplane, for example, which can easily be sold to another airline, does not represent a sunk cost.

Why are sunk costs so important? They create an essential asymmetry between a firm that is established in an industry and one that is not. The potential entrant is not in the same position as the firm already in the industry, for the firm already in the industry has expended funds that it cannot recover. In deciding whether to enter, a firm does not look at the level of current profits and prices, but at what prices and profits will be *after* entry. Even if prices currently are considerably above average costs (so profits are large), a potential entrant may well believe that the firm already in the industry will respond to entry not by exiting the industry but by lowering its price; at the lower price, entry no longer is profitable. Moreover, when sunk costs are significant, an entrant worries that it will not be able to recover all the expenditures it makes upon entry. Thus, it will be reluctant to gamble that the current firm will either exit or leave its prices at their currently high levels. Accordingly, sunk costs act as a barrier to entry and allow the established firm a degree of monopoly power that it could not exercise otherwise.

Because virtually all natural monopolies entail important sunk costs, the government cannot simply rely on the threat of potential competition. The fact that a single firm controls consumers' water or electricity gives rise to concern: the monopolist is in a position to exploit its consumers. The monopolist that is unconcerned about entry by other firms charges a price that maximizes its profits, the price at which the marginal revenue it gets from selling an additional unit is equal to the marginal costs (output Q^* in Figure 8.1). Its profit per unit of output is the difference between the price it charges and the average costs.

MULTIPRODUCT NATURAL MONOPOLIES   So far, we have focused on a natural monopoly producing a single commodity. If the industry is not to be subsidized, it must charge a price in excess of marginal cost.

On what principle should prices be set when a natural monopoly produces several commodities? Prices, on average, will still need to exceed marginal cost if the firm is to break even. Should the ratio of the price to marginal cost be the same for all the firm's products? Should higher charges on some services be used to subsidize other services?

For example, the U.S. Postal Service imposes uniform charges for delivering mail, even though the marginal cost of delivering mail to a rural household in North Dakota may be much higher than the cost of

delivering a letter in Chicago. If the post office is to break even, there must be a **cross-subsidy**, a subsidy from one user (product) to another user (product).

The issue is obviously very political; the elimination of cross-subsidies will affect some groups adversely. When pricing decisions are made politically, these groups will attempt to persuade those in charge to lower the prices to them, implicitly raising prices to others.

The analysis of pricing decisions involves both efficiency and distributional considerations. Economists have been particularly concerned with the efficiency costs of politically determined pricing policies. When prices are raised on some service, the consumption of that service declines, but a 1 percent price increase reduces demand more for some goods than for other goods. Goods for which demand is more sensitive to price increases are said to have **price elasticity**. Figure 8.2A shows an inelastic demand curve, for which a change in price does not result in a very large change in consumption, whereas in Figure 8.2B, the demand is very elastic: a change in price results in a large change in consumption.

If a natural monopoly is to break even (without government subsidies), it obviously must charge a price in excess of marginal cost. If the government increased price above marginal cost by the same percentage for all commodities, clearly consumption of goods with elastic demand would be reduced by more than consumption of goods with inelastic demand. Under some circumstances it can be shown to be desirable to charge prices such that consumption of every good is reduced by the same percentage (from what it would be if price equaled marginal cost). If the government wishes to do this, it should increase the price (above marginal cost) more for commodities whose demand is inelastic than for commodities whose demand is elastic.[3]

---

[3] This policy will minimize the deadweight loss resulting from price exceeding marginal cost. The problem of how to set prices for a multicommodity public monopoly was first solved in 1956 by Marcel Boiteux, who served as the director of Electricité de France, the government agency in France responsible for producing electricity. For an English translation, see "On the Management of Public Monopolies Subject to Budgetary Constraints," *Journal of Economic Theory* 3 (1971): 219–240. This question of the determination of prices for different commodities turns out to be equivalent to a similar question posed some twenty-five years earlier by the great British economist Frank Ramsey: If the government must raise a given amount of revenue by distortionary taxation, how should it raise the revenue? Should it, for instance, charge a uniform tax on all commodities, so the ratio of the price to marginal cost would be the same for all commodities? Would there then be no relative distortions? Ramsey showed that, as plausible as that might seem, a uniform tax was not the correct answer; it was preferable to charge a higher tax on a commodity whose demand was inelastic. Both Ramsey and Boiteux ignored the distributional issues that are central to most of the political debate. These were introduced into the analysis by M. Feldstein, "Distributional Equity and the Optimal Structure of Public Prices," *American Economic Review* 62 (1973): 32–36. In the context of taxation, see P. Diamond and J. Mirrlees, "Optimal Taxation and Public Production," *American Economic Review* 61 (1971): 261–278; A. B. Atkinson and J. E. Stiglitz, "The Structure of Indirect Taxation and Economic Efficiency," *Journal of Public Economics* (1972): 97–119; and "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics* 6 (1976): 55–75. Also see Chapter 21 for more on optimal taxation.

FIGURE 8.2

PRICING IN A
MULTICOMMODITY
NATURAL MONOPOLY

(A) With an inelastic demand,
an increase in the price above
marginal cost results in a rela-
tively small decline in output.
(B) With an elastic demand,
an increase in price above
marginal cost results in a
large decline in output.
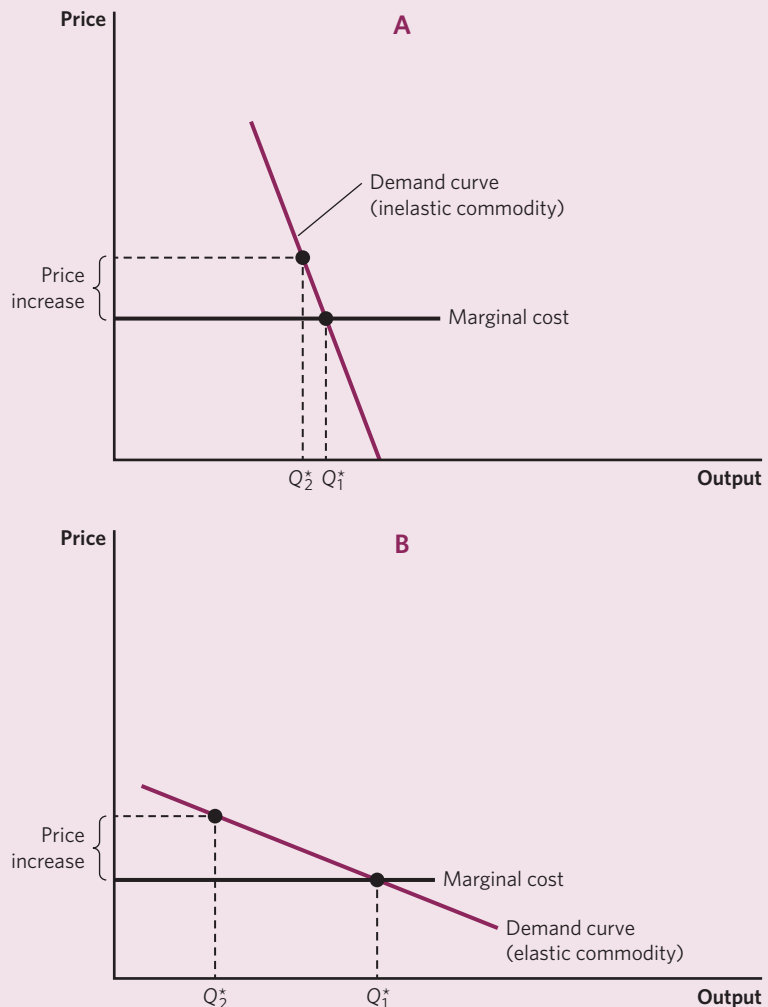
REGULATION AND TAXATION (SUBSIDIES)

When there is a natural monopoly with sunk costs, there is a danger that
the monopolist will take advantage of its position and charge a high price.
One way of addressing this concern, as we have seen, is for government
to take over production. However, there has been increasing concern that
governments do not do a good job at managing production. Rather than
attempting to produce the good itself, the government may leave production
to the private sector, but regulate prices to ensure that the firm does not take

advantage of its monopoly position. Moreover, it can use subsidies to encourage the firm to provide services that might not be profitable privately but are viewed as socially desirable, such as providing postal service to rural areas.

Those who advocate **regulation** and subsidies (or taxes) as remedies for market failures believe that they have three major advantages over public ownership. First, they allow for a more consistent and efficient national policy. Assume that it is desirable to locate firms in areas with high unemployment, and because of this, government-run firms are told to locate to such areas. However, it is better to provide a general subsidy—to firms for which the move has the least cost taking advantage of the subsidy—than to simply impose the burden on firms that happen to be government run.

Second, the utilization of tax and subsidy schemes allows a clearer estimate of the costs associated with pursuing a given objective. It may be desirable to reduce the level of pollution, but how much is it worth? It may be desirable to locate a firm in an area where there is high unemployment, but how much is it worth? It is often difficult to ascertain the additional costs of government enterprises' pursuing alternative objectives; providing direct government subsidies brings the costs more into the open and thus allows a more rational decision concerning whether the costs are worth the benefits.

Third, there is a widespread belief that incentives for efficiency are greater with private firms, even with regulation.

In the case of natural monopolies, the United States has long relied on regulation (although there is some government production, particularly of hydroelectric power and water), in contrast to Europe, which, until recently, relied on government ownership. Regulation is not without its problems: there are significant costs to administering the regulations, prudent implementation requires substantial government capacity, and almost any regulatory scheme gives rise to distortions—that is, deviations in their behavior from what efficient, competitive firms would do— as private firms try to maximize their profits, given the regulatory rules. Thus, if regulations allow a particular return on capital, there may be an incentive to invest too heavily in capital; and if they allow more generous depreciation allowances for one type of capital than another, this may also distort the investment decisions.

In spite of these regulatory problems, beginning in the 1970s and 1980s, there was a major movement throughout the world toward privatization: selling off government-owned enterprises and subjecting them to regulation. By and large, the advantages of gains in efficiency seemed to overwhelm any disadvantages associated with regulation.

In the subsequent years, however, understandings have become more nuanced, as the difficulties of successful regulation and privatizations have become clearer. Special interests try to influence the regulatory

GOVERNMENT PRODUCTION OF PRIVATE GOODS

Examples:	Postal Service, utilities
Rationale:	Market failure—lack of competition associated with natural monopoly
Problems:	Inefficiencies associated with government production
	Cross-subsidization issues
	Intervention of political concerns
Alternatives:	Regulation
	Ignoring problem (all known cures worse than disease)
Current trends:	Privatization
	Focusing government involvement on natural-monopoly core, encouraging competition where feasible
	Public-private partnerships

structure—there have been marked advances in the ability of markets to take advantage of deficiencies in regulations—to the detriment of the public; markets on their own often evolve in a way that undermines competition; and there have been notable failures in privatization.

Deregulation of electricity in California proved to be the quintessential failure: Enron, one of the companies that championed deregulation, managed to manipulate prices, causing massive shortages before it itself went bankrupt—the largest corporate bankruptcy up to that time. Once regulations were restored, the shortages, which conservatives had blamed on *excessive* regulation, disappeared.

Failures often have been due to irreconcilable differences between public policy objectives and private sector priorities. Private prisons, for instance, were more interested in saving on costs than in rehabilitation, because they made more money if those released soon returned to prison. In some cases, such as the privatization of British Rail, there is a broad consensus that consumers have not been served well—prices have increased while the quality of service has deteriorated. Some privatized airports seem more focused on selling goods than in providing facilities for efficient and comfortable arrivals and departures.

In some cases, such as with highways in Argentina, Colombia, and Chile, privatizations cost the government much more than anticipated because contracts were continuously renegotiated and public funds were used repeatedly to bail out franchise holders. In other cases, the privatized enterprises had to receive public assistance, sometimes surreptitiously, such as the U.S. Enrichment Corporation, which enriches uranium.

Many privatizations were unsuccessful because of a failure to understand the potential for abuse of monopoly power. Telecom deregulation did not produce the competitive market that many had hoped, and although airline deregulation resulted in strong competition in some segments of the market, in others, monopolies and duopolies led to high prices. Especially in economies making the transition from communism to the market, monopoly rents were simply transferred from the public to the private sector, many times to former state enterprise managers

who claimed they were now private sector entrepreneurs. The transfer of monopoly rents also occurred in market-based economies—Mexico's privatization of its telecom industry helped create one of the world's richest men. Multiple studies have demonstrated the differential impact of privatization to insiders versus outsiders: those with political or economic connections (insiders) have gained at the expense of those without special access to information or decision makers (outsiders).

To enhance competition, it has become common in the privatization process to differentiate among activities—for instance, the production, transmission, and sale of electricity; the provision of tracks (railroad infrastructure); and the provision of services. Coordination among these pieces often proves difficult, and if one segment is a natural monopoly, such as electricity transmission (the grid), it can exert undue influence in the workings of the system as a whole. Performance has been undermined further by the noncompetetive and nontransparent process by which the privatizations have been undertaken, as well as by incomplete privatization contracts, with unclear metrics for determining achievement of performance benchmarks and ineffective provisions for dispute resolution, resulting in extensive litigation. For example, private water companies have threatened to pull out if what they receive is not increased.[4]

NO GOVERNMENT INTERVENTION

Some economists, including Harold Demsetz of the University of California at Los Angeles and the late George Stigler of the University of Chicago, question whether it might be better in most cases to just let the private sector alone, even with natural monopoly. Monopolists are efficient; the only problem with them is that they charge too high a price, and, accordingly, produce too little. Arnold Harberger, however, in a famous calculation, estimated that the loss from monopoly pricing is relatively small (less than 3 percent of the value of output). Monopolies reduce production relative to the efficient level, but the resources not used by the monopolist go elsewhere in the economy. The loss is the difference in the marginal values of the two uses.

---

[4] For more on privatization of social services, see R. M. Blank, "When Can Public Policy Makers Rely on Private Markets? The Effective Provision of Social Services," *The Economic Journal* 110 (2000): 34–49. Two good studies of privatization to insiders versus outsiders are S. Estrin, J. Hanousek, E. Kocenda, and J. Svejnar, "The Effects of Privatization and Ownership in Transition Economies," *Journal of Economic Literature* 47 (2009): 699–728; and R. Frydman, C. Gray, M. Hessel, and A. Rapaczynski, "When Does Privatization Work? The Impact of Private Ownership on Corporate Performance in the Transition Economies," *The Quarterly Journal of Economics* 114 (1999): 1153–1191. A good general literature review of privatization evaluations is W. Megginson and J. Netter, "From State to Market: A Survey of Empirical Studies on Privatization," *Journal of Economic Literature* 39 (2001): 321–389.

NATURAL MONOPOLY

- Efficient production entails only a single firm.
- Market equilibrium will be characterized by a lack of competition.
- It provides rationale for government production or regulation.

The cumulative loss in efficiency from either regulation[5] or government production, these economists believe, may be much greater.

Most economists remain skeptical. They see larger losses from natural monopolies, partly because managers in industries not subject to competition (and not subject to the scrutiny of regulators) have a tendency to become lax, and partly because, in the absence of competition, incentives to innovate may be limited.

Whether for these or for other reasons, popular support for *eliminating* regulations of natural monopolies remains limited. However, there is still considerable interest in *narrowing* the scope of regulation. It is now recognized that there is more scope for competition than had been previously realized—for instance, several firms now provide cellular telephone service, and there is a multitude of generators of electricity. There is still some natural monopoly—only a single firm provides local landline telephone service or electricity to most homes. Today, regulation is focused on ensuring that there is competition when competition is viable, and that the parts of the system in which there is a natural monopoly do not abuse their monopoly power, either by leveraging their monopoly power to gain further control or by raising prices to enable high rates of return. Major reforms in telecommunications and in electricity regulation in 1996 reflected these changed perspectives.

GOVERNMENT FAILURES

Although market failures led to the institution of major government programs in the 1930s and 1960s, in the 1970s and 1980s the shortcomings of many such programs led economists and political scientists to investigate **government failures**. Under what conditions would government programs not work well? Were the failures of government programs accidents, or did they follow predictably from the inherent nature of governmental activity? Were there lessons to be learned for the design of programs in the future?

There are four major reasons for the systematic failures of the government to achieve its stated objectives: the government's limited information,

---

[5] Stigler has argued, further, that regulation may be ineffective, as the regulated group "captures" the regulators, partly because the only people who are well informed on the highly technical matters of regulation are the regulated parties, and partly because the regulators often get lucrative jobs in the regulated industry after they leave their regulatory positions. Although the regulations may be ineffective in limiting the regulated entity's profit, the regulations may still be highly distortionary. In recent years, there have been problems on the other side: elected regulators may follow populist policies, driving down prices to the point that the utility has little incentive to invest further. George J. Stigler, "Free Riders and Collective Action: An Appendix to Theories of Economic Regulation," *Bell Journal of Economics and Management Science* 5 (Autumn 1974): 359–365; and "The Theory of Economic Regulation," *Bell Journal of Economics and Management Science* 2 (Spring 1971): 3–21.

RENT CONTROL AND AGRICULTURAL PRICE SUPPORTS: CASE STUDIES IN GOVERNMENT FAILURE

Creation of Excess Demand

In the aftermath of the Great Depression and World War II, a housing shortage developed in New York City. The failure to expand supply to keep pace with demand led to an increase in prices, as any economist would have predicted. The political response did not, however, take into account these underlying forces. When lawmakers passed rent control legislation, they failed to anticipate its full consequences, overlooking the fact that apartments were supplied by those who could turn elsewhere for better investment opportunities if the return to investments in housing fell. Advocates of rent control thus failed to anticipate that the supply of rental housing would decrease, and that the quality of services provided by landlords would deteriorate. Even though the government attempted to control this deterioration by imposing standards on landlords, these attempts were only partially successful, and indeed exacerbated the decline in the supply of rental housing. There was little the city government could do to stop this, short of repealing the rent control statutes for new housing, which it eventually did, although numerous older buildings remain under rent control. Many more remain under "rent stabilization" legislation, which controls the rate of increase in rents.

Creation of Excess Supply

After World War II, Japan experienced substantial rice shortages. To ensure food security, the government controlled rice production and marketing until 1995. The result was predictable by economists but unanticipated by Japan's political leaders. Farmers generated large surpluses as they responded to these incentives, while per capita rice consumption dropped dramatically because of higher prices and a more diversified diet. Rather than cut price supports in response to this rice glut, which would have been extremely unpopular with the ruling party's farmer political base, the government decided to reduce rice production *and* maintain high rice prices. It initiated the *gentan* acreage reduction policy in 1970, which pays rice farmers *not* to grow rice by either leaving paddies uncultivated or by switching to crops such as wheat and soybeans (40 percent of rice paddies are now subject to acreage reduction). Rice production has been reduced but chronic surpluses persist; the government buys them to use for animal fodder and food aid. Rice prices have remained high but at the expense of the Japanese taxpayer, consumer, and potential rice exporter: acreage reduction has cost more than ¥7 trillion (more than $80 billion) cumulatively, excluding additional direct income support; the price of rice in Japan is much higher than the world price; and competition from imports has been minimized by imposing a 778 percent tariff. Government incentives have also encouraged inefficient part-time farmers, preventing full-time farmers from expanding their rice paddies to help lower their high production costs. The government is now contemplating abandonment of its production adjustment program and targeting its income support system to full-time farmers.

its limited control over private responses to its actions, its limited control over the bureaucracy, and the limitations imposed by political processes.

1. *Limited information.* The consequences of many actions are complicated and difficult to foresee. The government did not anticipate the precipitous increase in expenditures on medical care by the aged that followed the adoption of the Medicare program. Often, government does not have the information required to do what it would like to do. For instance, there may be widespread agreement that the government should help the disabled, but that those who are capable of working should not get a free ride at public expense. However, limited information on the part of government may preclude it from distinguishing between those who are truly disabled and those who are pretending.

2. *Limited control over private market responses.* The government has only limited control over the consequences of its actions. For example, we noted earlier that the government failed to anticipate the rapid increase in health care expenditures after the adoption of the Medicare program. One reason for this is that government did not directly control the total level of expenditures. Even when it set prices—such as for hospital care and doctors' services—it did not control utilization rates. Under the fee-for-service system, doctors and patients determine how much and what kinds of services are provided.

3. *Limited control over bureaucracy.* Congress and state and local legislatures design legislation, but delegate implementation to government agencies. An agency may spend considerable time writing detailed regulations; how they are drafted is critical in determining the effects of the legislation. The agency may also be responsible for ensuring that the regulations are enforced. For instance, when Congress passed the Environmental Protection Act, its intent was clear—to ensure that industries did not pollute the environment. However, the technical details, such as determining the admissible level of pollutants for different industries, were left to the Environmental Protection Agency (EPA). During the administration of President George W. Bush, there were numerous controversies over whether the EPA had been lax in promulgating and enforcing regulations, thus subverting the intentions of Congress. Two of the most contentious EPA rulings were failure to finalize standards for the allowable concentration of toxic arsenic in drinking water and weak enforcement of the Clean Air Act, particularly for violations by aging electric-generating stations.[6]

---

[6] For a detailed review of EPA enforcement during the administration of George W. Bush, see J. A. Mintz, "'Treading Water': A Preliminary Assessment of EPA Enforcement during the Bush II Administration," *Environmental Law Institute News & Analysis* 34 (October 2004): 10933–10953.

In many cases, the failure to carry out the intent of Congress is not deliberate, but rather a result of ambiguities in Congress's intentions. In other cases, problems arise because bureaucrats lack appropriate incentives to carry out the will of Congress. For instance, in terms of future job prospects, those in charge of regulating an industry may gain more from pleasing members of the industry than from pursuing consumer interests.[7]

4. *Limitations imposed by political processes.* Even if government were perfectly informed about the consequences of all possible actions, the political process through which decisions about actions are made would raise additional difficulties. For instance, representatives have incentives to act for the benefit of special interest groups, if only to raise funds to finance increasingly expensive campaigns. The electorate often has a penchant to look for simple solutions to complex problems; their understanding of the complex determinants of poverty, for instance, may be limited. The revolving door—through which both politicians and bureaucrats move almost seamlessly from the private to the public sector—creates obvious conflicts of interest. Too often, when in public service, they seem to be serving the private interests from which they came and to which they are likely to return.

Critics of government intervention in the economy, such as Nobel laureate Milton Friedman, best known for his advocacy of free markets while at the University of Chicago, believe that the four sources of government failure are sufficiently important that the government should be restrained from attempting to remedy alleged or demonstrable deficiencies in markets. Most countries, however, have devised political processes—rules and regulations governing democracy—that can enhance the performance of the public sector, and that at least ameliorate the problems posed by "government failures."

COMPARISON OF EFFICIENCY IN THE PUBLIC AND PRIVATE SECTORS

Anyone who follows the news media has encountered shocking stories of government inefficiency, often citing misguided procurement policies featuring $1000 toilet seats, $400 hammers, and the like. Similar accounts have been issued from the private sector, such as the quest of cigarette companies to create a smokeless, safe cigarette.

---

[7] This view has been particularly argued by George Stigler. See, for instance, his "Theory of Regulation," *Bell Journal* (Spring 1971): 3–21.s

NATIONAL PERFORMANCE REVIEW

The problems of government inefficiency, particularly arising out of procurement and personnel policies, were the subject of a major review under the direction of Vice President Al Gore. A report was issued in September 1993; its basic theme was that procurement and personnel policies had to be revised substantially in the process of *reinventing government.* The report emphasized greater reliance on performance measures, greater use of market incentives, and greater use of commercial procurement practices:

> Our government, built around a complex cluster of monopolies, insulates both managers and workers from the power of incentives. We must force our government to put the customer first by injecting the dynamics of the marketplace. The best way to deal with monopoly is to expose it to competition.

Elsewhere, the report argued that "competition is the one force that gives public agencies no choice but to improve."

Competition is enough to force the private sector to constantly "reinvent" itself and look for better and more efficient ways of doing what it does. However, it is political pressures—including the pressure of a huge government deficit—that provide the incentive to reinvent government, and it is often political pressure from special interest groups that prevents the process of reinvention from succeeding. There are many examples of this. For more than a hundred years, government has been responsible for meat inspection. Americans want to know that the meat that they are eating is safe. However, when the system of government inspectors was established following Upton Sinclair's graphic description of Chicago's stockyards,* all that was available was a system of visual inspection: Did the meat look rotten? Today, we know that most problems arise from microbes that are invisible to the eye, and there are advanced ways of both detecting and dealing with these microbes. However, despite the existence of systems that could do a better job at lower costs— and of private firms ready and able to perform these tasks—the meat inspectors' union worries that its workers will lose their jobs, and politicians worry that the union will mount a scare campaign, so the inefficient system of visual inspection survives.

---

*Upton Sinclair, *The Jungle* (New York: Doubleday, Page & Company, 1906).

The misallocation of resources attributable to America's financial sector in the years before the crisis was enormous; the full costs of the crisis that it helped bring on are in the trillions—greater than any "waste" from the public sector.

Comparing efficiency of public and private enterprises is typically difficult, because they often work in different parts of the economy and face different constraints: public post offices have an obligation to serve every part of the country, while private companies can select the most profitable ones to serve. When they do produce similar commodities like education, it is hard to measure both the inputs (the quality of students) and the outputs (tests adequately capture only some dimensions of student

achievement; creativity and citizenship values, both of which schools strive to promote, are typically left out). The result is that hard evidence on government inefficiency is difficult to come by.

Although the weight of the available evidence, both in the United States and abroad, suggests that government enterprises are less efficient than their private counterparts, some evidence shows that this need not be the case. The French-run public enterprises have long been held up as models of efficiency. For instance, the French electricity company developed a single design of a nuclear power plant, which they replicated throughout the country, pushing costs significantly below those associated with U.S. nuclear power plants, where there used to be many different designs. Within the United States, state-run liquor stores charge prices that are 4 to 11 percent lower than those charged by private retailers.[8] Administrative costs of the Social Security Administration are less than 1 percent of the benefits paid, but private insurance companies frequently spend as much as 30 to 40 percent of the amount provided in benefits in administrative and sales costs.

Admittedly, it is difficult to measure the productivity of many government workers, those who are engaged in administrative activity; there is no good measure of their output. However, there are some indirect indicators. Since 1992, the number of government public employees in the United States has been brought down dramatically, to the level of the early 1960s, and as a percentage of the civilian labor force, to a level comparable to the early 1930s; in the same period, there has been a huge increase in government service and an increase in the populations served—suggesting an increase in productivity.[9]

The U.S. Postal Service provides an example in the United States of how difficult it is to draw general conclusions. In areas of direct competition, such as overnight mail and parcel post, the Postal Service has not fared well in recent decades. On the other hand, in its main line of business, it has shown remarkable increases in productivity in the past twenty years—three times the pace of business sector productivity.

Of all the comparative productivity studies, one of the most telling was that between private and public railroads in Canada. One of the two major Canadian rail systems is private, the other public. The study concluded that there was no significant difference in the efficiency of the two systems. Evidently, competition between the two provides strong incentives for efficiency in both.[10]

---

[8] Sam Peltzman, "Pricing in Public and Private Enterprises: Electric Utilities in the United States," *Journal of Law and Economics* 14, no. 1 (1971): 109–147. This article discusses both electric utilities and liquor stores.
[9] Prior to the 1990s, productivity growth in the government may also have been greater than in the private sector. See Nancy Hayward and George Kuper's detailed study over the period of 1967 to 1978, "The National Economy and Productivity in Government," *Public Administration Review* 38 (1978); and U.S. Office of Personnel Management, *Measuring Federal Productivity*, February 1980.
[10] D. W. Daves and L. R. Christensen, "The Relative Efficiency of Public and Private Firms in a Competitive Environment: The Case of Canadian Railroads," *Journal of Political Economy* 88 (1980): 958–976.

SOURCES OF INEFFICIENCY IN THE PUBLIC SECTOR

There are several reasons why we might expect public enterprises to be systematically less efficient than private enterprises. These have to do with incentives and restrictions at both the individual and organizational levels.

ORGANIZATIONAL DIFFERENCES

ORGANIZATIONAL INCENTIVES Because public enterprises are not driven by the profit motive, they have little incentive to maximize productivity. Indeed, they are often driven by political concerns that work against productivity, such as providing jobs, especially in regions where there is an unemployment problem. In some countries, public enterprises may not even worry about taking a loss, as they cannot go bankrupt, and any losses are made up out of government revenues. They face, in other words, a **soft budget constraint**, and they often operate in an environment with limited competition.

PERSONNEL RESTRICTIONS Worry that public employees might abuse their position and power—to the detriment of taxpayers who might have to pay more for the services than they should—has resulted in the imposition of numerous constraints. Private firms can hire whomever they like and pay whatever salaries they like. The owners suffer if a firm pays someone more than they are worth, while the taxpayer suffers if a government agency pays someone more than they are worth. We find it particularly objectionable when a public official does not act fairly; equity is an essential part of public trust. Thus, we have imposed strong *civil service* rules, which are designed to ensure that the government hires and promotes the most qualified individuals and that their pay is appropriate. Although such rules serve an important function, they introduce rigidities: it is difficult for a government agency to fire an incompetent worker, and this attenuates incentives. It is difficult for the government to compete with private companies for the best brains; these often command a high wage premium, well beyond the civil service scales for someone with the same qualifications.

PROCUREMENT RESTRICTIONS Similarly, to prevent abuses in the government's purchase of billions of dollars of goods and services every year, procedures have been designed to ensure that the government is not taken for a ride, but their effect is often to raise costs. In buying a jet engine, for instance, these procedures reportedly result in cost increases of as much

as one-third. In many areas, the government insists on competitive bidding. But to do this, the government must specify in minute detail what it is purchasing. A T-shirt may take thirty pages of fine print, detailing the quality of thread, the shape, and so on. Because the specifications required by the government typically differ in several ways from those the T-shirt companies make for the private market, firms will have to have separate production runs to meet the government specifications. Relatively few companies will find complying with all the government regulations worthwhile; competition will thus be restricted, and the prices bid will reflect these high costs of complying with the government specifications and regulations. As a result, the government may end up paying substantially more than it would have to pay for comparable products already available to the public. Off-the-shelf purchasing—a well-informed consumer using the discipline of the market combined with product testing—may save considerable money. Procurement reform enacted in 1994 along these lines has saved billions of dollars.[11]

BUDGETING RESTRICTIONS   Another way government agencies differ from private firms is in budgeting, particularly in making long-term investments. It took the airlines many years and hundreds of millions of dollars to develop their airline reservation systems, but they could easily budget for what they knew was an important capital expenditure. The air traffic control system—which makes sure that airplanes do not crash into each other—is run by the Federal Aviation Administration. Keeping up with the immense increase in air traffic and updating the controllers' obsolete computers will require investments of billions of dollars. Congress makes appropriations only on an annual basis, however, and given the tight budgetary situation, it never appropriates enough to fund fully the rapid modernization. It does not look at the matter as a business would, assessing the return on the investment.

INDIVIDUAL DIFFERENCES

Many of these organizational differences have immediate impacts on individuals. Because individuals cannot be fired, and cannot be rewarded for good performance with the kinds of bonuses that private firms pay, there are neither the carrots nor the sticks to provide as strong individual incentives. Because public agencies have less incentive for efficiency or

---

[11] For instance, within two years of the reform's passage, the Defense Department reported saving $4.7 billion from the new procurement programs (U.S. Department of Defense, "Defense Acquisition Pilot Programs Forecast Cost/Schedule Savings of Up to 50 Percent from Acquisition Reform," News Release No. 138–96, March 14, 1996).

for ensuring that they are attentive to their "customers," what powers to provide incentives they do have often are not directed at those objectives, but rather at more political goals.

There is a whole set of traits that are normally associated with bureaucratic behavior. Bureaucrats may not receive larger paychecks or bigger dividends from increased efficiency, but they often seem to act as if they enjoy the power and prestige associated with being in charge of a larger organization. They thus may try to maximize the size of their **bureaucracy**; if the demand curve for their services has less than unitary elasticity, by reducing efficiency—increasing the price per unit service provided—they actually can increase total expenditures on their agency, and its size (see Figure 8.3). What stops bureaucrats from doing this is competition—competition between bureaucracies. W.A. Niskanen, a member of the Council of Economic Advisers in the Reagan administration and chairman emeritus of the Cato Institute, a conservative think tank, has argued that the increasing centralization of government bureaucracies—in the attempt to ensure that two government agencies do not perform duplicative functions—though intended to enhance efficiency, has reduced competition, thereby giving bureaucrats more scope to pursue their interests at the expense of efficiency and the public interest.

The problem to which Niskanen called attention—that government bureaucrats may act in their own interests, and not necessarily in the interests of the citizens whom they are supposed to serve—is an example of a general class of problems called **principal–agent problems**. The principal–agent problem is simply the familiar problem of how one person gets another to do what he or she wants. Here, the problem is, how do citizens (the "principals") get their employees, public servants (the "agents"), to act in their interests? The analogous problem in the private sector is, how do shareholders (the principals) get their employees, the managers and workers in the firms they own (the agents), to act in the shareholders' interests?

Principal–agent problems arise in all organizations, whether public or private. Managers always face the problem of ensuring that their employees' behavior conforms with their wishes; unless the firm is owned by managers, the owners always have a problem in ensuring that managers act in their interests. The problems of controlling employees are particularly acute in large organizations, and the problem of controlling managers is perhaps as serious in a large corporation in which there is no large shareholder as it is in a government enterprise. When British Petroleum was government owned, it acted little differently from any other large oil company with diffuse ownership, such as Texaco (although in some ways, both may have acted differently from a large oil company controlled by a single family, such as Getty). What difference does it make whether there are private shareholders or a single shareholder—the government? Some
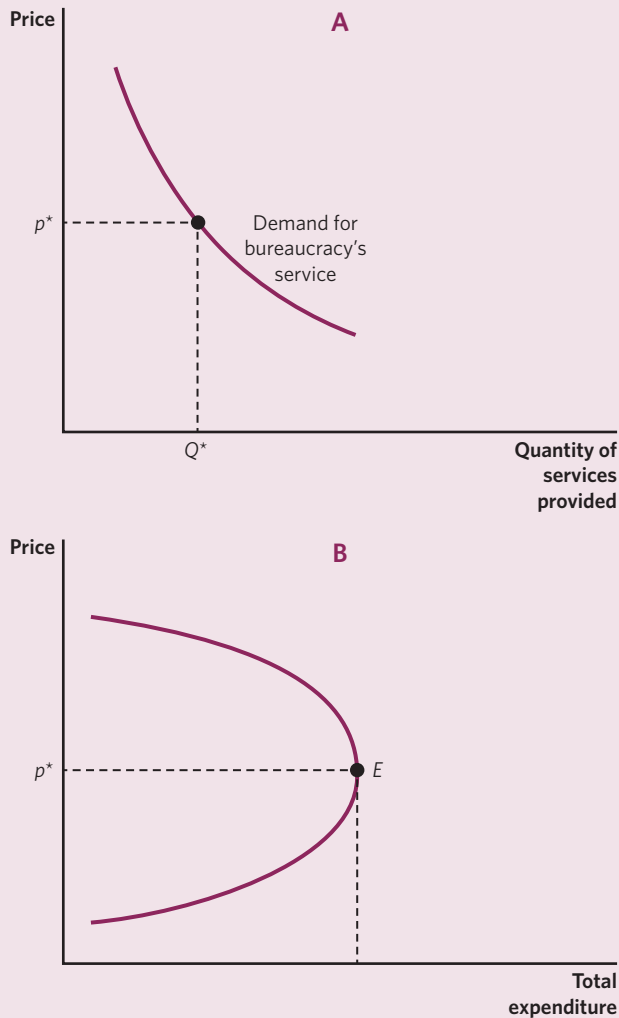
Price | A

p^*

Demand for
bureaucracy's
service

Q^* Quantity of
services
provided

Price | B

p^* • E

Total
expenditure

contend that managers of public enterprises may behave in much the same
way that managers of large private enterprises do. In both cases, manag-
ers have a large amount of discretion, allowing them to pursue their own
interests, often at the expense of the public interest (in the case of public
enterprises) or shareholder interests (in the case of private enterprises).
Payoffs, called "green-mail," out of the corporate purse to those attempt-
ing to take over a firm, have confirmed these views; such payments, as
well as the provisions that management has attempted to put into their
corporate charters making takeovers more difficult, have preserved man-
agement's prerogatives, but at the shareholders' expense.

EXPLANATIONS OF INEFFICIENCY IN THE PUBLIC SECTOR

Organizational differences

- Soft budget constraints (government subsidies, no bankruptcy)
- Role of political concerns
- Absence of competition
- Additional restrictions

    On personnel (civil service—hard to fire, pay competitive wages)

    On procurement

    On budgeting (hard to do long-term budgeting required for large capital investments)

Individual differences

- Absence of incentive pay, difficulty of firing reduces incentives (removes carrots and sticks)
- Principal–agent problems

    Pursuit of bureaucratic objectives—maximizing size of organization

    Excessively high levels of risk aversion—leading to a focus on following procedures (red tape)

In large organizations, principal–agent problems are never fully resolved. Incentive structures—rewards for "good" performance (often financial rewards) and punishments for "bad" performance (being fired)—represent the most effective ways of aligning incentives. Although both public and private sectors face problems in designing incentives that fully resolve agency problems, these problems seem more formidable within the public sector, partly because of restrictions on how public agencies can compensate their employees.

BUREAUCRATIC PROCEDURES AND RISK AVERSION

Bureaucrats' desire to increase the size of their budgets seems to provide an explanation of many aspects of bureaucratic behavior. Other aspects of bureaucratic behavior can best be explained by another important aspect of the incentives bureaucrats are given. Even though bureaucrats' pay may not be closely and directly related to their performance, in the long run, their promotion depends at least partially on observed performance.

Bureaucrats can absolve themselves of responsibility for mistakes by following certain bureaucratic procedures that ensure that all their actions are reviewed by others. Although this process of group decision making also reduces the claims an individual can make for any success, bureaucrats seem willing to make this trade-off. We say they are **risk averse**. This is what gives rise, in part, to the nature of bureaucracies: everything must pass through the appropriate channels (red tape). The emphasis is on procedural compliance rather than quality of results.

Two other factors contribute to the prevalence of bureaucratic procedures. First, many of the costs associated with engaging in risk-averse activities are not borne by bureaucrats themselves. Rather, they are borne by society as a whole, through the taxes required to pay the extra personnel. Further costs are imposed on those dealing with the bureaucracy in the form of delays, paperwork, and so on. (Indeed, there are those who claim that bureaucrats may actually enjoy the bureaucratic process.)[12]

---

[12] There are alternative theories (more psychologically or sociologically based) as to why bureaucrats behave bureaucratically.

Second, the prevalence of set routines that must be followed entailing the approval of any proposal by several different individuals has a positive aspect as well; it is not just a consequence of bureaucrats' pursuing their own self-interest. It follows naturally from the fiduciary relationship between government bureaucrats and the funds they allocate. That is, government bureaucrats are not spending their own money; they are spending public resources. It is generally accepted that an individual should have more discretion—should take greater care—in spending the money of others than in spending his or her own money. Again, what is implied by taking greater care is that certain routines are followed; these ensure that the funds are spent not according to the whim of any single individual. They also reduce the possibility of corruption. Because many individuals must give approval, it is usually not within the power of any individual to give a contract at an above-market price and thus receive a kickback.

It should be clear, however, that both government and large corporations rely on bureacratic procedures. There may thus be a greater difference between small enterprises and large ones, than between publicly owned and privately owned large enterprises.

Two examples of bureaucratic routines commonly used in the public sector are the use of cost–benefit analyses and environmental impact statements. The intent behind such procedures is clear. On the other hand, because the data on which an assessment can be firmly established are rarely available, the studies often become pro forma exercises with predictable outcomes. Occasionally they serve as the basis for attempts by opponents of a project to delay the project and thereby to increase the costs to the point at which the project is no longer economically feasible. There is a social loss in these delaying practices.

CORPORATIZATION

The last section explored a number of the reasons why government enterprises are often less efficient than their public counterparts. We have seen that a host of regulations and restrictions, on hiring and promotion, on procurement, and on budgeting for long-term investment—for all of which there may be good reasons—inhibit efficiency. But enterprises do not have to be turned over to the private sector—privatized—to address these concerns. The United States and other countries have experimented with forms of organization that lie between conventional public agencies and private companies, including **government corporations** and **performance-based organizations**.

There are many examples of government corporations. Among the most widely known are the Postal Service; the Tennessee Valley Authority, the nation's largest producer of electricity; the Saint Lawrence Seaway; and, until it was privatized in 1998, the U.S. Enrichment Corporation (USEC), which converts natural uranium into enriched uranium used in nuclear reactors and atomic bombs. These corporations are owned by the U.S. government, so typically their boards of directors and president are appointed by the President; however, they are intended to be apolitical, with terms of office that do not necessarily coincide with that of the president. Like ordinary firms, they raise their revenues by selling their products or services. Most importantly, they are freed from most of the restrictions imposed on government agencies—they can borrow and lend, and they have considerable discretion in pay and procurement. By and large, they act much like private corporations.

Typically, before a government enterprise is privatized, it goes through the intermediate stage of **corporatization**. This often entails moving the enterprise off budget to increase transparency and accountability. Rather than commingling revenue and expenses with general departmental finances, the enterprise produces its own financial statements to better reflect its true financial position. Subsidies for loss-making government enterprises are also more transparent, as they are explicit on-budget expenditures. Most of the efficiency gains seem to occur in this stage, although there is controversy as to why. Some argue that the freedom from government personnel, procurement, and budget restrictions is all that is required; under corporatization, effective incentive schemes can be put into place. Others argue that without the profit motive—derived from private ownership—these gains could not be sustained. Often the managers of government enterprises do well after privatization, becoming highly paid executives in the new private company and/or receiving hefty shares or options in the newly privatized company—and it is these economic returns that drive them to improve efficiency during the corporatization stage.

Performance-based organizations (PBOs) are government agencies that remain more firmly within the public sector, but in which agency officials are rewarded on the basis of performance. In the United Kingdom the Patent Office became a PBO, and a similar proposal has been made in the United States.

Many countries are involved in a lively debate: When should a government agency be privatized, corporatized, or converted into a PBO? In all three cases, the agency's output must be measurable. In the case of privatization or corporatization, the agency has to produce a product that can be sold, or there at least has to be a source of revenue related to its activities: the TVA sells electricity; the Postal Service, stamps. Many countries have

privatized their air traffic control systems, and the Clinton administration proposed corporatizing it; it could easily be financed either through a ticket tax or through landing or takeoff fees.

Why not privatize everything, or at least everything for which a charge could be imposed? The answer is that there are public objectives that may not be well addressed by a private firm, and that cannot be well addressed through regulation. One might not want a profit-maximizing firm in the business of producing enriched uranium that could be used to make atomic bombs—although, curiously, the Bush administration proposed doing precisely that, and the Clinton administration actually did it in 1998. If the Postal Service were privatized, concern would arise that it might raise prices in an attempt to exercise its monopoly power and that it might not serve rural areas as well. (Although the government might require that it serve those areas, stipulating how it should do so, and then monitoring the quality of the service, might be difficult.)

Issues of public interest are often complex, hard to measure, and difficult to reflect adequately in the design of PBOs. One wants the Patent Office to be not only fast, but also accurate; that is, one does not want it to deny patents that should be granted, or to grant patents that will be overturned by the courts. Unfortunately, the appeal process often takes years; accordingly, it may take years to know how good a job the Patent Office has done. Moreover, there are fundamental *public* issues associated with, for instance, the scope of a patent: Should a patent for a new genetically altered tomato be limited to that tomato breed, or to all genetically altered tomatoes, or to all genetically altered plants? Typically, such decisions are made now on a case-by-case basis. There is concern that if the Patent Office became a PBO, it might not make these decisions in a way that best reflects the national interest.

Privatization has been pushed even further at the subnational level. Some states already have privatized prisons (see case study, "Privatizing Prisons"), and Texas proposed privatizing the administration of welfare programs. The federal government said that Texas could not do so, on the grounds that the private welfare agency might have an incentive to deny benefits to those who were really eligible for benefits. Critics of the federal government's decision argue that an incentive scheme could be found that would address these concerns—for instance, by imposing a sufficiently large fine on the private welfare agency when its denial of benefits was overturned upon appeal. Texas subsequently privatized the state's food stamp eligibility program in 2005, resulting in what the U.S. Department of Agriculture called an unwarranted five-year slide in processing food-stamp applicants that left thousands of people without benefits for which they were eligible.

PRIVATIZING PRISONS

Prisons may seem among the least likely candidates for privatization. After all, a for-profit company would have every incentive to reduce services, such as the quality of food, to the bare minimum; the "customers" might complain, but they really have nowhere else to go. There is also a potential conflict of interest in turning over prison construction and management to corporations whose business model relies on growing prison populations to ensure high occupancy rates in their existing and prospective private prisons. Indeed, although private prisons are not new—they date back to colonial times—they have had a somewhat spotty history. Privatization of prisons has grown dramatically in the United States over the past twenty-five years, so that by the end of 2009, more than 120,000 inmates were held in privately owned or operated federal or state prisons.

It is not surprising that these private prisons have been the subject of considerable controversy. Some claim that costs have been contained, even as services, such as educational training and drug treatment programs, are improved. The leading corporation, Corrections Corporation of America, has done this partly by "reinventing" prisons—for instance, redesigning them so that fewer guards can supervise more people. They also claim that with less tension in the prison, not only are prisoners better off, but fewer guards are required, and absenteeism rates are far lower than in conventional prisons. Critics, however, argue otherwise: that costs are actually higher, violence is increased, and that there is a lower rate of rehabilitation.

Of particular concern is the perverse incentive to which private prisons give rise: they benefit from having legal frameworks with long mandatory jail sentences even for minor infractions.

Despite current levels of overcrowding and projected continued growth in the prison population, severe fiscal constraints will limit the construction of new public institutions. In spite of the controversy over private prisons, it is anticipated that the shortfall will be met largely by privately built or managed prisons.

SOURCES: Charles Thomas, Dianne Bolinger, and John Badalamenti, "Private Adult Correctional Facility Census," Private Corrections Project, Center for Studies in Criminology and Law, University of Florida, March 1997, p. iv; Management & Training Corporation, "Privatization in Corrections: Increased Performance and Accountability Is Leading to Expansion," MTC Institute, December 2009; Lauren E. Glaze, "Correctional Populations in the United States, 2009," *Bureau of Justice Statistics Bulletin* (December 2010); and American Civil Liberties Union, "Banking on Bondage: Private Prisons and Mass Incarceration," November 2, 2011.

Public–private partnerships (PPPs) have evolved over the past two decades as an alternative to purely public, purely private, or corporatized public provision of infrastructure and services, particularly in the power, water, and transportation sectors. PPPs try to allocate risks and rewards between the public and private sectors based on respective missions and capabilities. For example, governments often have land, or the authority to assemble land (when different parcels are owned by different individuals), but are typically short of the resources required to design, finance, construct, and operate extremely large and complex infrastructure improvements, commonly refered to as *megaprojects*. In contrast,

private firms do not have the legal or regulatory authority of government, but can mobilize large pools of funds and expertise.

The most common types of PPPs range from relatively simple service, management, and lease contracts for public services, such as operation of public sports and recreation venues, to very intricate financing and operational arrangements for large-scale public infrastructure such as build–operate–transfer (BOT) concessions and public–private joint ventures. The PPP model has been used around the world for the provision and operation of power plants, toll roads, and health facilities, as well as for the upgrading of muncipal water, sewage, and solid waste systems.

PPPs have produced mixed results. In some electricity projects, the prices paid by consumers have been exorbitant. In some cases, the partnership seems to involve the government bearing the downside risk and the private party getting the upside profits, with the private party threatening to withdraw if its returns turn out too small. The weaker the check of the market on abuses, the less transparent the bidding process, the greater the chances that the PPP will not succeed.[13]

A GROWING CONSENSUS ON GOVERNMENT'S ROLE IN PRODUCTION

The role of the government in production will remain an area of active debate. Today, there is growing consensus that government should not be involved in the production of ordinary private goods, and there is a consensus that government should not privatize national defense (although the United States did hire mercenaries during the Revolutionary War). Even in these areas, however, government can purchase many goods and services from private contractors. It has long purchased airplanes and tanks, but it is turning more to private firms for housing and other logistical support services. There also has been increasing reliance (especially during the Iraq and Afghanistan conflicts) on contractors such as Xe Services LLC (formally Blackwater USA and Blackwater Worldwide) for broadly defined security services, sometimes blurring the line between combat and noncombat support.

Some of the arguments against further privatization are political. Opponents of privatizing the TVA and the other government hydroelectric

---

[13] For an analysis of efforts to evaluate PPPs, particularly in respect to PPP effectiveness and value for money, see G. Hodge and C. Greve, "PPPs: The Passage of Time Permits a Sober Reflection," *Economic Affairs* 29 (2009): 33–39.

ALTERNATIVE ORGANIZATIONAL FORMS

Private firms
Government corporations
Performance-based organizations
Conventional government agencies
Public-private partnerships

- Criteria for privatization:
    Source of revenue (related to its functions)
    Possible to deal with "externalities" and other
        public interest issues such as safety and
        abuse of monopoly power in satisfactory
        manner, as through regulation
- Criteria for performance-based organization:
    Possible to measure performance
    Possible to deal with public interest issues in
        satisfactory manner

projects fear that prices would rise to market level, and a hidden subsidy would be eliminated. Others oppose based on the principle of "if it's not broken, don't fix it": many of these government enterprises seem to be reasonably efficient, and there is no clear case that privatization would improve efficiency. Opposition to privatizing—or even corporatizing—the air traffic control system arises from owners of corporate jets and from general aviation, which currently receive these services without paying their fair share of the costs—they receive a subsidy estimated at as much as $2 billion a year.

In other cases, however, there is real concern that with privatization, the broader range of public objectives would not be pursued—and it is in this arena that the debate is likely to be most lively: Should prisons, welfare agencies, schools, or the production of the material to make nuclear bombs be privatized? Or can most of the gains in efficiency be achieved simply by corporatizing, rewarding performance, and encouraging public agencies to think about those they serve as customers and clients?

To what extent can government, by imitating the private sector—for instance, by making more extensive use of incentive pay—achieve comparable efficiencies? And to what extent are observed inefficiencies in the public sector an inherent consequence of the nature of what the public sector does? It is, for instance, difficult to measure performance in many areas of administrative work; certainly, one does not want to measure performance by the number of pages of paper produced or processed. In many areas, there cannot be competition, or competition might be feasible but undesirable. Do we want two competing armed forces? Two competing judicial systems? To be sure, in some instances competition can be introduced, especially when the government has maintained a state monopoly—the examples in broadcasting, in telephone service, in health, and in education have been promising. Even after competition is introduced, however, there remain vexing questions: How do we ensure that the public and private firms are on a level playing field, that there are not hidden subsidies in the government's operation? The recent financial crisis brought out hidden subsidies even within the private sector: the large ("too big to fail") banks can obtain capital at a low rate, partly because providers of capital feel that it is very likely that, in the event of troubles, they will be bailed out.

There are other concerns besides efficiency. Although the Social Security system has lower transactions costs than private companies providing annuities, this public system also provides a product that the private sector does not—insurance against inflation. In many areas of finance, it has been hard to ensure that the private sector provides the products that help ordinary families manage the risks they face, and that the sector not take advantage of those who are financially less sophisticated. Some call for increased regulation, whereas others believe that government enterprises can, and have, provided financial products that are less exploitive and more effective in meeting needs of ordinary citizens.

Thus, although the presumption remains that for ordinary commodities, the private sector should be relied on, there are a wide range of goods and services for which the debate about how best to provide these services rages: Should there be more reliance on private provision, with strong regulation to prevent abuses, knowing full well that the regulation will be imperfect and may itself introduce distortions? Should there be more reliance on public provision? Should we turn to hybrid models, PPPs, or corporatizations? Each institutional arrangement has its successes and failures. Today, the challenge is to find the best institutional arrangement for a given country in a given instance, at a given moment of time.

REVIEW AND PRACTICE

SUMMARY

1. In the United States, the government has played an important role in production in several sectors, although its role is far more limited than in most other countries. Market failures provide an explanation for government intervention, but not an explanation for government production.

2. Natural monopolies are industries in which it is efficient to have only one producer; in such situations, there is unlikely to be effective competition in the market equilibrium. This lack of competition provides a major reason for government production of private goods. An alternative to public production is government regulation. Whereas the United States has, for the most part, addressed the problem of natural monopolies through regulation, until recently, in Europe, telephone and other natural monopolies were government controlled. More recently, many have been privatized, with mixed results.

3. There is some limited evidence that governments, on average, are less efficient than private enterprises in providing comparable services. There are important exceptions, though, suggesting that government enterprises are not necessarily less efficient than their private counterparts.

4. The government, however, is not necessarily the solution to private sector failures. The failure of many public programs can be attributed to four factors:

   a. The consequences of any action by the government are complicated and difficult to foresee.

b. The government has only limited control over these consequences.

c. Those who design legislation have only limited control over the implementation of the government programs.

d. Politicians may act to further special private interests; more generally, political processes are complicated and need not yield efficient outcomes.

5. Government enterprises differ from private enterprises in several respects: whereas private enterprises maximize profits, government enterprises may pursue other objectives; government enterprises often face soft budget constraints and limited competition; and they face additional constraints, in personnel policy (pay and firing), procurement, and budgeting. Although there may be good reasons for these restrictions, they nonetheless interfere with economic efficiency.

6. These differences lead to differences in individual incentives. Bureaucrats often try to maximize the size of their organization and to avoid risk.

At the same time, public organizations share with private firms the principal–agent problem, the problem of ensuring that their employees act in the interests of the organization, or more broadly, that managers and workers in firms act in ways that are congruent with the interests of shareholders, and that public servants act in ways that are congruent with the interests of citizens.

7. A number of organizational forms lie between conventional public agencies and private corporations, including government corporations and performance-based organizations. They may be able to achieve many of the efficiency benefits of private organizations, and at the same time pursue public interests more effectively than purely private firms subject to regulations. Much of the debate in the future will be about the extent of utilization of these organizational forms and whether private firms should enter into areas that were previously thought of as core government functions, such as prisons and social services. A hybrid public–private partnership model has emerged as a way to take advantage of the relative strengths of government and business while mitigating their respective weaknesses.

KEY CONCEPTS

Bureaucracy

Corporatization

Cross-subsidy

Government corporations

Government failures

Nationalization

Natural monopolies

Performance-based organizations

Price elasticity

Principal–agent problems

Privatization

Public–private partnerships (PPPs)

Regulation

Risk averse

Soft budget constraint

Sunk costs

Zero profit point

QUESTIONS AND PROBLEMS

1. In the past three decades there has been extensive privatization of public enterprises. The U.S. government sold Conrail, the French government sold off many of its banks, and the British government partly sold off its telephone services. In each of these cases, outline the major arguments in favor of and against privatization. Do you feel differently about the three cases? Why?

2. Under the Reagan administration, the Interior Department greatly increased the rate at which it leased offshore oil and gas. This had the effect of significantly reducing the prices that the government received for these leases. (Although the leases are sold by auction, there was only a single bidder for more than two-thirds of the tracts.) Discuss the distributional and efficiency consequences of this policy.

3. The Postal Service claims that one reason why it cannot provide services as cheaply as private firms is that it is required to provide services to rural areas but cannot charge them more than the urban areas. The private companies "skim" the low-cost markets. (Effectively, urban areas are subsidizing rural areas.) Discuss the efficiency and equity consequences of this kind of cross-subsidization.

   Some have argued that if it is desirable as a matter of national policy to subsidize rural post offices, the subsidies should be paid out of general tax revenue, not by the other users of the postal system. Discuss the advantages and problems of such an alternative subsidy scheme.

4. There are many private security firms, and many large housing developments have police protection provided by such private firms. Few towns, though, contract out their police departments. Why do you think this is so? What would be the advantages and disadvantages of doing so? Recently, however, many communities have contracted to have their prisons run by private firms. What advantages or problems might you anticipate from this?

5. The military buys most of its equipment from private contractors but does not use private contractors to staff its ships or fly its airplanes. What differences in the nature of the services provided might account for these differences?

6. There have been recurrent proposals for education voucher schemes, in which the government provides a voucher that can be used to purchase education from either a public provider (the local town) or a private provider. The GI Bill effectively provided such vouchers for veterans of the Korean War and World War II. In the 1996 presidential campaign, Senator Robert Dole proposed that the federal government finance a limited number of such vouchers. What do you see as the advantages and disadvantages of these voucher schemes? Are there some circumstances (some kinds of educational services) for which vouchers seem more attractive?

7. Discuss what organizational form (e.g., private firm, government corporation, or normal government production) you think might be appropriate for each of the following. In each case, discuss problems of designing appropriate incentives and effective regulatory systems.

   a. Public housing

   b. Production of enriched uranium to be used in atomic bombs or nuclear power plants

   c. Production of helium, sometimes used by the government for military purposes

   d. Air traffic control system

   e. The Patent and Copyright Office

   f. Prisons

   g. Job placement services

   h. Administering the welfare program

   i. Administering the food stamp program

8. Recently, there have been several proposals to privatize the Social Security system; some countries have actually privatized part or all of their social security systems. What arguments might be put in favor or against doing this?

9

PUBLIC CHOICE

Unlike expenditures on conventional private goods, which are determined through the price system, expenditures on public goods are determined through a political process. This chapter examines some models of that political process, bringing us to the border between political science and economics.

PUBLIC MECHANISMS FOR ALLOCATING RESOURCES

Chapter 3 explained how the market economy depends on the price system to arrive at efficient resource allocations in the production of private goods. The price system provides incentives for firms to produce goods that are valued and a basis for allocating among consumers the goods that are produced. We often speak of the important role prices play in conveying information, from consumers to producers concerning the value they

attach to different commodities, and between producers and from producers to consumers concerning the costs of production and the scarcity of these commodities.

Equilibrium in private markets is determined at the intersection of the demand curve and the supply curve. When, for one reason or another, the demand for a commodity increases, its demand curve shifts up, the price rises, and this induces firms to produce more. Thus, the price system conveys information about a change in individuals' tastes to firms. Similarly, when, for one reason or another, a commodity becomes less costly to produce, its supply curve will shift down, the price will fall, and individuals will be induced to shift their consumption toward the now-cheaper commodity. Again, the price system has conveyed information about the change in technology from firms to consumers. Indeed, one of the central results of modern welfare economics, as was pointed out in Chapter 3, is that in a competitive economy, the resulting resource allocations are efficient.

Decisions about resource allocations in the public sector are made in quite a different manner. Individuals vote for elected representatives; these elected representatives, in turn, vote for a public budget; and the money itself is spent by a variety of administrative agencies. Thus, there is a major difference between how an individual decides to spend his or her own money and how, for example, Congress decides to spend the public's money. The vote of a member of Congress is supposed to reflect the views of constituents, not just the representative's own views. In deciding how to vote, members of Congress face two problems: first, they must ascertain the views of their constituents, and second, because these views are likely to differ, they must decide how much weight to assign to various positions.

THE PROBLEM OF PREFERENCE REVELATION

Individuals may express their views about the desirability of one private good versus another by a simple action—by buying the good or not—but there is no comparably effective way that individuals can express their views about the desirability of one public good versus another. Elections of public officials convey only limited information about voters' attitudes toward specific public goods; at best, they convey a general notion that voters prefer more or less government spending.

Even if individuals were asked directly about their preferences, would they truthfully and meaningfully reveal them? In recent years, politicians have increasingly turned to polls to ascertain voters' preferences, and although they have found the polls to be useful, they have come to treat the results with extreme caution. For instance, in today's environment of

FOCUS QUESTIONS

1. In what ways does *collective* decision making, such as determining the level of public goods, differ from standard decision making within a household? What is the problem of eliciting preferences? When individuals differ in what they want, say, about the level of expenditures on a public good, how are those differences resolved? What is meant by the problem of "aggregating preferences"?

2. Why may there not be a well-defined outcome when majority voting is used to resolve differences in views? Is there any voting procedure that yields a well-defined outcome in all situations? When there is a well-defined outcome, is it necessarily efficient?

3. What are alternative ways for determining the level of public goods expenditures? Are there ways that ensure an efficient level of expenditures on public goods?

4. What are some of the ways in which politics affects the outcomes of public decision making about resource allocation?

exploding government deficits, polls consistently show strong sentiment toward reducing the deficit. However, no consistent picture emerges concerning the trade-offs: some polls suggest that voters would be willing to pay higher taxes or accept expenditure cuts to reduce the deficit, but other polls (confirmed to some extent by voting patterns) suggest otherwise. Although in polls voters consistently say they believe that the government should spend less on assistance to foreign countries, when asked how much should be spent, they give a number considerably in excess of what the United States is currently spending.

Unless individuals are faced with concrete trade-offs, under which they actually have to give up something to get more, say, of another good, it is difficult to get them to think hard about their choices. In addition, there are special problems in inducing individuals to reveal truthfully their preferences concerning public goods. If what they have to pay does not depend on their answer, then there is a tendency to ask for more of the good: one would normally like more public goods as long as one does not have to pay for them. However, if what an individual says affects how much he or she has to pay, there is an incentive for the individual to pretend that he or she enjoys the good much less than he or she really does—the individual knows that the answer will have a negligible effect on the total amount supplied, and he or she would like to be a free rider.[1]

In private decisions, the decision maker knows his or her own preferences. In public decisions, however, the decision maker must ascertain the preferences of those on whose behalf he or she is making the decision. This is the first important difference between public and private resource allocations.

INDIVIDUAL PREFERENCES FOR PUBLIC GOODS

Collective decision making is difficult because different individuals have different views, for instance, about how much should be spent on public goods. They differ for three reasons. Sometimes there are simply differences in tastes. Just as some individuals prefer chocolate ice cream and others vanilla, some people prefer public parks and education, whereas others prefer private goods, such as video games and cars. This book will not have much to say about these matters of tastes.

The other two sources of differences are incomes and taxes. Richer individuals have higher incomes, so normally they prefer to spend more on all goods, both public and private. When the government spends more on

---

[1] Economists have devised some clever ways of getting around these problems *in principle*, but in practice, no government has actually implemented them. See the appendix to this chapter.

public goods, however, richer individuals often have to pay a relatively large share of the additional cost. In the case of private goods, rich and poor individuals typically pay the same price; with public goods, in effect, richer individuals typically have to pay a higher price. The **tax price** is the additional amount an individual must pay when government expenditures increase by one dollar. The tax price multiplied by total government expenditures equals the individual's tax payment. A higher tax price by itself means that richer individuals would want a lower level of expenditures on public goods. With an income effect leading to a higher desired demand, however, and a price effect leading to a lower desired demand, the net effect is ambiguous.

To observe these effects, consider first a situation in which there are N people and each must pay the same amount, regardless of income. Under this system of **uniform taxation**, the tax price is just $1/N$ and the tax payment is G/N.[2] With **proportional taxation**, everyone pays the same percentage of income. The tax price can be easily calculated. If $\overline{Y}$ is *average* income, $N\overline{Y}$ is total income; if t is the tax rate, then $tN\overline{Y}$ is total government revenue. This must equal government expenditures:

$$G = tN\overline{Y}.$$

Thus, the tax rate is

$$t = \frac{G}{N\overline{Y}}.$$

The tax *payment* of an individual with income Y_i is

$$tY_i = \frac{GY_i}{N\overline{Y}}.$$

If government expenditures increase by a dollar, the individual's incremental tax—the tax price—is just $Y_i/N\overline{Y}$. Thus, an individual with average income ($Y_i = \overline{Y}$) faces a tax price of $1/N$, an individual with above-average income ($Y_i > \overline{Y}$) faces a higher tax price, and an individual with below-average income faces a lower tax price.

A *progressive* tax system is one in which tax payments increase more than proportionately with income; a *regressive* tax system is one in which they increase less than proportionately. Accordingly, the tax price for a high-income individual under a progressive tax system is typically greater than $Y_i/N\overline{Y}$.
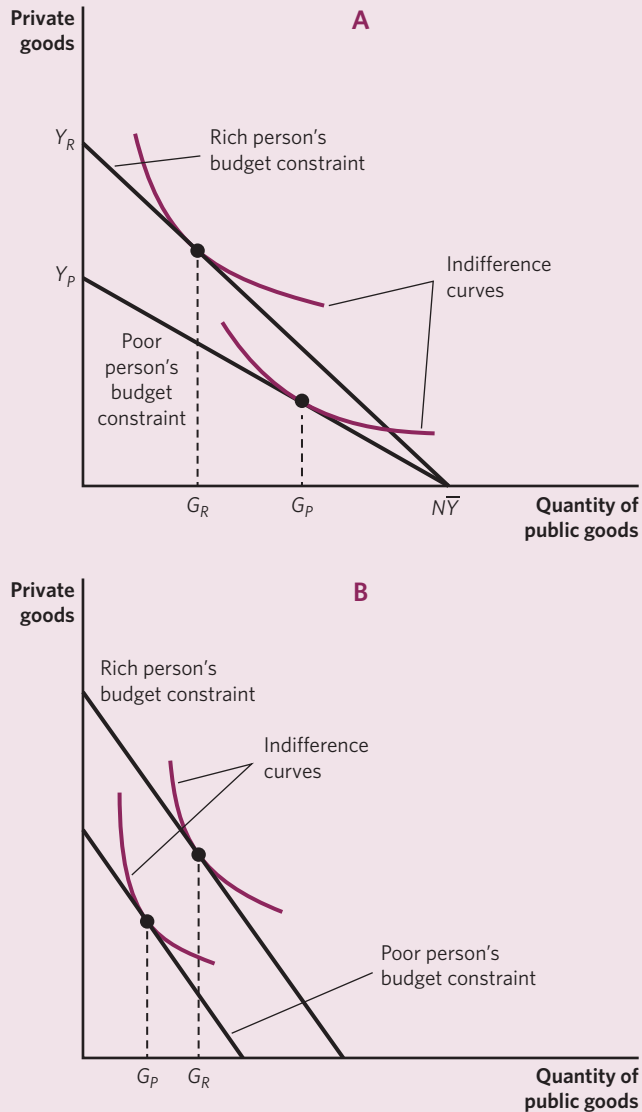
---

[2] No government today imposes uniform taxation—that is, taxes that do not depend on income. However, many "clubs," which can be thought of as voluntary collective organizations, do charge uniform fees, and there are some dedicated taxes (e.g., gasoline taxes that are used for road construction) that do not depend directly on income.

Given the individual's tax price, we can derive his or her preferred level
of public goods expenditure, as illustrated in Figure 9.1. Individuals with
different incomes face different budget constraints; the preferred levels of
public goods expenditure are at the tangencies of the indifference curves
with the budget constraints.

Different individuals will differ with respect to their preferred level
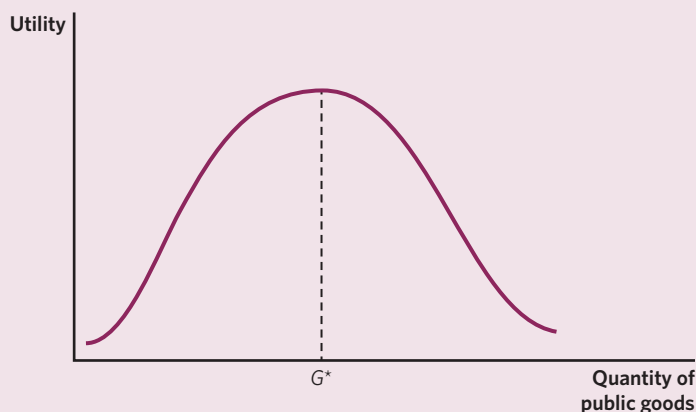of expenditures. With proportional taxation, poorer individuals face

UTILITY DEPENDS ON LEVEL OF GOVERNMENT EXPENDITURE
Utility is maximized at the point of tangency between the indifference curve and the budget constraint. The farther away the actual level of expenditures from the preferred level of expenditures, G^*, the lower the level of utility.

lower tax prices, and on that account, their preferred level of expenditures, G_P, is higher. Poorer individuals have lower incomes, however, and with lower incomes they demand less public as well as private goods. The net effect is ambiguous. Figure 9.1A illustrates the case in which the substitution effect (lower tax price) dominates the income effect, so the poorer individual does prefer a higher level of public goods than the richer person.

With uniform taxation, there is only an income effect, so high-income individuals will prefer higher levels of public expenditure (Figure 9.1B); with progressive income taxation, lower-income individuals will face a lower tax price than with proportional taxation, so their preferred level of expenditures will be even higher than with proportional taxation.

Figure 9.2 shows how utility depends on the level of government expenditures. The individual's most preferred level of expenditures occurs at G^*, but utility is maximized under a budget constraint, at the point of tangency with the indifference curve. The further away the actual level of expenditures is from the preferred level of expenditures, G^*, the lower the level of utility.

Figure 9.3 indicates the relationship between the level of utility and the level of public goods expenditure for three different groups—the rich, the poor, and the "middle," assuming a particular tax system. Each has its own preferred outcome, and utility decreases as expenditures deviate either above or below that level. For expenditures above the middle, the marginal benefits of increased public expenditure are less than the marginal costs the individual bears in additional tax payments, whereas the converse holds for expenditures below the preferred level.

UTILITY AS A FUNCTION OF *G*

The figure reflects the preferences of three groups in the population: the poor, the middle class, and the rich. In this example, the rich prefer higher levels of expenditure to the middle class, who prefer higher levels than the poor.
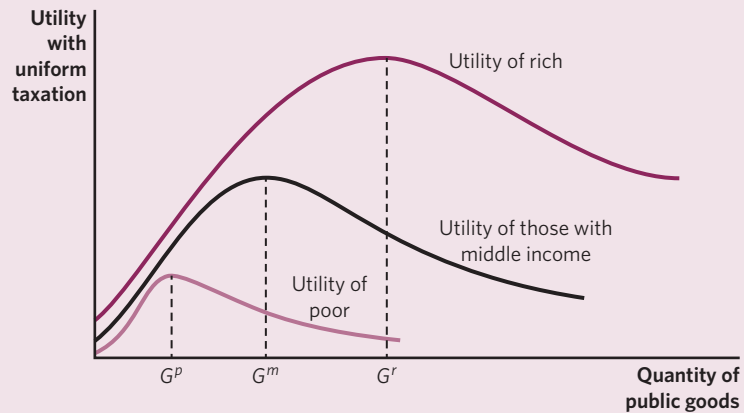
THE PROBLEM OF AGGREGATING PREFERENCES

In the private market, the firm does not have to balance the claims and interests of one group against those of another. If an individual is willing to pay a price for a commodity that exceeds the marginal cost of production, it pays for the firm to sell the commodity to the individual. Decisions are made on an individual basis. In the public sector, on the other hand, decisions are made collectively. For example, when a politician votes to increase expenditure on some public good, it is not as if the politician has to pay for the good himself or herself. The politician's vote is intended to represent the interests of his or her constituents, but their opinions are not likely to be unanimous: some individuals would like more military spending, others less; some individuals would like more expenditures on welfare, others less.

The problem of reconciling differences arises whenever there must be a collective decision. Popular political discussions often refer to what "the people" want. Because different people want different things, however, how can a social decision be made from these divergent views? In a dictatorship, the answer is easy: the dictator's preferences dominate. There is no such easy resolution in

CENTRAL PROBLEMS OF PUBLIC CHOICE

Preference Revelation:

- Ascertaining the desired level of public goods of each individual.

Aggregating Preferences:

- Different individuals have different preferred levels of public expenditure.
- Preferred level depends on both individuals' income and the tax system.
- Other things being equal, rich typically prefer higher levels.
- However, the cost of increased public expenditure may be higher for the rich.

a democracy, though. A number of different voting rules have been suggested, among them unanimity voting, simple majority voting, and two-thirds majority voting. Of these, perhaps the most widely employed rule for decision making in a democracy is simple majority voting.

MAJORITY VOTING AND THE VOTING PARADOX

We all know how majority voting works. Suppose you and two friends are trying to decide whether to go to a movie or a basketball game. You take a vote: if the movie gets two votes, you go to the movie; if the basketball game gets two votes, you go to the game. Sometimes majority voting does not lead to such a clear outcome, though, when there are more than two alternatives. A majority voting equilibrium requires that there is one alternative that can win a majority in a contest against *any* alternative. As early as the eighteenth century, the famous French philosopher Nicolas de Condorcet noted that *there may not exist any majority voting equilibrium.* The problem may be seen in the following simple example, in which there are three voters and three alternatives, denoted A, B, and C. A could be going to a movie, B to a basketball game, C renting a video; or A could be spending more money on health care for children, B reducing the deficit, C cutting taxes.

Voter 1 prefers A to B to C.

Voter 2 prefers C to A to B.

Voter 3 prefers B to C to A.

Assume we vote on A versus B. Voters 1 and 2 vote for A, so A wins. Now we vote on A versus C. Voters 2 and 3 prefer C to A, so C wins. It appears that C should be the social choice. C wins against A, which wins against B. But let us now have a direct confrontation between C and B. Both Voter 1 and Voter 3 prefer B to C. This is referred to as the **voting paradox**, or the *paradox of cyclical voting.* There is no clear winner. B beats C and C beats A but A beats B. Assume we began by saying we were going to first vote on B versus C, and put the winner against A. B beats C, and then A beats B. But just to check that we had made the right decision (A), we decide to put A against C. C beats A. So we think C is the winner. But then we check that by challenging C with B. B beats C—which was our original vote. B again appears to be the winner. But just to check, we again challenge it with A. A again beats B, as we knew from our earlier vote. The voting process thus goes on and on.

Often, to avoid these voting cycles, democracies organize their decision making as a sequence of votes—for instance, first they will

vote A against B, and the winner of that vote will be put against C, with the final determination depending on the outcome of that vote, with no further contests. In that case, it may be very important to control the agenda—the order in which the votes occur. For instance, as just depicted, we know that A would defeat B, and C would defeat A, so C would be the decision. But suppose, instead, we structure the election as first a contest between B and C, and then the winner of that contest against A. A would win that election. Finally, suppose we structured the election as first a contest between A and C, and the winner of that election against B. Then clearly B would win. Thus, the winner of each of these elections is determined solely by the order in which the pairwise comparisons were made.

Note, too, that if individuals realize there is going to be a particular sequence of votes, they may wish to vote strategically. That is, in the first round of the vote, Voter 1 may not vote his or her true preferences on, say, A versus B, but think through the *consequences* of that for the eventual equilibrium. The voter may vote for B, even though he or she would prefer A, knowing that in a contest between C and B, B will win, whereas in a contest between A and C, C might win. Because this voter prefers B to C, he or she votes initially for B.

This analysis leads to two questions. First, are there voting rules that will ensure a determinate outcome for any vote? Second, are there any circumstances under which simple majority voting will yield a determinate outcome? We take up these two questions in the next two sections. It turns out that the voting paradox cannot be resolved through voting rules, but there are indeed circumstances in which majority voting yields clear decisions.

RANK-ORDER VOTING FAILS: INDEPENDENCE OF IRRELEVANT ALTERNATIVES

- John and Jim prefer a swimming pool to a library to a tennis court.
- Tom prefers a tennis court to a swimming pool to a library.
- Lucy and Jill prefer a tennis court to a swimming pool to a library.

Majority voting: Three prefer a tennis court to a swimming pool.

Three prefer a tennis court to a library.

Five prefer a swimming pool to a library.

SWIMMING POOL WINS

Rank order (lowest sum of ranks wins): Tennis, nine; swimming pool, eight; library, thirteen.

TENNIS COURT WINS

Rank order in choice between tennis court and swimming pool (library not an option): Tennis, seven; swimming pool, eight.

SWIMMING POOL WINS

Adding "irrelevant choice"—library—alters the outcome.

ARROW'S IMPOSSIBILITY THEOREM

An endless cycle of voting is clearly an unsatisfactory state of affairs. It is natural to ask, then, whether there is any other political

mechanism—any other set of rules for making social decisions—that eliminates this problem. An ideal political mechanism should have four characteristics:

1. *Transitivity*. If the rule shows that A is preferred to B, and B is preferred to C, then A should be preferred to C. As we have seen, simple majority voting lacks this essential property. Without this property, we can get into cyclical voting.

2. *Nondictatorial choice*. There is a simple way of avoiding voting cycles: give all decision-making powers to a dictator. As long as the dictator has consistent preferences, then there will never be a voting cycle. In a democratic society, however, a meaningful political mechanism must ensure that the outcomes do not simply reflect the preferences of a single individual.

3. *Independence of irrelevant alternatives*. The outcome should be independent of irrelevant alternatives; that is, if we have to make a choice between, say, a swimming pool and a tennis court, the outcome should not depend on whether there is a third alternative, such as a new library.

4. *Unrestricted domain*. The mechanism must work no matter what the set of preferences and no matter what the range of alternatives over which choices are to be made.

In looking for a system that would satisfy all four of these properties, a number of alternative rules have been examined, but each fails one or more of the requirements. For instance, *rank-order voting* in which individuals rank the alternatives, then the ranks assigned by all individuals are added together, and the alternative with the lowest score wins, does not satisfy the "independence of irrelevant alternatives" criterion.

The quest for an ideal system came to an end with the findings of Nobel Laureate Kenneth Arrow of Stanford University. He showed that there was no rule that would satisfy all the desired characteristics. This theorem is referred to as **Arrow's impossibility theorem**.[3]

Arrow's impossibility theorem has one further interesting and important implication. We often hear expressions such as "The government seems to be acting in an inconsistent manner," or "Why doesn't the government determine its priorities and then act on them?" This language personifies the government, treating the government as if it were an individual. Language is important: although we all know that the government is not a single individual, speaking of it as if it were frequently leads us to think of it in such facile terms. We come to expect that government

---

[3] Kenneth Arrow, *Social Choice and Individual Values,* 2nd ed. (New York: John Wiley & Sons, 1963).

should act consistently like a rational individual. Arrow's impossibility theorem, however, suggests that, unless some individual is granted dictatorial powers, the government should not be expected to act with the same degree of consistency and rationality as an individual.

In the earlier example in which there was no majority voting equilibrium, we saw the importance of control of the agenda. We also saw that it can be beneficial for individuals to vote strategically—that is, to vote not according to their true preferences, but to take into account how the outcome of the current vote will affect the final outcome. Just as Arrow established that there does not exist any way of adding together the preferences of different individuals to satisfy all the desired characteristics of a choice mechanism, it has been shown that there does not, in general, exist any voting system[4] in which individuals will always vote their true preferences.

SINGLE-PEAKED PREFERENCES AND THE EXISTENCE OF A MAJORITY VOTING EQUILIBRIUM

Although Arrow's impossibility theorem shows that there is no voting rule that *always* satisfies the desirable properties of a social choice mechanism described earlier, under some conditions, the simple system of majority voting yields a determinate outcome.

Figure 9.3 showed the level of utility as a function of the level of expenditure on public goods. There, each individual has a single peak to his or her preference profile. This property of **single-peaked preferences** is enough to guarantee the existence of a majority voting equilibrium. Note that the peak need not be "interior" but may lie on the "end," so that preferences such as those in Figure 9.4A are also consistent with single-peakedness.

On the other hand, preferences such as those illustrated in Figure 9.4C are not consistent with single-peakedness. Both 0 and G_1 are (local) peaks. Unfortunately, such examples arise naturally in considering many public choice problems.

For instance, consider the problem of an individual's attitudes toward expenditures on public education. If the level of expenditure on public education is below a certain minimum level, a rich individual may prefer sending his or her children to private schools. If so, any increase in expenditure on public schools simply increases his or her taxes; the rich individual gets no direct benefits. Thus, his or her utility decreases with

---

[4] A voting system is any set of voting rules by which a group of individuals tries to reach a decision—for instance, by dropping from consideration the alternative with the lowest number of votes or by giving individuals several votes and allowing them to assign as many as they like to each alternative.
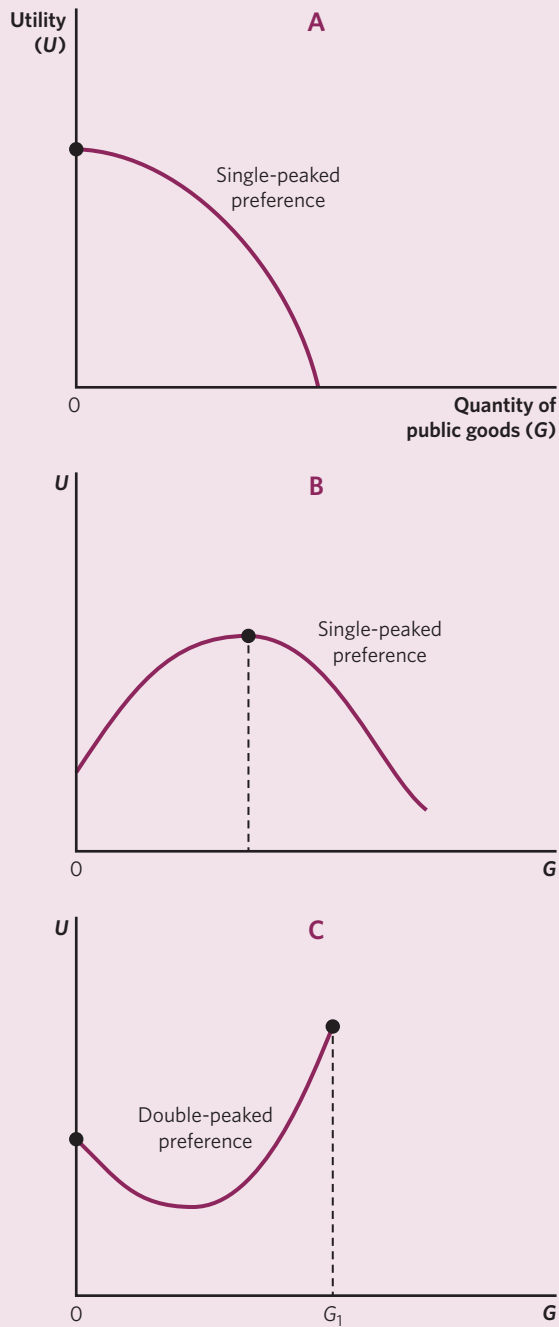
SINGLE-PEAKED AND DOUBLE-PEAKED PREFERENCES
(A and B) With single-peaked preferences, there always exists a majority voting equilibrium. (C) Without single-peakedness, there may not exist a majority voting equilibrium.

TABLE 9.1 ALTERNATIVE TAX SCHEDULES

	FRACTION OF INCOME		
PAID IN TAXES	A	B	C
Poor	20%	18%	17%
Middle	20%	18%	21%
Rich	20%	23%	22%

government expenditures up to a critical level at which the individual decides to send his or her children to public school. For increases beyond that level, the individual derives some benefit. Of course, beyond some point, the increases in taxes more than offset the benefits. For this individual, a high level of expenditure is preferred to no expenditure, but no expenditure is preferred to an intermediate level of expenditure. There may be no majority voting equilibrium in this case.

Although preferences for a *single* public good (with no private good option, unlike education) are usually single-peaked, when we have to rank choices involving more than one public good, those rankings are seldom single-peaked.[5] To obtain single-peakedness, we have to restrict ourselves to voting on one issue at a time.[6]

Equally important, for most distribution issues there will not be a majority voting equilibrium.[7] This can be seen most clearly in considering the structure of income taxation. Suppose we are voting among three income tax schedules that are designed to raise the same amount of revenue. For simplicity, we assume that there are three groups of individuals with equal numbers—the poor, the middle class, and the rich—and that they vote in solid blocks. The three tax schedules are denoted as A, B, and C in Table 9.1. Tax schedule A is strictly proportional; it takes the same fraction of income from each individual. The poor and the middle class then get together and propose tax schedule B. This reduces the taxes they pay but taxes the rich much more heavily. Clearly, tax schedule B will win a majority over A. Now, the rich propose to the poor: "Because you are more needy, why don't we lower your taxes somewhat more? At the same time, we'll adjust the tax schedule at the upper end, to reduce the inequities associated with excessive taxation." Thus, they propose tax schedule C, which, relative to B, lowers the taxes on low and high income and raises them on middle income, so that now both the middle- and upper-income individuals pay a larger proportion of their income in taxes than do the poor.

[5] See G. Kramer, "On a Class of Equilibrium Conditions for Majority Rule," *Econometrica* 41 (1973): 285–297.
[6] See S. Slutsky, "A Voting Model for the Allocation of Public Goods: Existence of an Equilibrium," *Journal of Economic Theory* 11 (1975): 292–304.
[7] See D. K. Foley, "Resource Allocation and the Public Sector," *Yale Economic Essays* 7 (1967): 45–98.

TABLE 9.2 PREFERRED LEVELS OF EXPENDITURE ON PUBLIC GOODS

LUCY	TOM	JIM	JOHN	JILL
$600	$800	$1000	$1200	$1400

Jim is the median voter.

Clearly, tax schedule C wins a majority over B. Now, however, the middle class proposes going back to straight proportional taxation. Because both the upper- and middle-income individuals prefer schedule A, A defeats C. We again get a cyclical pattern of voting.[8]

THE MEDIAN VOTER

When preferences are single-peaked, we asserted that there is a well-defined majority voting equilibrium. What does it look like? And how does it correspond to the *Pareto efficient* equilibrium that we described in Chapter 5?

When preferences are single-peaked, we can rank individuals by their preferred levels of, say, expenditure on the public good, from the individual who prefers the least to the individual who prefers the most. The *median* individual is the individual such that half prefer less and half prefer more than he or she does. In Table 9.2 Jim is the **median voter**. *The outcome of majority voting corresponds to the preferences of the median voter.* In this case, it is Jim's preferred level ($1000) that wins. The reason is simple: if any level of expenditure below $1000 is voted on against $1000, Jim plus all of those who want more than $1000 vote for $1000; because Jim is the median voter, a majority cast their vote for $1000. If any level of expenditure above $1000 is voted on against $1000, Jim and all of those who want less than $1000 vote for $1000. Again, $1000 wins.

THE INEFFICIENCY OF THE MAJORITY VOTING EQUILIBRIUM

Because the median voter determines the level of expenditure on public goods, to ascertain whether there is too much or too little expenditure we

---

[8] If we restrict the set of tax schedules over which voting occurs, for instance, to tax schedules with an exemption level and a fixed marginal tax rate (called *flat-rate tax schedules*), there may be a majority voting equilibrium. See T. Romer, "Individual Welfare, Majority Voting, and the Properties of a Linear Income Tax," *Journal of Public Economics* 4 (1975): 163–185.

need only examine how the median voter votes, and contrast that with the conditions for efficiency discussed in Chapter 5. The median individual is assumed to compare only the benefits he or she receives with the costs that he or she bears. The median voter's benefits are lower than total social benefits (which includes all the benefits that accrue to others), but so are his or her costs. Whether there is too much or too little expenditure on public goods thus depends on whether the median voter's share of total (marginal) costs is less or greater than his or her share of total (marginal) benefits.

For a wide variety of public goods, with proportional or progressive taxation majority voting will result in an oversupply of public goods. To see this, assume there are N individuals. With uniform taxation, the tax price for each individual would be $1/N$; with proportional taxation, it would be $Y_m/\overline{Y}N$, where Y_m is the income of the median voter and $\overline{Y}$ is average income; and with progressive taxation, the tax price would be still lower. With a symmetric distribution of income, as illustrated in Figure 9.5A, the income of the median individual equals the average income—that is, $Y_m = \overline{Y}$. However, most income distributions are skewed, as shown in Figure 9.5B. There are a few very rich individuals, which increases the average income. As a result, average income exceeds the income of the median individual, $Y_m < \overline{Y}$, so that with proportional taxation the tax price is even less than $1/N$.

The question then is: What fraction of the total marginal benefits accrues to the median voter? If the marginal benefits are uniform, then the median voter gets $1/N$ of the total marginal benefits; so, with uniform taxation, the median voter would get $1/N$ of the total social benefits and bear $1/N$ of the total costs. Therefore, the median voter would vote for an efficient level of expenditure. With proportional or progressive taxation, however, the median voter's share of the cost would be smaller than his or her share of the benefits, and he or she would vote for excessive expenditures—that is, for a level of expenditure at which the sum of the marginal benefits is less than the total marginal cost to society.

Some forms of public expenditures are actually enjoyed disproportionately by the poor: the rich may make less use of public parks because they have large yards of their own. According to the median voter theory, there is an especially large tendency for an oversupply of such goods. In many cases, it is difficult to ascertain the balance: the median voter gets a smaller share of the benefits and bears a smaller share of the costs.

In the examples discussed so far, the median voter corresponds to the voter with the median income. This is often, but not necessarily,
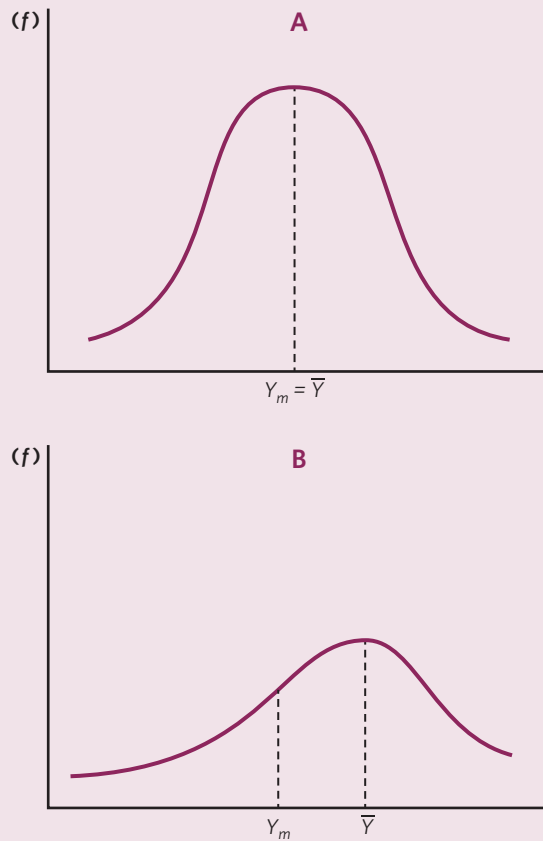
INCOME DISTRIBUTIONS
(A) Shows a symmetric income distribution; the median income, Y_m, equals the average income. (B) Shows a skewed income distribution more typical of the United States. Because there are a few very rich individuals, the average income exceeds the median income.

the case. Higher-income individuals normally demand more of all goods, including public goods, *if they face the same prices*; but the tax price faced by higher-income individuals is typically higher, so in some cases, very–high-income individuals may actually prefer lower levels of public expenditures on certain goods than do middle-income individuals.

Median voter theory says that to understand collective choices, we should focus our attention on the median voter and how a particular policy affects that individual. Consider what the median voter theory says about support for, and consequences of, abolishing the deductibility of state and local income and property taxes for purposes of the federal income tax, a proposal that is frequently raised in the context of tax reform. Most individuals do not itemize; that is, in calculating their taxable income, they

> ## MAJORITY VOTING
>
> Proposal that gets a majority against all other wins.
>
> *When majority voting equilibrium exists, it reflects preferences of the median voter.*
> - In a two-party system, both parties will converge to the position of the median voter.
> - Majority voting equilibrium is not, in general, Pareto efficient.
>
> *Majority voting equilibrium may not exist.*
> - Proposal A defeats B, C defeats A, but B defeats C.
> - Majority voting equilibrium exists if preferences are single peaked.
>
> *Arrow's Impossibility Theorem:* There does not exist an alternative voting mechanism that avoids problems of majority voting (and satisfies certain other desired properties).

subtract a "standard deduction" rather than listing all the allowable deductions, including local and state taxes. Accordingly, they would be unaffected directly, although the increased federal revenue, which could be used to reduce the deficit or taxes, would be a benefit. Thus, the median voter should support this reform. However, it has been defeated repeatedly, suggesting that there are other considerations that go into determining outcomes in political processes, as we note later in this chapter.

One of the worries of state governors has been that if state taxes are made nondeductible, the "tax price" for state expenditures will go up. Now, in effect, the federal government subsidizes the state. If a New York voter who pays 28 percent of his or her income to the federal government in income taxes, pays $1000 more in state taxes, and can deduct that expenditure from income, his or her federal tax liability goes down by $280, and the *net cost* to him or her is only $720. With a higher tax price, though, demand for state-provided goods and services would normally be expected to decrease. The median voter theory says "not to worry": tax deductibility has no effect on the level of state-provided goods and services because the median voter in the state does not itemize.

THE TWO-PARTY SYSTEM AND THE MEDIAN VOTER

We noted earlier that an elected representative bears a negligible fraction of the costs of, and receives a negligible fraction of the benefits from, an increase in government expenditure. What can economic theory say about how he or she should vote? A natural supposition is that the politician wishes to stay in office and that, accordingly, this individual wishes to maximize his or her votes, given the position taken by his or her rival. A vote-maximizing voting strategy can easily be defined as follows: Assume there are two parties, "R" and "D." Party R takes the position of Party D as a given. Focusing on a single issue, the level of expenditure, denote by G_R the position of Party R (that is, the level of public expenditure advocated by the party) and by G_D the position of Party D.

For each value of G_D there is an optimal (that is, vote-maximizing) position for G_R.

Under the hypothesis that each party seeks to maximize its vote given the position of its rival, what will each party do? Let G_m be the preferred level of expenditure of the median voter. Suppose Party D chooses $G_D > G_m$. Then, if Party R takes a position between G_m and G_D, it will get all the voters who prefer an expenditure level less than or equal to G_m, and some who prefer slightly more; thus Party R gets over 50 percent of the votes and wins. In response, Party D will choose a position, G_D', between G_m and G_R, which wins against G_R. Then, however, Party R chooses a position, G_R', between G_D' and G_m. The process continues until both parties stand for the same position: that of the median voter (G_m). (See Figure 9.6.)

This result is consistent with the widely observed allegation that with our two-party system, voters get no choice: both parties take a "middle-of-the-road" position. This is precisely what the theory predicts.[9]

Some important limitations of the theory, however, need to be borne in mind. First, we noted earlier that in general, there may not be a majority voting equilibrium. One does exist if individuals have single-peaked preferences. In the present context, this requires that we should be able to arrange issues along a single dimension—for example, conservative–liberal. If, however, there are a variety of dimensions—some individuals are liberal on some issues and conservative on others—then the median voter is not well defined, and there may be no equilibrium to the political process.

Second, we have ignored questions of participation in the political process. There are, for instance, costs associated with becoming informed

---

[9] This general theory is due to Harold Hotelling, a pioneer in mathematical economics who taught at Columbia and North Carolina State universities. See H. Hotelling, "Stability in Competition," *Economic Journal* 39 (March 1929): 41–57. For a comparative portrait of the median voter over time and across countries, see: J. E. De Neve, "The Median Voter Data Set: Voter Preferences across 50 Democracies," *Electoral Studies* 30 (December 2011): 865–887.

SOCIAL CHOICE THEORY

The hypothesis that government actions can be explained as the outcome of individuals' acting rationally in their own self-interest, in response to the political "rules of the game," is a central tenet of social choice theory. James Buchanan, who taught at George Mason University until his death in 2013, received a Nobel Prize in 1986 for his pioneering work in this area. In this view, designing the rules of the game—the constitution—is critical. An important part of the rules of the game is the imposition of constraints on government (such as limits on deficit financing). Without some form of constraints, the incentive of the majority to redistribute income in their favor, away from the minority, or of special interest groups to try to induce elected representatives to act in *their* interests, rather than in the general interest, is too great to be resisted.

Even acts of Congress that seem to go against the special interest groups appear in a different light when viewed from this perspective. Consider the 1986 tax reform, which attempted to strip out of the tax law many of the favorable provisions that special interest groups had successfully lobbied for inclusion. Professor Buchanan argued that this action should be viewed from the self-interest of Congress: the tax system had become so overladen with special provisions that the cost of granting further special benefits was rising sharply. The tax reform enabled Congress to start with a clean slate: now there was greater opportunity for introducing *new* special benefits. The greater opportunity for Congress to dispense special favors was of immense benefit to congressional representatives.*

One issue addressed by Nobel Laureate Gary Becker, of Chicago and Stanford universities, is how to explain the seemingly disproportionate influence of certain small interest groups. Although they constitute less than 2 percent of Americans, farmers have succeeded in getting huge subsidies from the federal government. The answer Becker suggests is that with small groups, the free rider problem that we encountered in Chapter 5 is smaller. "Bribing" representatives to support one's special interest is a public good: all wheat farmers benefit from a wheat subsidy; all steel or car producers benefit from trade barriers that keep out less expensive foreign steel or cars. The smaller the group, the easier it is to persuade all members to contribute to the cost of lobbying. Each of these programs has losers, and not only are the losers far more numerous, but their aggregate losses exceed the benefits of the special interest groups—although each of the losers loses a little. Opposing the special interest groups is also a public good, and each opponent has an incentive to be a free rider.

*James M. Buchanan, "Tax Reform as Political Choice," *Journal of Economic Perspectives* 1 (Summer 1987): 29–35.

and voting. These costs are sufficiently great relative to the perceived benefits that, for example, even slight changes in the weather, which makes it slightly less pleasant to go outside to vote, have significant effects on voter participation. In particular, voters whose preferences are near the median have little incentive to be politically active, particularly if they believe that the political process will reflect their preferences anyway. Thus, it may be in the interests of those who are more extreme to attempt to pull their party away from the center. This tendency for greater political activism at the

extremes may partially offset the median-directed tendencies noted earlier. There is a tension between the centripetal force of the median voter model and the centrifugal force of the base voter mobilization model.

ALTERNATIVES FOR DETERMINING PUBLIC GOODS EXPENDITURES

We have identified several major problems with majority voting, the most commonly employed way of making collective decisions: there may not be a determinate outcome, and even when there is, it may not be efficient; and voters may vote strategically, not revealing their true preferences. Even if there is no ideal system, are there perhaps systems that resolve one or the other of these problems? Economists have looked for such alternative systems.

LINDAHL EQUILIBRIUM

The most famous alternative system is called the *Lindahl solution*, after the great Swedish economist Erik Lindahl, who first proposed it in 1919.[10] He was looking for a system that would yield efficiency; he paid little attention to the other problems listed earlier. Lindahl's system attempts to mimic, as far as possible, the way that the market works in providing private goods. Remember, market equilibrium for private goods is described by the intersection of the demand and supply curves. All individuals face the same price. The sum of the quantities they demand is equal to the sum of the quantities supplied by firms.

One way we can characterize the efficient level of public goods is the intersection of the "collective" demand curve (formed by adding vertically each individual's demand curve) with the supply curve. The demand curves are generated by asking the individual how much of the public good he or she would demand if he or she were to pay so much for each unit produced; that is, in Figure 9.7A if the first individual faced a tax price of, say, p_1, he or she would demand G^*.

---

[10] E. Lindahl, "*Positive Lösung, Die Gerechtigkeit der Besteuerung*," translated as "Just Taxation—A Positive Solution" in *Classics in the Theory of Public Finance*, ed. R. A. Musgrave and A. T. Peacock (New York: St. Martin's Press, 1958). Application of the Lindahl equilibrium for determining public goods expenditures is still debated today, see for example, J. Dávila, J. Eeckhout, and C. Martinelli, "Bargaining over Public Goods," *Journal of Public Economic Theory* 11 (December 2009): 927–945; and A. van den Nouweland and M. Wooders, "Share Equilibrium in Local Public Good Economies," *Journal of Mathematical Economics* 47 (May 2011): 376–381.

A Individual 1

B Individual 2

C Collective demand curve—
the vertical sum of individual
demands
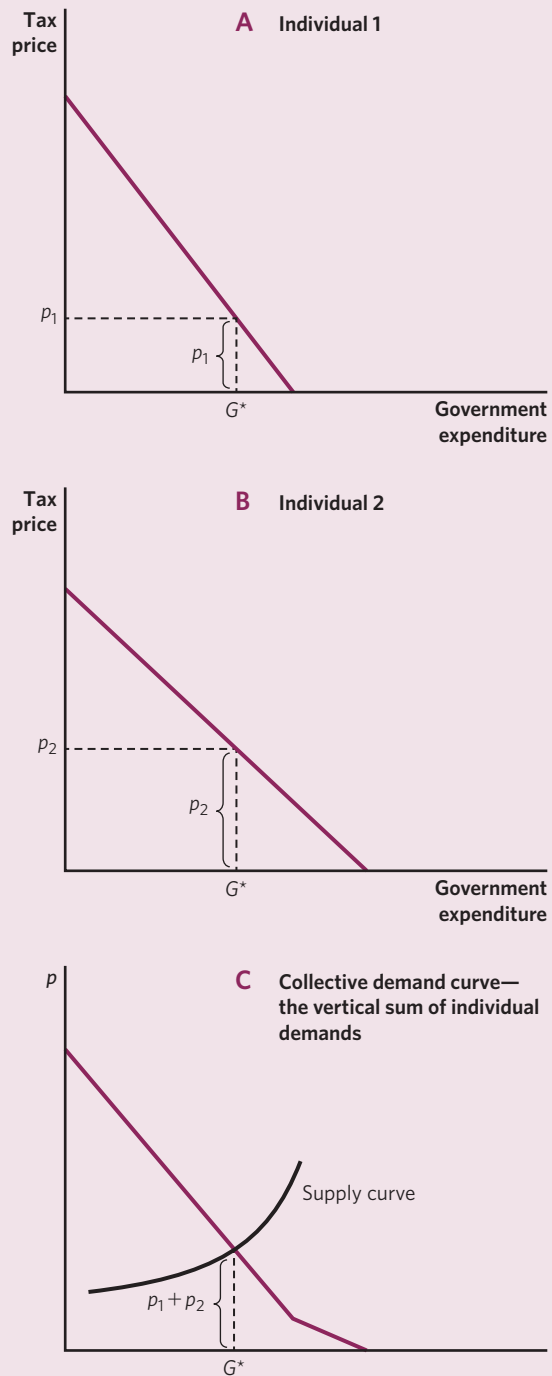
The **Lindahl equilibrium** is illustrated in Figure 9.7C. We add vertically the demand curves for Individuals 1 and 2 illustrated in Figure 9.7A and 9.7B, obtaining the collective demand. The Lindahl equilibrium occurs at the intersection of this curve with the supply curve. Price, along the supply curve, measures the marginal cost of production. p_1 measures the first individual's marginal benefit (marginal willingness to pay for an additional unit of government expenditure) at G^*, and what he or she has to pay, and p_2 measures the same thing for the second individual. At G^*, the sum of the marginal benefits (the total marginal willingness to pay) just equals the marginal cost. The Lindahl equilibrium is thus a set of tax prices (the amount each individual has to pay if one more unit of the public good is produced) adding up to the marginal cost of production, such that, given those tax prices, every individual prefers the same level of expenditures, G^*. Because at the Lindahl equilibrium the sum of the marginal benefits equals the marginal cost, the Lindahl equilibrium is Pareto efficient.

We noted earlier that, in fact, there were a whole range of Pareto efficient resource allocations, with one individual better off in some and another better off in others. Almost by definition, there cannot be unanimity about which of these points is preferred. The Lindahl equilibrium picks one of the Pareto efficient points, but individuals who are disadvantaged by this particular Pareto efficient point will not agree to the use of this mechanism for determining the allocation of public goods; indeed, they would prefer Pareto inefficient allocations as long as the level of utility they obtain is higher.

The most telling criticism of the Lindahl solution is that *individuals do not have an incentive to tell the truth because their tax price increases as their stated demand does*—that is, the higher their stated demand (given the demand statements of others), the higher the equilibrium expenditures on public goods will be. Higher expenditures on public goods, of course, necessitate higher equilibrium tax prices. The demand curves that are used in the Lindahl analysis were drawn under the hypothesis that individuals face a given tax price; they believe that nothing they say will alter what they have to pay per unit of public expenditure. If they understand the Lindahl mechanism, however, they will realize that what they say does alter what they have to pay per unit of public expenditure, and thus they will not truthfully reveal their demands.

COMPARISON OF PUBLIC CHOICE MECHANISMS

Majority voting
- Equilibrium may not exist.
- When equilibrium exists, in general, it is not Pareto efficient.

Lindahl equilibrium
- Equilibrium always exists.
- Equilibrium is always Pareto efficient.
- Individuals do not have incentive for truthful revelation of preferences.

Let us briefly review the two sets of processes by which collective decisions concerning public goods could be determined, majority voting and the Lindahl equilibrium. Voting, as we have seen, may not yield an equilibrium; when it does, it is, in general, not Pareto efficient. We saw earlier that no democratic mechanism fully resolved the problem of the nonexistence of a voting equilibrium. The Lindahl equilibrium will be efficient, but individuals do not have an incentive to be honest in providing the information required to implement it. The next natural question is: Are there more general plausible ways of organizing collective decision making that yield efficient outcomes, and that do not have the problems we have encountered in the case of majority voting and the Lindahl equilibrium? Although economists have devised complicated schemes that have *some* desirable properties, or that may work in *some* circumstances, it appears that there is no *perfect* solution to the problem of collective decision making in democratic societies. The appendix to this chapter describes one scheme that, although it induces individuals to reveal honestly their preferences, has other problems that perhaps explain why it is not commonly used.

POLITICS AND ECONOMICS

The preceding discussion of the political process differs markedly from the kind of analysis one might typically find in a political science course. There, one might discuss the roles of special interest groups and political institutions. A full discussion of the relationship between economic theories of the political process and other theories of the political process would take us beyond the scope of this book. In the following pages, we touch on some economic interpretations of certain political phenomena.

WHY DO INDIVIDUALS VOTE?

In many elections, voter participation rates are low and sensitive to such chance occurrences as changes in the weather. The reason for this, as we have noted, is that the benefits of voting for the individual are low—there is little chance of one person affecting the outcome. The alternatives may differ so little that the outcome is inconsequential, and, even though the costs of voting are relatively low, they are not low in relation to the benefits. Indeed, in a fully rational calculation, no one would vote: the probability that an individual's vote would make a difference to the outcome (because, in most cases, the individual cares only about whether his or

her candidate wins or loses, not the magnitude of the win or loss) is essentially zero. However, individuals *do* vote.[11]

This paradox is resolved, in a somewhat tautological manner, simply by assuming that individuals get utility out of voting, or, more generally, out of participation in the political process. More to the point, considerable time and energy are devoted to inculcating into our children notions of civic responsibility, and among these civic responsibilities is the responsibility to be an informed voter.

The same considerations imply that when the individual votes, he or she may not act in the narrowly self-interested manner that we have assumed in our discussion so far. Individuals may support representatives who vote for closing some loophole in the income tax system because it would result in a more equitable distribution of the tax burden, even though their personal tax liability might thereby be increased.

ELECTIONS AND SPECIAL INTEREST GROUPS

The political models we have discussed thus far have assumed that all individuals are well informed about the consequences of all alternatives under consideration, that all individuals vote, and that they cast their votes on the basis of the implications that each alternative has for their own (private) welfare.

Many believe that this does not provide an adequate description of the political process. Although constitutionally each person has one vote, some votes seem more effective than others. The outcome of the political process, in this view, reflects the political power of special interest groups.

Assessing the validity of these views is beyond the scope of this chapter. Instead, we focus on a more narrow set of questions: What can economic theory say about the kinds of interest groups that are likely to be effective? How can we reconcile the effectiveness of special interest groups with the fact that each individual in the United States has only one vote?

The answer to these questions is related to our discussion in Chapter 5 of the public interest as a public good. We saw there that the efficient management of the public sector was a public good. Similarly, when others choose elected officials who are competent and who reflect values similar

---

[11] In U.S. presidential elections since 1932, between 49 and 63 percent of the voting-age population has voted. In elections of U.S. representatives to Congress, participation has been somewhat lower, at 33 to 59 percent (U.S. Census Bureau, *The 2011 Statistical Abstract*, Table 418). In local school board elections, often less than 10 percent of the eligible voters vote.

to our own, we benefit as well. This may reduce our incentive to become informed voters, and encourage us instead to "free ride."

At the same time, we should note that the free rider problem may not be as serious in small groups as it is in large ones. It is easier, for instance, to form an interest group of a small number of steel producers to attempt to persuade Congress to restrict steel imports than it is to form an interest group of the large number of steel users who would be adversely affected by such restrictions. Each of the producers has more to gain than each of the consumers has to lose, even though the aggregate gains of producers may, in fact, be less than the aggregate losses to consumers.

Trade unions, recognizing the nature of the free rider problem, have long sought closed shops, in which all workers must support the activities the union believes to be in their collective interests. Once they have this power, they can attempt to use it not only at the bargaining table but also in the political arena, where they act as a special interest group.

THE POWER OF SPECIAL INTEREST GROUPS

How are interest groups able to exercise power? There appear to be at least three mechanisms. First, as we noted earlier, individuals have little incentive to vote or to become informed concerning the issues. Interest groups can attempt to lower the costs of becoming informed and voting, particularly for voters who are likely to support them. They do so by making information (supporting their own views) readily available and providing transportation, child care, and other assistance on polling day. Interest groups mobilize their constituencies through a combination of targeted information campaigns and aggressive "get out the vote" efforts.

Second, we noted the difficulty that politicians have in obtaining information about the preferences of their constituents. There is no simple demand-revealing mechanism for public goods as there is for private goods. Interest groups attempt to provide such information. Politicians may lack the technical information required to make informed political decisions—for example, they may not know the consequences of continued imports of cheap foreign cars or clothing. Interest groups are a primary source of information, and it is through providing information that they often exercise influence.

The third mechanism is through direct and indirect bribery of the politician. Direct bribery does not occur often, at least in most jurisdictions in the United States. (Presumably, this may not be due to the purity of our politicians so much as to the costs associated with being caught.)

Indirect bribery, however, is important: special interest groups provide financial and other forms of support for politicians who support their interests; this support is viewed to be essential in running a successful political campaign.

Since Congress enacted the first federal campaign finance legislation in 1867 prohibiting federal officials from requesting contributions from navy yard workers, there have been many attempts at campaign finance reform. Key laws over the past century include the Tillman Act of 1907, which prohibited corporations and national banks from contributing money to federal campaigns; the Federal Corrupt Practices Act of 1925, which strengthened disclosure requirements and increased expenditure limits; and the Federal Election Campaign Act of 1971 (together with the 1971 Revenue Act and many subsequent amendments) to provide stricter limits and greater transparency in campaign financing, an institutional framework for enforcing federal campaign finance laws, and federal financing of presidential elections.

Nonetheless, the impact of many of these statutes has been steadily eroded by legislative loopholes, poor enforcement, and court challenges, usually based on freedom of speech. For example, the Supreme Court ruling in *Citizens United v. Federal Election Commission*, 558 U.S. 310 (2010), determined that the First Amendment's right to free speech applies to corporations, permitting them to advocate expressly for or against candidates without any government-imposed limits on corporate campaign expenditures (see case study, "Campaign Finance Reform").[12]

OTHER ASPECTS OF THE POLITICAL PROCESS

Several other aspects of the political process affect outcomes. In the United States, each of the two parties holds primary elections to determine who represents them in the general election. Voter turnout in each reflects disproportionately the "extremes" in the party: more conservative Republicans or more liberal Democrats. Although each party has an interest in nominating a candidate who will win—and thus one who is a "centrist"—there is an offsetting centrifugal tendency in the primaries.

This has been exacerbated by the process of gerrymandering—designing congressional districts in which one or the other party is dominant. Winning the primary in such a district is almost tantamount to winning the election.

---

[12] Federal Election Commission, *The FEC and the Federal Campaign Finance Law*, February 2004 (updated February 2011), and *The Federal Election Campaign Laws: A Short History*, www.fec.gov; and National Public Radio, *A Century Of U.S. Campaign Finance Law*, www.npr.org.

CAMPAIGN FINANCE REFORM

Behind the power of many special interest groups is the power of money—especially money to support candidates they favor. Traditionally, the Democratic Party has had less access to money, and its members have been particularly concerned that the power of money has created an unlevel playing field, where the voices of rich individuals and corporations are heard more clearly than the voices of the poor and middle class. However, incumbents of any party, who typically have better access to financial support, are usually less enthusiastic about reform than those trying to get elected. This helps explain in part why Congress, at least in the past, has had only limited enthusiasm for strong reforms.

Some reforms have focused on limiting the amount of money that any organization can give to any candidate. Such reforms typically have been subverted: if a corporation cannot give money directly, it may still encourage its employees to contribute. If they cannot give money to a candidate, they give money to the candidate's party. Even if they were restricted in their contributions to the candidate's party, they could give money to some group identified with a position of the candidate. If the National Rifle Association were limited in the amount it gave to pro-gun representatives, for instance, it could still mount a campaign against their opponents who advocate gun control.

Some critics of campaign finance reform worry that it will create an unlevel playing field of a different sort. Groups that can mobilize thousands of volunteers and huge letter-writing campaigns, such as senior citizen advocates or labor unions, will still be able to exercise political influence.

There is also a worry that any restrictions on contributions to parties or causes would be, in effect, a restriction on First Amendment (free speech) rights—in a democracy, individuals should not be restricted in what they say, how they say it, or to whom they say it. Campaign finance mainly goes to support the dissemination of views; although, to be sure, questions may be raised at the extent to which a thirty-second spot ad on TV promotes intelligent decision making.

One set of reforms that has already been implemented entails public disclosure of campaign contributions. The hope is that candidates who receive large amounts of money, say, from the financial sector might feel embarrassed supporting legislation that is excessively favorable to that sector. Even this disclosure requirement has been undermined recently, though, with the creation of secretive "SuperPacs" that provide support for candidates and positions.

The second reform entails trying to create a more level playing field by partially matching with public funds money raised privately by the champion money raisers. Such actions do not circumscribe the right of any individual to say what he or she wants, but ensures that others with differing views can also get their perspectives across.

One consequence of gerrymandering is that a party can take control of, say, the House of Representatives even if it gets far less than a majority of votes. This happened in the 2012 election, where the Democrats garnered more than one million more votes, but the House was controlled by Republicans. Other countries, concerned that such outcomes are undemocratic, have introduced systems of proportional representation, in which

the number of parliamentarians is proportional to votes received, rather than a "winner-takes-all" system.

We have noticed the problems posed by low voter turnout. Some countries have tried to address this by requiring everyone to vote (or, more accurately, go to the voting booth). Even with small fines, much higher voter turnout is achieved, with the hope that the electoral outcomes are more reflective of the views of the citizenry.

We noted earlier that, in the American system, politicians can be partially "bought" by campaign contributions. There are other ways by which their behavior gets distorted to reflect the interests of special interests. Of particular concern is the "revolving door," in which politicians obtain jobs after their public service in the industries over which they passed legislation while in office. There is a clear conflict of interest, and although there are rules in places to limit the extent to which this occurs, they have proved imperfectly effective.

Finally, the media play an important role in forming opinions, both of political leaders and ordinary citizens. In some countries, the media are controlled by a small elite, representing particular views (say, of the very wealthy). In such circumstances, news and information can get distorted. Although in most countries politicians believe that they are unfairly treated, in many, there is a real problem of lack of media diversity. Some developing countries have responded by providing public support for community radio stations; many developed countries provide broad-based support for various forms of media. Information—especially accurate information—is a public good, and, as we have repeatedly seen, private markets undersupply public goods.

THE ALTRUISTIC POLITICIAN?

An alternative view holds that many politicians do not behave in as self-interested a way as we have assumed throughout this chapter. Just as individuals behave altruistically as private citizens and give to charity, so too do they behave as public citizens, in their capacity as elected officials. In our society, considerable status and respect are accorded to public officials and public service. Effective government depends on the quality of these public servants.

However, being disinterested does not suffice to ensure that politicians will make wise decisions. As has been noted, even a well-intentioned public official may have a difficult time ascertaining the preferences of his or her constituents. Further, even if the electorate would like to choose altruistic representatives who vote exclusively on principle, attempting

to ensure that there is an "efficient" supply of public goods, they face a difficult problem: as long as not all individuals running for office are disinterested, the voters must select between those who are and those who are not. If voters believe that it is better to be a "disinterested" public servant than a selfish politician, then self-interested politicians will all attempt to resemble disinterested public servants. How are voters to choose among them on the basis of the limited information they normally have available?

In numerous instances, politicians seem to act "on principle," voting in a way that is inconsistent not only with their own narrowly defined self-interest, but also with the wishes of their voters. They thus risk not getting elected (even though the electorate often respects such shows of "independence" and "principle").

Although altruistic behavior on the part of politicians is to be preferred to corruption, or even to lesser forms of self-aggrandizement, economists have long worried about the reliability of seemingly altruistic behavior in the pursuit of the common interest. Indeed, it was the seeming capriciousness of the actions taken by political leaders (often allegedly in the public interest) that led Adam Smith to suggest that there was a better way that the public interest might be served: by each individual's pursuing his or her private interest. Unfortunately, although Adam Smith's invisible hand may work well for most goods, it does not work well for public goods. Still, at least to a limited degree, self-interest on the part of politicians—their desire to get re-elected—serves an important function: it induces them to try to elicit accurately the preferences of their (voting) constituents and to vote for measures that reflect those preferences. It is this form of self-interest that lies behind the analysis of the two-party model presented earlier.

THE PERSISTENCE OF INEFFICIENT EQUILIBRIUM

In recent years, the prices of commodities—from cars to corn to aluminum to uranium to tomatoes—have been higher as a result of trade restrictions, or the threat of imposition of trade restrictions. When special interest groups manage to impose trade restrictions or to obtain subsidies for themselves, the resulting resource allocations not only violate generally accepted standards of equity and fairness but also are frequently inefficient. There are alternative allocations that could make everyone better off. Why, in the face of this, don't individuals get together and propose one of these Pareto dominating alternatives, to which, presumably, all

would agree? There is no universally agreed-upon answer to this puzzle. Several "partial" answers may be suggestive.

First, as we have already seen, the public interest is a public good. Because the efforts to maintain a good government must come from private individuals, there will be an undersupply of this (as any other) privately provided public good.

Second, many of the distributive implications of public programs undertaken at the behest of special interest groups are far from obvious—and this is deliberately so. It is unlikely, for instance, that the American voters would deliberately vote to transfer resources (give a public gift) to *rich* rice farmers. For these individuals to receive a transfer at the public expense, they must be included in a broader-scale program, of which they appear to be almost accidental beneficiaries. Thus, rich rice farmers become advocates of federal aid to rice farmers, singling out, in their public rhetoric, the benefits that would accrue to poor rice farmers from such a program. A Pareto improvement might, for instance, entail giving each rice farmer a fixed sum, which would leave the farmer free to move into some other occupation where productivity might be higher. Although such direct grants could be structured to make everyone better off, they would expose the true distributive implications of the program—that is, that most of the benefits accrue not to poor rice farmers, but to rich ones. Because the likely result would be weakened political support for subsidies to rice farmers, this Pareto improvement would not receive the backing of rich rice farmers.

REVIEW AND PRACTICE

SUMMARY

1. Collective decision making, such as determining the level of public goods, differs from standard decision making within a household in two important ways. First, there is a problem of eliciting preferences. If the amount that individuals have to pay depends on their statements, they may tend to understate their true preferences. If the amount that individuals have to pay does not depend at all on their statements, they may tend to overstate their true preferences. Second, there is a problem of resolving differences in preferences: How much should be provided if different individuals desire that the government should, for instance, spend different levels on providing public goods?

2. Majority voting is the simplest way by which such differences are resolved. Unfortunately, a majority voting equilibrium may not exist.

3. Arrow's impossibility theorem demonstrates the impossibility of finding an alternative, nondictatorial political mechanism resolves this problem of majority voting and satisfies certain other properties that one would desire of any political mechanism, such as independence of irrelevant alternatives.

4. The majority voting equilibrium exists if preferences are single-peaked.

5. Preferences for a single public good will usually be single-peaked. Preferences will not be single-peaked if:

  a. There is more than one public good, and the vote is taken over packages, rather than over a single good at a time.

  b. Voting is over a publicly provided private good, for which there exists a private alternative, such as education.

  c. Voting is over distributional questions, such as the structure of the income tax schedule.

6. The majority voting equilibrium, when it exists, reflects the preferences of the median voter.

7. In a two-party system, there will be a convergence of positions of the two parties toward that of the median voter.

8. The majority voting equilibrium does not, in general, result in an efficient supply of public goods; there may be either an undersupply or an oversupply.

9. The Lindahl equilibrium is the level of public goods provision in which the sum of the tax prices equals the marginal cost of production. Although the Lindahl equilibrium is Pareto efficient, there is no incentive for individuals to tell the truth concerning their preferences.

10. In many elections, voter participation is low. In fact, economists are puzzled why individuals bother to vote at all, as the expected private benefit—given the low probability that they affect the outcome—is typically less than the private cost.

11. Special interest groups often exercise strong influences over the outcome of political processes.

KEY CONCEPTS

Aggregating preferences
Arrow's impossibility theorem
Lindahl equilibrium
Median voter
Preference revelation
Proportional taxation
Single-peaked preferences
Tax price
Uniform taxation
Voting paradox

QUESTIONS AND PROBLEMS

1. Assume that some individual's marginal valuation of public goods increases. What does this do to the Pareto efficient level of public expenditures? If this individual is not the median individual, what will happen in a two-party system to the equilibrium level of expenditure on public goods? If the equilibrium was originally Pareto efficient, will it still be?

2. Assume that all individuals have identical preferences but some individuals are wealthier than others. Assume there is a single public good and a single private good.

  a. Show diagrammatically how you derive the demand curve for the public good, as a function of the tax price charged the individual.

  b. Assume that the demand function is of the form

$$G = \frac{kY}{p},$$

where k is a constant (less than 1), Y is income, and p is the tax price. This says that when income doubles the demand for public goods doubles, but when the tax price doubles the demand is cut in half. If the tax price is proportional to the individual's income (as with proportional taxation), how will demand for public goods differ among those with different incomes?

3. Assume instead there is uniform taxation, so all individuals face the same tax price. Recall that along each individual's demand curve, the price equals the marginal rate of substitution. Thus,

$$\text{MRS} = p = \frac{kY}{G},$$

the marginal rate of substitution is proportional to income. Assume that income is symmetrically distributed, so that mean income equals the median. Explain why the majority voting equilibrium will be Pareto efficient. Now, assume that income is not symmetrically distributed, but rather is skewed toward higher incomes, as in Figure 9.5B. Will the majority voting equilibrium still be efficient? Will there be an under- or over-supply of public goods?

4. Demand curves are said to be *income elastic* if the demand for the good increases more than proportionately with income. For instance, with the demand curve

$$G = \frac{kY^2}{p},$$

the demand for public goods increases with the *square* of income. Draw the marginal rate of substitution as a function of income for a fixed level of expenditure on public goods. Assume income is symmetrically distributed. What is the relationship between the average value of the marginal rate of substitution and the marginal rate of substitution of the median individual? What does this imply about the equilibrium supply of public goods under majority voting with uniform taxation?

5. In the text, we suggested that for well-off individuals with uniform taxation, preferences for education were not single-peaked. Why might preferences for local parks and urban public transportation systems (e.g., buses and subways) also not be single-peaked?

6. Is the median voter always the voter with the median income? Give examples.

7. How might the majority voting model be used to explain the growth of government expenditures?

   a. Should changes in median or average income better explain increases in the demand for government services?

   b. What should be the effect of an increase in the costs of producing public good caused by government inefficiency? Would it make a difference if the increase in cost is a result of government paying above-market wages (wages higher than those paid comparable workers in the private sector)? (Does your answer to the last question depend on whether the median voter is a government employee?)

   c. Why might you expect that if income per capita remains the same but the number of individuals in the economy increases, the demand for public goods would increase?

8. One popular voting scheme is rank-order voting, in which individuals assign a rank (1, 2, 3) to the possible alternatives; the assigned ranks are then added up, and the alternative with the lowest sum wins. Consider a choice among four alternative ways of spending public funds: a library, a ski slope, a swimming pool, and a garbage dump. Can you construct an example in which the outcome (the most preferred alternative) is, say, a library, if the vote is among the first three alternatives, whereas the outcome is a ski slope, if the vote is among all four alternatives? This voting scheme thus violates the principle that the chosen outcome should be independent of irrelevant outcomes (the garbage dump was not chosen in either situation).

9. Median voter theory says that to predict changes in collective decision making, one should focus on the median voter. Between 1973 and 1993, average incomes in the United States increased, whereas the income of the median family remained roughly stagnant. (Since then, the median income has increased slightly, but not enough to erase the increased gap of the previous two decades.) How might an economist focusing on median voter behavior and an economist focusing on average incomes differ in their predictions concerning changes in the level and composition of public expenditures?

APPENDIX: NEW PREFERENCE-REVELATION MECHANISMS

This appendix describes a simple procedure that induces individuals to reveal their demands truthfully, provided there is no collusion among individuals. Everyone is asked to give his or her demand curve for public goods, just as in the Lindahl equilibrium. As before, the equilibrium will be at the intersection of the collective demand (formed by adding vertically the demand curves of each individual) and the supply curve. For simplicity, we assume that the marginal cost of production of the public good is constant, so the supply curve is horizontal. Now, however, there is a different rule for determining the individual's tax liability.

We first add up the demand curves for *all other* individuals (vertically). The collective demand curve of all *others* intersects the supply curve at G_0 in Figure 9.8. This is what the level of public goods would be if an individual said that he or she got no value out of the public good. This individual is told that for each unit beyond G_0 that the government produces, he or she will have to pay the difference between the marginal costs of production and the collective valuation (demand) of all others. If the equilibrium entailed an output of $G_0 + 1$, the individual would have to pay AB, the distance between the marginal cost curve and the others' collective demand curve.

The individual is in a position to determine the level of public goods simply by announcing how much they are valued to him or her. Clearly, this individual will try to increase G to the point at which the marginal cost of increasing G is equal to his or her marginal benefit. This can be seen in two alternative ways. First, in Figure 9.8B, we have plotted the marginal cost to the individual from each additional unit of production beyond G_0, given others' demands. Because the individual's marginal cost is the difference between the cost of production and others' demand, the marginal cost of the $G + 1$st unit is equal to AB. In Figure 9.8B, we have also drawn the individual's demand curve; the individual will wish point G^* to be chosen, where his or her *demand curve* intersects his or her *marginal cost* curve.

We now show that each individual has an incentive to reveal honestly his or her demand for public goods, and that the equilibrium is Pareto efficient. To see this, we look at the individual's budget constraint. The individual faces a budget constraint as depicted in Figure 9.9. The extra amount that the individual has to give up for each extra unit of public goods beyond G_0 is the marginal cost minus the others' collective demand (marginal valuation). Thus, the individual sets his or her marginal rate of substitution equal to the marginal cost minus the others' collective

Price

A

Collective demand
of all others

Marginal cost of
the public good

A

B

G_0 $G_0 + 1$ **Quantity of
public goods**

Price

B

Marginal benefit
to individual
(his or her
demand curve)

Marginal cost
to individual
(given demand
curve of others)

A

B

G_0 $G_0 + 1$ G^* **Quantity of
public goods**
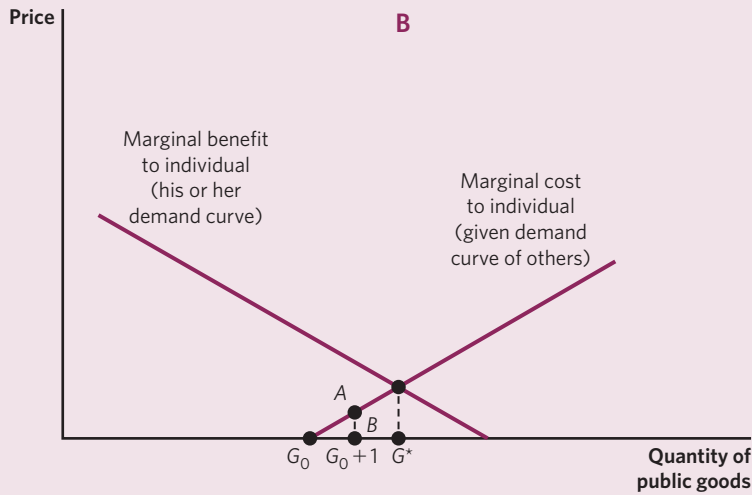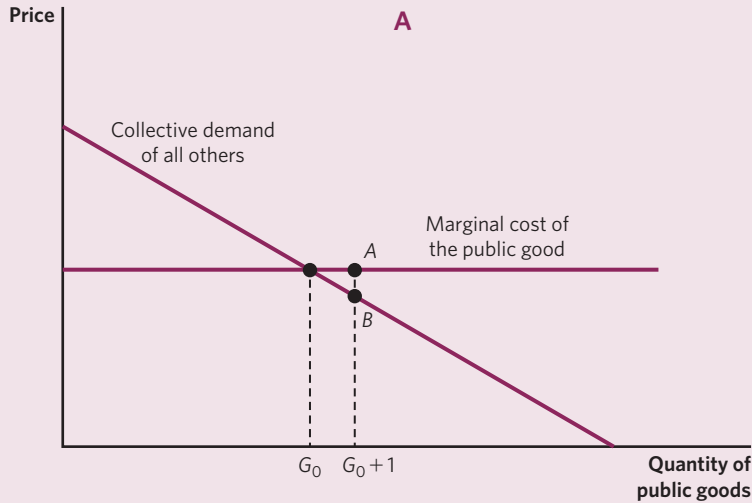
FIGURE 9.8

**NEW PREFERENCE-
REVELATION MECHANISM**
(A) Shows the collective demand
of all but one individual for the
public good (the sum of their
marginal valuations) and the
marginal cost of production.
If the last individual placed no
value on the public good, the
level of production of the public
good would be G_0, where the
sum of the marginal valuations
equals the marginal cost. As the
level of expenditure increases
beyond G_0, the last individual
is required to pay, for each
additional unit, the difference
between the collective (marginal)
valuation of all others and the
marginal cost. Thus, if $G_0 + 1$
is produced, this individual
must pay the amount denoted
by *AB*. (B) Shows, at each
level of output, the *marginal*
cost that the (last) individual
must pay for each extra unit
of output. Thus, to have the
economy go from producing G_0
to producing $G_0 + 1$ requires
that this individual pay *AB*. It
also shows the last individual's
marginal valuation of the public
good (his or her demand curve).
The individual's most preferred
level of expenditure is where
his or her marginal benefit from
increased expenditures (given
by the demand curve) exactly
equals his or her marginal cost—
that is, G^*. If the government
sets public expenditures at G^*,
the individual will be induced
to reveal his or her demand
truthfully.

demand—point *E* in Figure 9.9. It is clear that the individual has no incentive to misrepresent his or her preferences. If the individual asked for any level of public goods other than *G**, he or she would be worse off.

Assume now that each individual honestly announces his or her demand curve. Recall that in constructing the demand curve, the tax price for each individual (the slope of the budget constraint) was set equal to the individual's marginal rate of substitution. Hence, when the demand curves are added vertically, the sum of the marginal rates of substitution are just the sum of the tax prices, and at the Pareto efficient allocation that equals the marginal cost (the marginal rate of transformation):

$$MRS_1 + MRS_2 + \cdots = MC.$$

In other words, each individual's marginal rate of substitution is equal to the marginal cost of the public good minus the sum of the marginal rates of substitution of others (the sum of their tax prices). For instance, for the first individual,

$$MRS_1 = MC - (MRS_2 + MRS_3 + \cdots).$$

This is exactly the point we described earlier, at which the marginal cost to the individual of further increases in government expenditure (which equaled the marginal cost of production minus the sum of others' demand prices at the given quantity) equaled the marginal benefit to the individual (his or her marginal rate of substitution). We have just shown that by

honestly revealing his or her demand curve, the individual maximizes his or her own utility, and the allocation of resources to public goods will be Pareto efficient.

In spite of the attention that revelation mechanisms such as the one we have just described have received from economic theorists, there is considerable controversy about their relevance. There are several objections to them, besides their possibly high administrative costs and the fact that they are susceptible to collusion (two individuals could get together, agree to distort what each said, and each be better off).

Like the Lindahl equilibrium we described earlier, these mechanisms ensure that the condition for a Pareto efficient allocation—that the sum of marginal rates of substitution equals the marginal rate of transformation—is satisfied. However, some individuals might prefer another Pareto inefficient allocation that gives them a higher level of utility. Hence, it is not obvious that they would agree to a decision to adopt this mechanism, knowing that they would thereby be disadvantaged. Finally, the mechanisms do not, in general, guarantee a balanced budget. Although the sum of the marginal valuations (marginal rates of substitution) does equal the marginal cost, the total amount paid may well differ from the total costs of the public good.

PUBLIC EXPENDITURE IN PRACTICE

In this part, we show how the theoretical models we developed in previous chapters can be used to analyze public expenditure in practice, focusing on national defense, health care, education, welfare, and social insurance. These particular programs were chosen for two reasons. First, they are among the most important programs for many countries, including the United States. In terms of dollars spent, they account for approximately three-fourths of federal expenditures, as well as of total public expenditures in the United States. Second, the examination of these particular programs brings out most of the critical issues in expenditure analysis; other programs can be analyzed using the basic framework and tools of analysis that we develop here. The first two chapters are devoted to explaining our basic approach to the analysis of public expenditures: Chapter 10 develops a general framework, and Chapter 11 shows how the benefits and costs of different government programs can be quantified both before and after the money has been spent. Chapters 12 through 16 then apply these concepts to specific sectors.

10

FRAMEWORK FOR ANALYSIS OF EXPENDITURE POLICY

FOCUS QUESTIONS

1. What are the major steps in the analysis of a public expenditure program?

2. What are some of the reasons why the actual effects of a government program are different from those that are intended or apparent at first sight? What is meant by the *incidence* of a program?

3. Why are some programs said to be inefficient?

4. How in practice are the distributional impacts of a program assessed?

5. What is meant by the trade-off between equity and efficiency? Is there always such a trade-off?

6. Why might an understanding of the political process be relevant for an understanding of the design of government programs?

At least since the beginning of the 1990s, there have been calls for change in major segments of the American economy: its health care and education systems; major reforms in welfare and Social Security; cutbacks in some programs like defense; and expansions in others such as those aimed at developing new technologies. This chapter provides a framework for thinking systematically about such policies: questions that need to be asked, and methods that can be employed to help answer them.

Policy makers need such a framework to address increasingly complex issues. Indeed, the complexity of most government programs is so great that Congress delegates responsibility for working out most of the details (within guidelines Congress has set up) to the executive branch. In a process called *rule making*, agencies of the executive branch (such as the Environmental Protection Agency or the Department of Transportation) spell out these details, and the public is given time to comment. The Office of Management and Budget (OMB) within the Office of the President provides guidance to the agencies in how to go about this process. Currently, OMB guidance closely reflects the framework discussed in this chapter.

This framework for analyzing public expenditures provides guidelines for applying the efficiency and equity criteria presented in Chapters 3 and 7. It is not a simple formula that can be applied blindly to all problems, but rather a list of considerations that should be raised. Some may be more relevant to certain government programs than to others. The kinds of questions we are ultimately interested in addressing are:

- Why is there a government program in the first place?
- Why does the government program take on the particular form that it takes?
- How does the government program affect the private sector?
- Who gains and who loses as a result of the government program? Are the gains greater than the losses?
- Are there alternative programs that are (Pareto) superior to current government programs (that is, in which some individuals can be made better off without adversely affecting anyone else)? Are there alternative programs that have different distributional consequences but at the same time achieve the program's primary objectives? What are the impediments to the introduction of these alternative programs?

We begin by breaking down the analysis of public expenditures into ten steps: (1) the need for a program; (2) market failures addressed by the program; (3) alternatives to the program; (4) particular design features of the program; (5) private sector responses; (6) efficiency consequences; (7) distributional consequences; (8) equity–efficiency trade-offs; (9) public policy objectives; and (10) the political process.

NEED FOR A PROGRAM

It is often useful to begin the analysis of a public program by investigating the program's history and the circumstances under which it arose. Who were the individuals or groups who pressed for its passage, and what were the perceived needs that it supposedly addressed?

For instance, when the bill establishing the Social Security program was passed in 1935, the United States was in the midst of the Great Depression. Up to that time, few employers provided adequate pensions for their employees, and the private market for annuities (insurance policies that provide individuals with a given annual income from retirement until death, regardless of how long they live) was undeveloped; many individuals had failed to save adequately for their retirement, and many who had saved had found their savings wiped out by the stock market

crash in 1929. The failure to have adequate savings was not as irrational and improvident as it appears to us today; in those days, many individuals continued to work until they died. They needed life insurance to provide for their families after their demise, but did not need pensions for themselves. In the Great Depression, however, many of these individuals lost their jobs and had no unemployment insurance. It was widely felt that society had to make some provision for them and that it was preferable to do so on a systematic basis, rather than just to solve the immediate problems of the time.

MARKET FAILURES

The second step in the analysis of public programs is to attempt to relate the need, the source of demand, to one or more of the market failures discussed in Chapter 4: imperfect competition, public goods, externalities, incomplete markets, imperfect information, and macroeconomic disequilibrium. In addition, we saw in Chapter 4 that even if the economy is Pareto efficient, there are two further arguments for government intervention: first, that the distribution of income emerging from the market economy may not be socially equitable; and second, that an individual's own perceptions of his or her welfare may be an inappropriate or inadequate criterion for making welfare judgments. There are merit goods, which the government should encourage, and merit bads, which the government should discourage or prohibit.

In some cases, the nature of the market failure is obvious: national defense is a pure public good, and, as we argued earlier, in the absence of public provision, such goods will always be in undersupply. In other cases the answers are not so obvious, and economists may not agree about the nature of the market failure. Some economists believe that education is a public good, for example. However, most economists argue that it is essentially a private good (in the technical sense defined in Chapter 5) and that to find an explanation for its public provision one must look elsewhere; for instance, at capital market imperfections, at the distributive consequences of public provision, or at education as a merit good, essential for the functioning of a democratic society.

The fact that there is a demand for the public provision of some good or service does not in itself imply that there has been a market failure. Some demands for public provision arise from an inadequate understanding of the market and of the government's capabilities for making things better. Identifying whether there is or is not a market failure is an essential step in identifying the appropriate scope for government action.

HIGHER EDUCATION IN THE UNITED STATES

In the United States, higher education illustrates several of the alternative forms of government involvement. It is publicly produced: every state has its own system of universities, colleges, and junior colleges. Although direct aid to private universities is limited in the United States, in other countries like Canada, it is common, and is granted on the basis of the number of students enrolled. In the United States the federal government provides considerable aid to research universities through a variety of programs of support for basic and applied research.

Most federal support to higher education, however, takes the form of support to the consumers—the students. Even though there have been no general programs of support, there have been five major selective programs:

1. Since World War II, a large number of veterans have attended colleges and universities at government expense, most under the original GI Bill, and later under the Montgomery GI Bill and Post–9/11 GI Bill. The federal government also offers merit-based Army, Air Force, and Navy Reserve Officers' Training Corps (ROTC) scholarships, as well as Iraq and Afghanistan service grants for students whose parent or guardian died as a result of military service in either country.

2. Government provides targeted grant programs, such as Pell grants for low-income students, and Teacher Education Assistance for College and Higher Education (TEACH) grants for those studying to become elementary and secondary school teachers.

3. Government offers a variety of subsidized and unsubsidized student loans using either federal funds channeled through schools (Direct Loan Program) or funds from private lenders that are guaranteed by the federal government (Federal Family Education Loan Program).

4. Federal Work-Study Program provides part-time employment for enrolled students to help them pay for their educational expenses.

5. Federal government offers tax incentives for education-related expenses: the American opportunity tax credit for students pursuing an undergraduate degree or other recognized education credential, and the lifetime learning tax credit for those taking postsecondary education or vocational courses.

ALTERNATIVE FORMS OF GOVERNMENT INTERVENTION

Once a market failure has been identified, a variety of government actions might address the problem. The four major categories of government action are: (1) public production; (2) private production with taxes and subsidies aimed at encouraging or discouraging certain activities; (3) private production with government regulation aimed at ensuring

that firms act in the desired way; and (4) some form of public–private partnership (PPP).

If the government decides to bear responsibility for production, it must decide on how the output is to be allocated. It can charge for the good at market prices; it can charge for the good at something approximating the cost of production, as it typically does for electricity; it can charge for the good, but the charges can be much less than the cost of production, as it typically does for higher education; it can provide the good free of charge and uniformly, as it does for elementary school and secondary school education; or it can allocate the good or service in some way corresponding to a perceived need or benefit. In countries like Britain, where medicine is free, it is obviously not provided equally to all individuals. Needs differ, so the decision as to who gets how much of the available supply of medical services is left to doctors (operating within guidelines set up by the government, in consultation with them).

Similarly, if the good is to be produced privately, the government must decide whether to (a) contract directly for the commodity but retain responsibility for distributing it; (b) provide a subsidy to producers, with the hope that some of the benefits will be passed on to consumers through lower prices; or (c) provide a subsidy to consumers. If some form of subsidy is desired, government must decide whether it should be provided through the tax system or through a direct grant. If a subsidy is granted, the terms have to be decided upon, for example, how restrictive eligibility standards should be. Currently, governments engage in all these possible forms of action.

If the government opts for a hybrid PPP approach, then it must decide on the nature of its relationship with the private sector, particularly regarding mutual roles and responsibilities as well as allocation of risk. Arrangements can be as simple as service, lease, and management contracts with the private sector, whereby the government retains asset ownership but outsources daily operations. They also can be as complicated as build–operate–transfer (BOT) contracts, whereby the private sector builds, for example, a toll road, operates it for an extended period of time to recoup its initial investment plus a market return to the capital it expended—perhaps 20 to 30 years—and then transfers the asset to the government.

ALTERNATIVE FORMS OF GOVERNMENT INTERVENTION

1. Public Production
  - Free distribution
  - Distribution at below cost of production
  - Distribution at cost

2. Private Production
  - Government subsidies to (taxes on) producers
  - Government subsidies to (taxes on) consumers
  - Direct government distribution
  - Government regulation

3. Public–Private Partnership
  - Outsourcing procurement of goods and services
  - Service, lease, and management contracts for operations and maintenance
  - Concessions such as BOTs for capital investment
  - Joint ventures (financial and in-kind)

The importance of identifying alternative programs is increasingly recognized. Frequently, new programs can be devised that attain the objectives of older programs at less cost and more effectively. "Social innovation" is no less important than technological innovation. Today, there is increasing emphasis on the use of markets and market-like mechanisms.

THE IMPORTANCE OF PARTICULAR DESIGN FEATURES

The detailed provisions of a program—for instance, the precise statements concerning eligibility standards—are often crucial in determining the efficiency and equity consequences of the program. Fairness and efficiency require making a number of distinctions that, though clear in principle, are difficult to apply in practice. The distinction between those who are hungry and those who are not is an important one, but devising a program to provide food for the hungry requires some easy way of identifying who the hungry are. Too narrow a definition will result in many of those who are needy not receiving aid. Too broad a set of eligibility standards will result in many individuals who are not needy receiving aid, much to the objections of other taxpayers who are having to contribute to these individuals' support. Thus, because of the impossibility of identifying perfectly those who are truly deserving of aid, there is a trade-off, when designing regulations, between two types of errors: denying aid to those who are deserving and granting aid to those who are not deserving (see Figure 10.1). Different individuals may judge the importance of these two kinds of errors differently.

The design of eligibility standards has further effects, as individuals may alter their behavior to gain eligibility or to receive larger benefits. There has been concern, for instance, that welfare programs that provide funds only to single mothers discourage marriage. The Supplemental Nutrition Assistance Program (SNAP; formerly the Food Stamp Program),[1] which provides assistance to people with low available income, offers another example of altered incentives. To calculate the amount of income available for spending on food, expenditures on housing are subtracted from the individual's take-home pay, so the individual who spends more on housing receives more in food assistance. This may alter

---

[1] Not only has the name of the program changed, but paper coupons ("food stamps") have been replaced by the electronic benefits transfer (EBT) issuance system, similar to a debit card system.

behavior: because of the particular way it has been designed, SNAP— a program intended to encourage better nutrition among the poor—may encourage more expenditure on housing. There is little hard evidence to suggest that either of these concerns is significant.

PRIVATE SECTOR RESPONSES TO GOVERNMENT PROGRAMS

One of the central features of a mixed market economy like that of the United States is that the government has only a limited degree of control over it. The private sector, for instance, may react to any government program in such a way as to undo many of its alleged benefits. For example, when the government increases Social Security benefits, the welfare of the aged may not increase in the long run by the full corresponding amount; individuals may be induced to reduce their own savings for retirement, and children may be induced to provide less support for their aging parents. Public support may thus **crowd out** private support, eroding the impact of the program. Again, emperical observation to date provides little evidence that this has occurred in a significant way. In fact, there might actually be **crowding in**, for example, when government basic research encourages private sector applied research.

In considering the consequences of a government program, one needs to look at the long-run consequences, after all producers and consumers have adjusted their behavior, as well as the immediate impact. One of the major impacts of rent control, for instance, is that the supply of new housing dries up, the effects of which are felt only gradually.

Calculating the full private sector responses is often one of the most difficult and contentious aspects of analyzing a government program. To what extent, for instance, will a government subsidy to builders of lower-income housing result in higher profits, thus benefiting the building industry? To what extent will competition in the industry bid these profits away, lowering the price and increasing the supply, thus benefiting the intended beneficiaries? The answers depend on views concerning the housing and construction markets. How competitive is the building industry? If it is competitive, what is the elasticity of supply? What is the elasticity of demand?[2]

As we have already noted, the effects of a government program may hinge critically on seemingly innocuous design features. Economists look for *marginal incentive* effects. Two programs giving the same average subsidy can have quite different marginal incentive effects. For example, SNAP, with a $200 cap on benefits per month, may have no *marginal* effect for those individuals who spend more than $200 on food. An analysis of the magnitude of the demand and supply responses from the private sector—and thus of the effects on price and quantities—must pay careful attention to these marginal incentives.

Recent research in economics has suggested that the economist's standard model may not always work as well as traditionally thought. Individuals may pay less attention to some forms of incentives than the standard theory would have predicted, or react in an unexpected way. Many economists have thus tried to understand unanticipated responses (or no response at all) to incentives through the lens of behavioral economics, a relatively new discipline that explores the psychology underpinning household and firm responses to government policies and programs. The hope is that by exploring public finance through the lens of behavioral economics, we can more accurately predict private sector reactions to government initiatives spanning the domain of public sector economics, including public expenditure, taxation, social insurance, retirement savings accruals, social safety nets, and income redistribution.[3]

---

[2] Elasticity of supply is defined as the percentage change in quantity supplied as a result of a 1 percent change in price; elasticity of demand is defined as the percentage change in quantity demanded as a result of a 1 percent change in price.

[3] See, for example, J. King, S. Mullainathan, and W. Congdon, *Policy and Choice: Public Finance through the Lens of Behavioral Economics* (Washington, DC: Brookings Institution Press, 2011).

EFFICIENCY CONSEQUENCES

The next steps in expenditure policy analysis entail identifying the efficiency and distributional consequences of each alternative program and assessing the extent to which alternative programs can meet the objectives of public policy.

Government programs may result in inefficiencies both in the production of a good or service and in levels of consumption. In Chapter 8 we suggested that the government's decision to produce a good or service itself, to purchase the good or service from private firms but distribute it itself, or to have private firms produce it and market it subject to government regulation may significantly affect the costs associated with producing and delivering the given good or service.

We also suggested that when consumers had an element of choice, the competition among providers would likely increase the efficiency with which the goods or services were provided, as well as make what was produced more responsive to the needs and desires of consumers. These arguments are less persuasive if consumers have limited information concerning the product they are purchasing like medical care, or if consumer concern about costs is reduced because the government pays all (or a substantial part of) the costs, again, as in the case of medical care.

INCOME AND SUBSTITUTION EFFECTS AND INDUCED INEFFICIENCY

For many programs, it is useful to distinguish between **substitution effects** and **income effects**. Whenever a government program lowers the price of some commodity, there is a substitution effect: the individual substitutes the cheaper good for other goods. For example, with tuition subsidies for higher education, individuals substitute education for other goods on which they might have spent their money. On the other hand, grants to individuals that make them better off but do not alter the relative prices of different commodities result in an income effect: an individual changes his or her expenditure pattern after becoming better off. In many cases, there are both an income effect and a substitution effect, and both alter the individual's behavior. Normally, however, it is only the substitution effect that we associate with *inefficiency*.

To illustrate, assume that the government gives an individual SNAP assistance to buy $10 worth of groceries every week. Prior to this, the

INCOME EFFECT

Giving free food has an income
effect but no substitution effect:
its effects are identical to giving
an individual extra income.
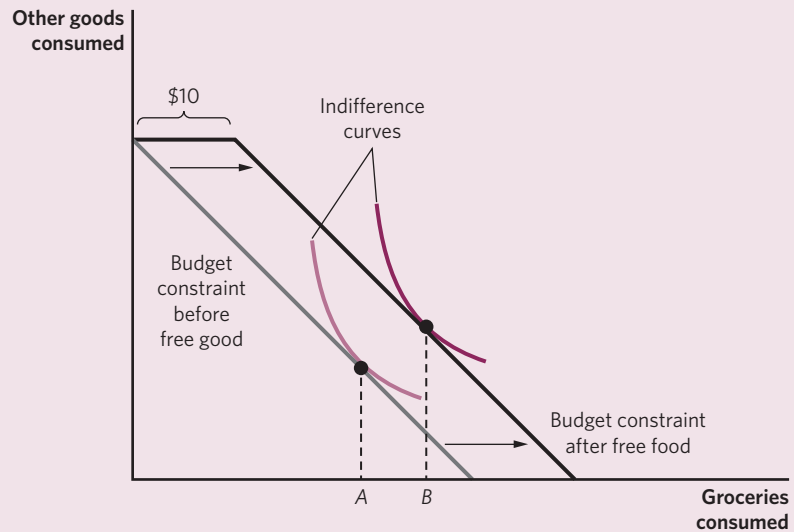
individual's budget constraint was the lower one in Figure 10.2, which
shows a substitution effect: by giving up $1 of groceries, the individual
could acquire $1 more of other goods. SNAP shifts budget constraint to
the right, so if the individual now wants to consume more than $10 worth
of groceries, he or she still must give up $1 of other goods for each extra
dollar of groceries consumed. There is no substitution effect, but there is
an income effect: the individual now has an extra $10 to spend. The effect
on food consumption is the same as giving the individual an equivalent
amount of income (except when the individual would prefer to consume
less than $10 worth of food each week). SNAP has altered the individual's
behavior to consume more food (*B*) than previously (*A*). Notice, however,
that the individual does not increase food consumption by the full $10; the
individual spreads this extra income between food and all other goods,
just as he or she would have with $10 more of income. Because there is no
substitution effect, there is no inefficiency associated with SNAP.

To see the substitution effect, assume, in contrast, that the govern-
ment has agreed to pay for 30 percent of food purchases. This lowers the
cost of food. The new budget constraint is shown in Figure 10.3. Now
there is a substitution effect. Food is cheaper relative to all other goods, so
the budget constraint rotates as shown. Note that in Figure 10.2, by con-
trast, the new budget constraint associated with SNAP is parallel with the
original budget constraint. (There is also an income effect in Figure 10.3:
because with cheaper food, the individual not only consumes more food,
but can also consume more of all other goods.)
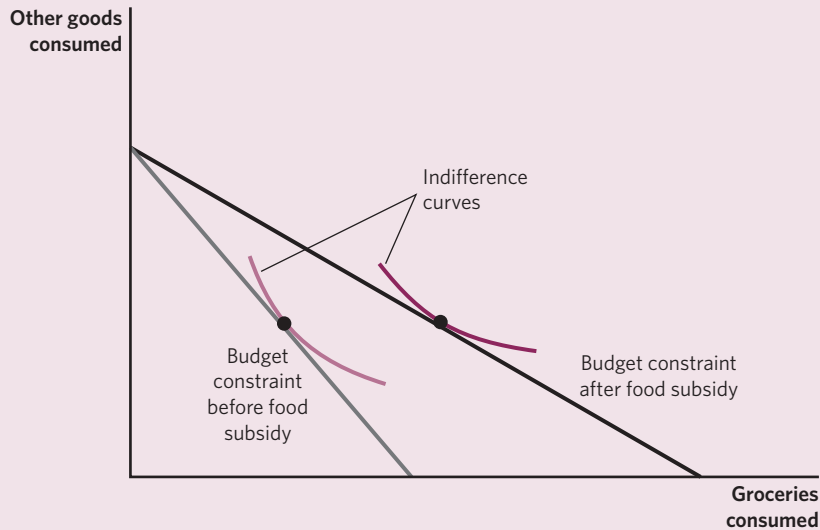
It is important to distinguish between income and substitution effects.
In some cases, the government may want to mitigate a market distortion,
such as monopoly power; or it may wish to encourage or discourage a par-
ticular economic activity, in which case, it may want a large substitution
effect. For example, if there is a belief that poor individuals do not attach
sufficient importance to housing, and the government wishes to improve
the quality of housing they purchase, then a program in which the gov-
ernment pays a fraction of housing expenditures (which has, as a result, a
substitution effect) will be more effective than a flat housing grant, which
(unless it is very large) has only an income effect. The same rationale
would apply to address the government's concern that parents might not
put a strong enough emphasis on the nutrition of their children.

On the other hand, if the government is primarily concerned with how
well off different individuals are if there are no market failures, then pro-
grams that do not alter marginal incentives are preferable; such programs
do not cause the inefficiencies associated with the substitution effect.

Returning to the SNAP case, we can see how a change in the design of
the program avoids the inefficiency generated by the substitution effect.
When the program was established in 1964, participants purchased food
stamps at a discount from their face value, so the government paid a frac-
tion of the costs of stamps: food stamps worth $100 might cost a poor person
$70, and he or she may be allowed to buy, say, up to $2000 of food stamps.
If he or she purchases $1000 worth of food stamps, the total subsidy would
be $300. Today, the government simply gives a low-income individual a

fixed amount of assistance (in the form of a credit on the recipient's elec-
tronic benefits transfer [EBT] card rather than in paper coupons); as long
as the amount given is equal to or less than the amount the individual
would spend on food anyway, this is equivalent to an income grant.

As we have seen, this version of food subsidies has only an income
effect, whereas the earlier version has a substitution effect as well. The
substitution effect introduces an inefficiency: the true cost of groceries—
the amount of other goods that society must give up to obtain an extra
unit of food—remains unchanged. For each additional dollar of food con-
sumed, society must give up $1 worth of other goods. Under the original
version of the food stamp program, however, individuals had to pay only
30 cents for a dollar's worth of groceries.

Figure 10.4 shows how the new form of the program can cost the gov-
ernment less, and leave poor individuals receiving the subsidy just as well
off as before. *BB* represents the budget constraint before any food subsi-
dies. The line *BKB'* represents the budget constraint under the original
form of the food stamp program, under which the government pays for a
fixed fraction of the costs of groceries, up to some limit. After that limit is
reached (represented by the point *K*) individuals must pay the full price
of groceries. The individual chooses the point *E*, at which his or her indif-
ference curve is tangent to the budget constraint. The magnitude of the
subsidy is the difference between what the individual has to pay and what
society has to forgo; it corresponds to the vertical distance between the
before-subsidy budget constraint and the after-subsidy budget constraint
at the equilibrium level of consumption of groceries, the distance *AE*.

Figure 10.4 also includes a budget constraint for assistance given in the form of a fixed amount to be spent on food. How do we know that the individual is just as well off? Because the income grant was set so that the budget constraint it produced would be tangent to the indifference curve through *E*. By definition of indifference curves, then, the individual is just as satisfied at *H* as at *E*. But the income grant costs the government less. Again, the cost of the program is represented by the vertical distance between the before-subsidy and after-subsidy budget constraints. The size of the grant, *HF*, is smaller than the size of the 30 percent subsidy, *AE*.[4] The reason for this is simple enough: when individuals must pay the full price of food at the margin (that is, when they have to pay $1 for $1 more worth of groceries), they value the increased consumption of groceries by precisely what they have to forgo in other consumption goods. But when individuals are given a 30 percent subsidy, they then purchase groceries up to the point at which they value $1 worth of groceries at 70 cents, which is the cost to them of the $1 worth of groceries.

Note, however, that under the new form of assistance, individuals consume less food than under the old form. If the purpose of the program is to encourage food consumption—because, for instance, government believes that individuals, in maximizing their own utility, will not consume enough food—then the old form, under which the government in effect lowers the price of food to the poor, is more effective.

DISTRIBUTIONAL CONSEQUENCES

It is not always easy to ascertain who really benefits from a given government program. Consider, for instance, the Medicare program, under which government finances most medical care for the aged. The aged clearly benefit greatly from the program, but to some extent, the federal aid substitutes for money that families of the elderly would have contributed (public expenditures thus crowd out private expenditures), and to that extent, the true beneficiaries of the program are not the elderly but their children. With this sort of analysis, economists seek to identify a property they call the **incidence** of a government expenditure program or tax; that is, they seek to answer the question of who ultimately benefits from, is hurt by, or bears the burden of the program or tax.

---

[4] To see this, note that the budget constraint segment *LB″* is parallel to the budget constraint *BB*—the individual has to pay the full marginal cost of food, so the trade-off between food and other goods is unchanged. Thus, *AG* = *FH* (the vertical distance between two parallel lines is the same everywhere). The inefficiency associated with the 30 percent subsidy is thus measured by *EG*—the government must spend extra that amount to leave the individual just as well off as he or she would have been with an income grant. *EG* is the deadweight loss associated with the inefficient subsidy.

Government programs often induce a variety of responses from the private sector that result in changes in prices. Thus, a program's effects can extend well beyond the people affected directly, and often the beneficiaries are different from those intended. There has been considerable concern that, at least in the short run, federal subsidies for private housing for the poor increase the price of housing, making the true beneficiaries the slum landlords, not the poor.

The effect of a government subsidy is illustrated in Figure 10.5, which shows the demand and supply curves for housing. In the short

FIGURE 10.5

SHORT-RUN AND LONG-RUN INCIDENCE OF EXPENDITURE PROGRAM

(A) In the short run, a subsidy may increase price more than quantity. Thus, landlords may benefit from a housing subsidy given to help the poor acquire better housing. (B) In the long run, the output response will be larger and the price response smaller.
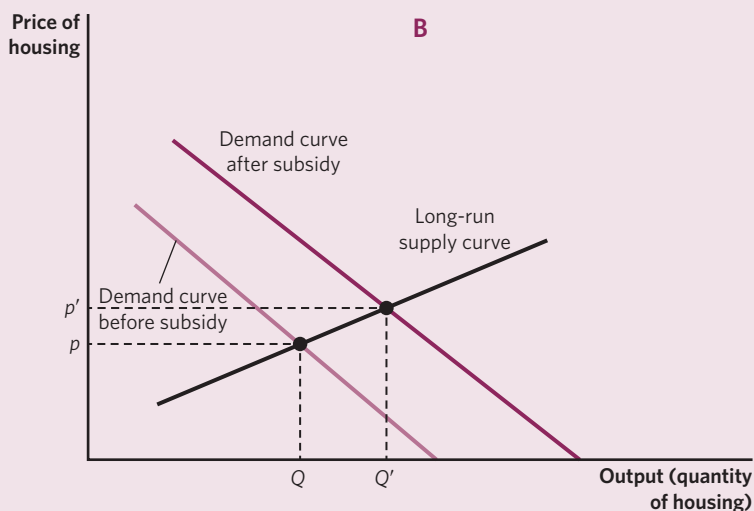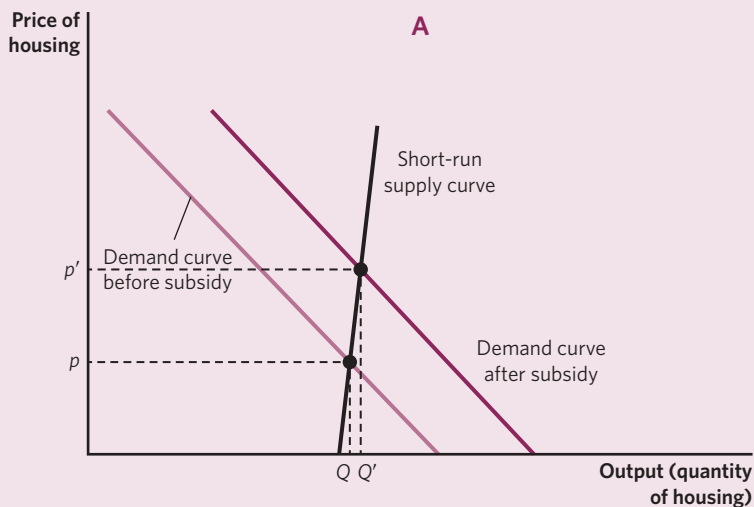
run (Figure 10.5A), the supply of housing is assumed to be very inelastic because it takes some time for new housing to be constructed. Assume the government has passed a general subsidy for housing, the effect of which is to increase the demand for housing (the demand curve shifts up). In the figure, almost the entire subsidy is reflected in the increased price of housing; the actual level of housing services provided increases very little. In the long run, of course, the supply response is likely to be larger; hence, in Figure 10.5B, the long-run supply curve is fairly flat, showing that a small percentage increase in the price, given enough time, elicits a fairly large increase in the supply of housing. In the short run, the beneficiaries of housing subsidies are the current owners of houses; renters find that virtually their entire subsidy is reflected in higher rents (the shift from p to p'). In the long run, however, renters are better off, as the increase in the quantity of housing supplied (from Q to Q') serves to limit the price increase.

Mass transit subsidies provide another example: Who benefits from a new subway system? At first glance, the answer seems obvious: subway riders. However, this may be incorrect. Those who own houses or apartments near the subway will find that their houses and apartments are more sought after; the increased demand for these residences will be reflected in the rents that the owners can charge (and in the market value of the houses and apartments). The commuter who owns no real estate finds that he or she is better off because of the better subway service, but worse off because of the higher rents, and the two effects are likely to cancel out. The true beneficiaries are the property owners near the subway lines.

The subway example illustrates a general principle: the benefits of government programs are often **capitalized** in the value of scarce assets associated with obtaining those benefits (land near the subway stops). In that case, the true beneficiaries are those who owned the asset at the time the program was announced (or passed, or when it came to be believed that the program would pass). By the same token, the costs of a program are often capitalized, so a tax on land is reflected in the value of the land; the true costs are borne by those who owned the asset at the time the tax was announced (or passed, or when it came to be believed that the tax would be passed). When those who benefit from a government program are different from those whom the program was intended to help, we say that the benefits have been **shifted**, or that the *actual incidence* (those on whom the benefits actually fall) is different from the intended one. Considerable research in recent years has been devoted to determining the actual incidence of government programs.

EVALUATING THE DISTRIBUTIONAL CONSEQUENCES

As we have noted, different individuals receive different benefits from a given government program. Although it is obviously not possible to identify how much *each* individual benefits, it may be important to know how different groups in society are affected differentially. Which groups we focus on may vary from program to program, and benefits may vary within a particular income group. Thus, a program of rebates for heating-oil expenditures for people whose income falls below a particular level obviously benefits the poor more than the rich, but it benefits some poor (those who consume a lot of heating oil, those who live in the Northeast) more than others (those who live in the Sun Belt). If the variability of consumption of heating oil among the poor is very large, this rebate program may be viewed as an unfair way of helping the poor, unless those who consume a lot of heating oil are viewed as particularly deserving of assistance.

INCIDENCE OF EDUCATION TAX CREDITS

In the 1996 election campaign, President Clinton proposed a $1500 tuition tax credit for the first two years of college for students who obtained a B average. The B average requirement was intended to encourage students to work hard. With the tuition tax credit, a middle-income family that spent $2000 in college tuition on their freshman daughter would be able to reduce their tax bill by $1500. It was as if they paid their full taxes, and then the federal government sent them a check for $1500. By subtracting the $1500 directly from the tax payment, this "round trip"—money going to the government and then coming back—is avoided.

Although the intent of the tuition tax credit was to help middle-class families with children in college, to increase enrollment, and to encourage better school performance, there was considerable controversy about the true incidence.

- Many states charged less than $1500 for community colleges. There was a concern that they

would (perhaps gradually) raise their tuition—after all, with $1500 coming from the federal government in tax credits, individuals could afford to pay more for tuition.

- Colleges could offset the increased tuition costs for those whose parents did not receive the tuition tax credit by giving them scholarships. However, in cases they do not, students would find college less, not more, affordable. Because most A and B students already go to college, the program might not increase enrollment much among them, but it might reduce enrollment among C students. Alternatively, teachers might worry that by giving a low grade—even a C—they might shut off a student's chance of staying in college by cutting off the tuition tax credit, so there may be grade inflation. With grade inflation, though, better students may have less incentive to work hard.

- If states did raise their tuition, then at least some of the benefits would accrue to state governments,

rather than to the taxpayers. This is the "true" incidence. Education still might be helped if the state spent most of the increased tuition revenue on education; on the other hand, the state might reduce its educational expenditures, and give a tax cut to its taxpayers. Then the true incidence would be on the states' taxpayers in general, not just on those who had children in college. If the state reduced taxes at the top, then what had appeared to be a middle-class tax break would become a tax break for upper-income individuals as a result of shifting.

• Similarly, private schools might be induced to raise their tuitions, enabling them to pay higher faculty salaries or support more research. Again, the incidence is markedly different from that intended.

As a result of concern about some of these perverse incentive effects—as well as complications in implementation—the B requirement was dropped when the Hope and Lifetime Learning tax credits were enacted as part of the 1997 Taxpayer Relief Act.* Since 1998, use rates of these two tax credits have increased steadily, with the following results: while both enrollments and tuition fees have gone up, there is no evidence such increases have been due to the tax credits; and in 2008, although 10 percent of tax credit claimants had an adjusted gross income (AGI) of less than $25,000 or more than $100,000, the remaining 90 percent of tax benefits went to the intended target population of middle-income taxpayers with an AGI between $25,000 and $100,000.

---

*The Hope tax credit was expanded and renamed the American Opportunity Tax Credit in 2009.

SOURCES: Department of Education, *Funding Education Beyond High School: The Guide to Federal Student Aid 2009–10*, www.FederalStudentAid .ed.gov; Internal Revenue Service, *IRS Publication 970—Tax Benefits for Higher Education*, www.irs.gov; and Benjamin Rue Silliman, "College Tuition Tax Credits: An Examination of the Impact of the American Opportunity, HOPE, and Lifetime Learning Tax Credits Since 1998", *Proceedings of ASBBS* 18, no. 1 (February 2011): 279–288.

In other cases, we may attempt to identify how producers are affected differentially. This typically is the focus of analysis in the evaluation of programs, such as agricultural price supports, aimed at aiding particular industries. In still other cases, such as the Social Security program, we may be concerned with the differential impact on the present elderly versus the impact on the young—the elderly of the future. We refer to these impacts as the program's **intertemporal distribution effects**—distribution effects over time. In still other cases, we may wish to identify the regional impact or the impact on cities versus suburbs, or urban versus rural areas.

When a program's benefits accrue disproportionately to the poor (they receive more than their contribution to the costs of the program through the tax system), we say that its distribution effect is **progressive**. If the benefits accrue disproportionately to the rich, we say that the program's distribution effect is **regressive**.

There are often controversies about who are the real beneficiaries of a program, and one's perspective on its distributive impact is determined in large part by the group on which one is focusing. For instance, government support for higher education is often viewed as enabling the children of the poor to go to college, and thus is viewed to have a positive

redistributive impact. However, children of the middle and upper-middle classes are more likely to avail themselves of a higher education. Thus, *general* subsidies—such as reduced tuition for all students—disproportionately benefit children of middle- and upper-income families. Indeed, by some calculations, they benefit more than their share of taxes—educational subsidies to higher education are thus regressive. This is in contrast to *targeted* subsidies—such as scholarships for children from low-income families. Even then, it is not clear that parents' income provides the appropriate focus of attention: the beneficiaries of education are not the parents but the children, since it is they who will receive higher wages as a result of their increased level of education.[5] Those who hold to this view often favor student loan programs.

Let us contrast the distributional consequences of direct state support for universities (allowing them to charge a low tuition) with the distributional consequences of a student loan program. Those who avail themselves of higher education will, on average, have a much higher income than those who do not. A loan program may thus be more progressive than the current system, in which even low-wage high school dropouts are called on to provide some support for higher education. Loan programs introduced in 1993, which allowed repayments to be related to students' incomes, could have increased progressivity still further, as students who wind up making higher incomes in effect pay more than those who receive low incomes, but in practice, relatively few student loans are of this type. As this example makes clear, one's view of the distributional impact of a government program depends not only on what groups one focuses on but also on the available alternatives to a given program. The relevant choice is seldom one program versus no program, but one *type* of program versus another. Thus, the present state system of aid to higher education *may* be more progressive than a totally private education system, but its distributional impact may look less favorable when contrasted with a system of income-contingent loans for higher education.

FAIRNESS AND DISTRIBUTION

Political discussions commonly focus on the equity of various proposals, with each side claiming that its proposals are more fair. Notions of fairness, unfortunately, are not always well defined, and different

---

[5] With middle- and upper-income parents who would have sent their children to college anyway, the true beneficiaries may be the parents who save the money they otherwise would have spent, but to the extent that parents use this money to increase the bequest they leave to their children, it is the children who really benefit.

individuals may have conflicting views of what is fair. A middle-class couple who love children but have decided, for financial reasons, to limit the number of children they have to two may feel that it is unfair for them to have to support a child from a family of ten children whose parents do not want to use modern birth control methods and cannot afford to send their children to college without government assistance. A couple who have saved $40,000 to put a child through college may feel that it is unfair that they are not entitled to receive a government grant or loan when their next-door neighbors, with the same income but who have put nothing aside for their children's education, enjoy expensive vacations every winter and are entitled to a government grant, which depends on assets as well as family income.

An unmarried person and a family with both spouses working may both think it unfair that their expected returns from Social Security are so much lower than those of an individual whose spouse does not have a job outside the home. However, an individual whose spouse does not work outside the home may feel that it is fair that he or she receive more, because the family has not had the benefit of a second income.

> ## CONSEQUENCES OF PUBLIC PROGRAMS
>
> 1. Government programs may crowd out private actions; equivalently, private actions may largely offset public actions, resulting in a small net effect.
> 2. Government programs give rise to income and substitution effects. The substitution effect is related to the magnitude of the marginal incentives.
> 3. Inefficiencies in public programs are related to the magnitude of the substitution effect.
> 4. The incidence of a program describes who ultimately benefits from, or is hurt by, the program. The actual incidence is often markedly different from the intended or apparent incidence.
> 5. The benefits of a program may be capitalized, in which case the true beneficiaries are those who own the asset in which they are capitalized at the time the program was started (or announced).

EQUITY–EFFICIENCY TRADE-OFFS

Because of the ambiguities associated with using the term "fair," economists try to avoid it in their analysis; rather, they focus on identifying the impact of programs. Economists begin their analysis of any program by looking for Pareto or near-Pareto improvements: changes in the program that make someone (or some groups) better off without making anyone (or almost anyone) worse off. Rent control, it is argued, fails to benefit renters in the long run, as the supply of housing dries up. There are better ways of helping low-income individuals obtain housing. Welfare programs would serve the beneficiaries better if taxpayers invested a little more money in training and education; in the long run, and the tax burden resulting from supporting the welfare population might actually be reduced. There may be alternative market-based ways of dealing with pollution, such as fines and tradable permits (see Chapter 6),

which can achieve higher levels of pollution reduction than a system of strict regulation at lower costs—benefiting both the environment and the economy.

Unfortunately, although there is considerable scope for such Pareto or near-Pareto improvements, in many expenditure programs trade-offs exist between the objectives of efficiency and equity (redistribution of income or benefits to the needy). It may be possible to design a more progressive expenditure program, but only at some cost. An increase in Social Security benefits may be desirable from the perspective of certain distributional goals, but the increased benefits may lead to earlier retirement, and the higher taxes required to finance them may decrease work incentives. Higher unemployment compensation may provide increased income to some who are among the most needy, but unemployment insurance may make some individuals feel disinclined to find another job.

Disagreements about the desirability of different programs often arise from disagreements not only about values—the relative importance of equity versus efficiency considerations—but also about the nature of the trade-offs, how much loss of efficiency would result from an attempt to change a program's structure of benefits to make its distributional impact more progressive.

Figure 10.6 shows the equity–efficiency frontier for a hypothetical program and the indifference curves for two individuals. In Figure 10.6A, Scrooge is much less willing to give up efficiency for a gain in equity than is his brother, Spendthrift. E_1 represents the point on the trade-off curve that is optimal as Spendthrift sees it, whereas E_2 is optimal from the point of view of Scrooge. Not surprisingly, Scrooge chooses a point with higher efficiency but lower equity than does Spendthrift. Thus, the source of the disagreement about policy is a difference in the values held by the two individuals.

On the other hand, Figure 10.6B depicts a situation in which the differences about policy arise from differences in judgments concerning the nature of the trade-off. Scrooge thinks that to get a slight increase in equity, one must give up a lot of efficiency. On the other hand, Spendthrift thinks that one can get a large increase in equity with just a slight loss in efficiency.

For instance, if the main reason that unemployed individuals do not obtain jobs is that there are no jobs available, then the size of unemployment insurance may have little effect on their search. If unemployment insurance has little effect on job search, there is not much trade-off between efficiency and equity, and the frontier is consistent with Spendthrift's perceptions. However, if job search is very sensitive to unemployment compensation, there is a significant trade-off, and the equity–efficiency frontier is consistent with Scrooge's perceptions.
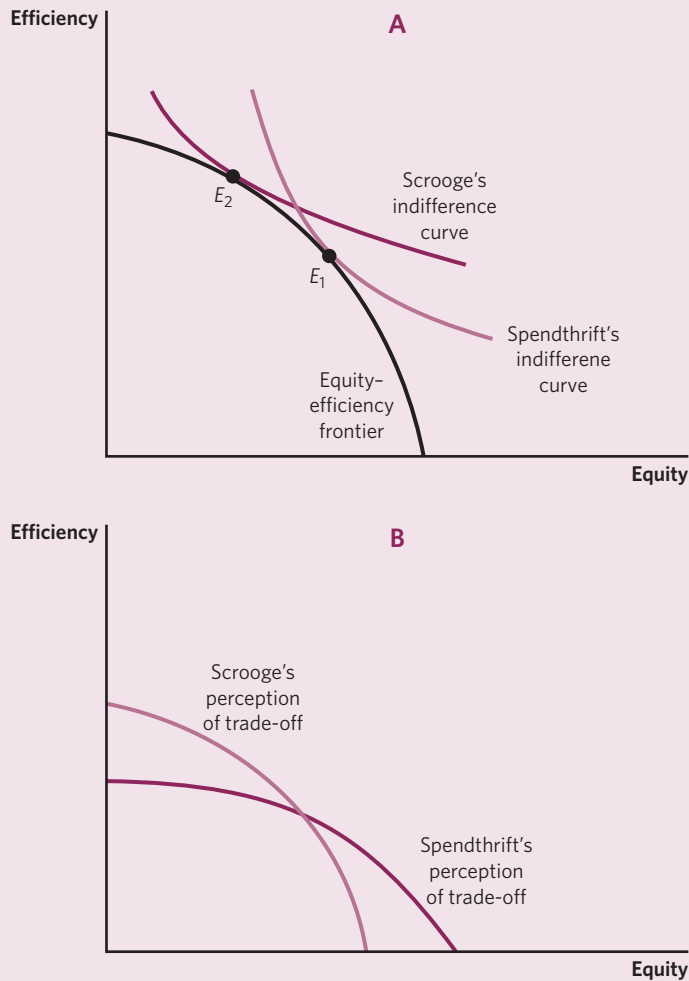
The equity–efficiency trade-off is encountered repeatedly in the evaluation of the detailed provisions of any government program. The decision to charge tolls on a bridge means that those who benefit from the bridge (that is, those who use it) have to bear its costs. To many people, this is desirable for equity reasons; it is unfair to make someone who does not drive over the bridge pay for it. However, there is an efficiency cost in money and time: the wages of toll collectors and the time of motorists. Moreover, if some drivers are discouraged from using the bridge (when it is below capacity), there is a further efficiency loss from underutilization.

EQUITY–EFFICIENCY TRADE-OFF

1. Sometimes programs can result in Pareto improvements, making some people better off without making anyone worse off.

2. More typically, there are trade-offs between equity and efficiency; more progressive tax systems reduce marginal incentives to work.

When programs are inefficient and not well designed, we can get more of both equity and efficiency. Later in this book, for instance, we describe attempts to help more Americans go to college by providing tax credits. Tax credits are, however, not typically the best way of accomplishing those objectives. Furthermore, they do not benefit the very poor—those from families that are so poor that they pay little, if any, taxes. Thus, there are alternative ways of providing assistance that are both more equitable and more efficient.

This example illustrates another theme: there is no single measure that captures all the dimensions of equity. Some programs may help more those of the middle class; others help those at the bottom. Both may improve "equity," but in two quite different senses, and these pose some of the hardest political choices.

PUBLIC POLICY OBJECTIVES

The discussion so far has focused on two bases for evaluating public programs: their effect on economic efficiency and their effect on distribution. However, government policy may be concerned with a broader range of objectives. For instance, government may be concerned with the extent to which individuals of different racial, ethnic, and class backgrounds are mixed together in schools. It may be concerned not just with the income of the poor, but also with the physical appearance of the housing in which they live. When these alternative objectives are fairly well defined, the government can still make use of a variety of instruments for attaining them. It can, for instance, still make use of private producers by imposing regulations on them or setting standards that must be met for individuals or firms to be eligible to receive subsidies. Thus, the government has specified that institutions receiving federal grants must comply with certain affirmative action regulations.

In some cases, however, it may be difficult for the government to specify clearly (and in advance) all of its objectives, or to articulate them in the form of a set of regulations or standards. There is widespread belief that private producers, in the absence of well-articulated and enforced regulations, will simply pursue profit-maximizing behavior, regardless of alternative objectives that they may affirm. In such circumstances, there is an argument for the government to assume direct responsibility for the activity. To the extent

that this is true, however, it may be difficult for Congress or the executive branch to clearly specify the objectives it wishes the bureaucrats responsible for implementing its programs to pursue. In that case, the bureaucrats will be left with considerable discretion, and the discrepancies between how they exercise that discretion and the intent of, say, Congress may be significant.

Similarly, there is concern that whenever the government finances an activity, it will almost inevitably impose a set of regulations, some of which may have adverse effects, particularly on economic efficiency; thus, many of the alleged efficiency advantages of private production may be lost. These concerns have been raised, for instance, in discussions of school voucher programs, which would provide students with funds that could be used at any school, private or public.

POLITICAL PROCESS

In a democracy, the design and adoption of any public expenditure program involves many individuals and groups with various objectives and various beliefs about how the economy works. The program that eventually is adopted represents a compromise among their views; it probably will not conform to the views of any one individual and may seem to be inconsistent with any single set of objectives. If two chefs, say, disagree about the appropriate liquid to add to a sauce, one arguing for lemon juice and the other for cream, the compromise solution of adding a little of both may be disastrous, with results inconsistent with any culinary objective.

The study of the political process by which a particular expenditure program was adopted may be insightful for several reasons. First, we may be able to understand why the program looks the way it does. Consider the government program to stabilize farmers' prices. This program addresses a market failure: the inability of individuals to obtain insurance for many of the important risks they face, including the risks associated with the variability of prices.[6] A closer examination of the price stabilization program, however, suggests that if that were the only objective, it would be designed in a quite different way. What farmers care about is *income* risk—not just the variability of price, but the variability of all the factors that go into determining net income, including output and costs. In some cases, the price stabilization program may actually increase income variability. In fact, reducing risk is probably not the true objective: the

---

[6] Futures markets now enable farmers to divest themselves of some of the risks associated with price variability.

real objective is to transfer resources (income) to farmers from the rest of the population. If that is the objective, there are more efficient ways of transferring resources to the farmers; outright grants would be preferable to the present program. However, if that objective were made explicit—if the transfers were made conspicuous—they might not get approved. Voters in urban districts might strongly oppose them, even though they do not oppose the present form of inefficient subsidies, simply because they are not fully aware of the nature of the transfers.

This example also highlights the dilemma that there are many different views of fairness. Some believe that anything they have earned is *theirs*, and it is *unfair* for the government to take it away; others recognize, perhaps more reasonably, that no one could have gotten to where he or she is without a vast array of help from others—we all depend on a variety of public services, including the rule of law, without which businesses could not flourish. Differences in these perspectives—and the difficulties of each group to understand others' perspectives—are a major impediment to reaching political consensus around a variety of issues.

Particular provisions of public programs are likely to have strong distributional consequences for particular groups in the population. If one group can be suitably organized, it will attempt to induce the political process to adopt provisions that are to its benefit. In Chapter 6, we discussed the regulations that provide for scrubbing the smoke emitted from burning coal. These regulations may have an enormous effect on the relative demand for hard (or western) coal and bituminous coal, and hence on the incomes of both miners and coal producers in different parts of the country. The shape of environmental legislation and regulation may be affected as much by these particular distributional consequences as by overall efficiency considerations.

A second reason why it may be helpful to study an expenditure program's adoption process is that in democracies, programs respond at least in part to the desires and perceptions of voters. Because programs have to be explained and "sold" to voters, there is a premium on simplicity. In addition, programs often look different from the way that economists think they should be designed because voters often do not understand the true incidence of a program. For instance, most voters think that half the cost of Social Security is paid for by contributions from the employer; most economists believe that the true incidence is the same as it would be if Social Security contributions were paid entirely by workers. In this case, the confusion over incidence has few consequences, but in many other programs, this confusion can have significant impacts, as we shall see in later chapters.

Equally important, however, the political process is shaped by special interests, and the programs that emerge reflect those special interests typically more than the rhetoric that is used to justify them. Consider America's farm programs discussed earlier. Such programs are often justified as helping poor family farmers—but the way that they are designed provides most of the subsidy to large farms, many of which are corporate farms, not family farms at all. Many ways of redesigning the programs would target more of the money on small, family farms, but such reforms are not likely to get the political support they need to be adopted.

Indeed, typically special interests try to enlist others in their cause, trying to persuade them that they are the real beneficiaries. The most egregious example of this was the corn ethanol subsidy, which was mainly of benefit to the ethanol producers, including Archer Daniels Midland (ADM), a large agro-business. ADM persuaded corn farmers that they were the beneficiaries, and with their support, this costly subsidy was adopted in 1978. In fact, though, in the early days of ethanol, the amount of corn used was so small that it had a negligible effect on corn prices and producers. ADM also tried to persuade environmentalists that it was good for the environment, but they were not so easily duped. The subsidy was supposed to be short-lived, just enough to help a new industry get started, but, in fact, it lasted for more than thirty years before being abolished at the end of 2011.

Finally, the design of programs may affect the extent to which they are subjected to political pressures or corruption. Corruption is an increasing concern in many countries. It can take a variety of forms. In modern democracies, special interests contribute to campaigns, often in an attempt to "buy" legislation that favors them; in many countries, bureaucrats use their discretionary powers to extend favors in return for bribes. In New York City, there have been extensive reports of bribes to building inspectors, more to ensure that they inspect the building in a timely way (so there will not be costly interruptions to construction) than to give approval to a substandard building. The more discretion that is left to bureaucrats, the more potential there is for the exercise of political influence and corruption.

Accordingly, in evaluating alternative policies, one needs to take into account the political process, what the legislation might look like after it has been subjected to the political process, and what the consequences of the program will be, knowing that it will be administrated by bureaucrats, probably not unlike those administering other government programs, and subject to the same kinds of incentives.

REVIEW AND PRACTICE

SUMMARY

There are ten major elements in the analysis of public expenditure programs:

1. Identifying a need, the source of demand for the government program

2. Identifying a market failure (if it exists) and ascertaining whether what is at issue is a concern for (the consequences of) the distribution of income or the provision of a merit good

3. Identifying alternative programs that might address the perceived problems

4. In ascertaining and evaluating the impacts of alternative programs, paying attention to the importance of particular design features

5. Identifying private sector responses

6. Identifying the efficiency consequences of alternative programs

7. Identifying the distributional consequences of alternative programs

8. Identifying the trade-offs between equity and efficiency considerations

9. Identifying the extent to which alternative programs achieve public policy objectives

10. Identifying how the political process affects the design and implementation of public programs

KEY CONCEPTS

Capitalization
Crowding in
Crowding out
Incidence
Income effects
Intertemporal distribution effects
Progressive
Regressive
Shifting
Substitution effects

QUESTIONS AND PROBLEMS

1. Explain how the following actual design features have an important effect on the consequences of government programs:

   a. The income ceiling for eligibility for SNAP benefits is reduced by expenditures on housing.

   b. Until recently, whether an individual between ages sixty-five and seventy was eligible for Social Security benefits depended on his or her income calculated on a month-by-month basis.

   c. An ex-spouse becomes eligible for Social Security benefits only if the marriage lasted at least ten years.

   Can you think of other instances in which particular design features have seemingly unintended consequences?

2. Who may be the actual beneficiaries of the following government program or proposed programs? That is, taking into account how individuals respond to the government program, who is actually better off as a result of the program?

   a. Medicare

   b. Housing subsidies for the poor

   c. Education loans

   Can you think of other programs whose actual beneficiaries may differ from those the program seemingly intended?

3. In Chapters 12 to 16, we will use the framework we have discussed in this chapter to analyze several different government programs. Before reading those analyses, see if you can answer the following questions for each program:

   a. What were the original sources of demand for the program? What perceived need was it intended to address?

   b. What are the market failures that gave rise to the program?

c. What are the possible forms of government intervention? Are there particular design features that have had, or currently have, an important impact on the program's effectiveness? How do private sector responses weaken, or reinforce, the intended effects of the program? What is the true incidence of the program?

d. What are the major efficiency consequences of the program?

e. Does the program entail any effective redistribution of income?

f. Are there important instances of trade-offs between equity and efficiency in the program's design?

g. What are some alternatives for meeting the program's objectives? To what extent might they do a better job, for example, by reducing distortions and increasing the equity of the programs?

h. How has the political process affected the nature of the present program?

4. Draw the budget constraint between housing and "other consumption" for an individual on SNAP, where the amount of food assistance the individual receives depends on his or her income net of housing costs. Does it make a difference whether the individual is consuming an amount of food equal to, less than, or greater than the SNAP allotment?

5. State governments effectively subsidize tuition in state universities and colleges. How might this affect the amount of education that individuals get? Is there a substitution effect? Is there a market failure that this program might be addressing? Are there alternative ways of addressing it?

11

EVALUATING PUBLIC EXPENDITURE

The preceding chapter set out the basic framework for analysis of government expenditure policies. In many cases, government wants more than a qualitative analysis; it needs a quantitative analysis. It needs to know not only that there is a rationale for government action; it also needs to know whether the benefits of the particular government action (project, regulation) will exceed the costs. For example, should the government:

- Build a bridge—and, if so, of what size?
- Construct a dam—and, if so, of what size?
- Institute more stringent regulations for flammability of mattresses?
- Institute more stringent regulations for licensing drugs?
- Extend the Washington, DC subway system?
- Declare certain portions of the Cape Cod seashore a national park?

After expenditures are authorized, government also needs to know whether public funds are expended efficiently and effectively. It needs to assess past government expenditure to improve future government performance.

This chapter describes how the government goes about making these evaluations. First, however, it is instructive to consider how a private firm makes decisions concerning which projects to undertake.

PRIVATE COST–BENEFIT ANALYSIS

Private firms continually must decide whether to undertake investments. The procedures they follow can be characterized in four steps:

1. *Identify the set of possible projects to be considered.* If a steel firm wishes to expand its production capacity, there may be a number of ways it can do this. There may be alternative technologies available for smelting iron ore, and there may be a number of alternative specialized forms of steel that can be produced. The first stage, then, is to list the various major alternatives.

2. *Identify the full consequences of each alternative.* The firm is concerned primarily with its inputs and outputs. Thus, it will determine the labor, iron ore, coal, and other materials required for each production alternative; it will assess the quality of steel that will be produced under each alternative; it will determine the quantity of various wastes that will be produced.

3. *Assign a value to each input and output.* The firm will estimate the costs of various kinds of labor (with various skills) over the lifetime of the plant; the costs of other inputs, such as coal and iron ore; the prices at which it can sell the steel (which will depend on the quality of the steel produced, which may in turn vary from project to project); and the costs of disposing of wastes.

4. *Add up the costs and benefits to estimate the total profitability of the project.* The firm will undertake the project with the highest profit (the maximum difference between benefits and costs)—provided, of course, that profits are positive (taking appropriate account of the **opportunity costs**, the return the firm's resources could obtain elsewhere). If profits for all contemplated projects are negative, the firm will undertake no project; it will invest its funds elsewhere.

PRESENT DISCOUNTED VALUE

The procedures just described seem simple and straightforward. Only one part requires some elaboration. The benefits and costs of the steel mill occur over an extended period of time. Surely the firm is not indifferent

when it comes to choosing between receiving a dollar today and receiving one in twenty-five years. How are the benefits and costs that accrue at different dates to be valued and compared?

The method used is based on the premise that *a dollar today is worth more than a dollar tomorrow.* Suppose the interest rate is 10 percent. If the firm receives $1 today, it can take that dollar down to the bank, deposit it, and have $1.10 at the end of the year. Thus, $1 today is worth $1.10 next year. The firm is just as well off receiving $1 today as $1.10 next year. If the firm invests the $1.10, it will have $1.21 at the end of the following year. Accordingly, the firm is indifferent between receiving $1 today and $1.21 in two years' time.

To evaluate projects with receipts and expenditures in future years, the firm multiplies those receipts and payments by a **discount factor**, a number (less than 1) that makes the future receipts and payments equivalent to current receipts and payments. The discount factor is smaller the further into the future the benefit is received. The discount factor for payments in one year is just $1/1 + r$, where r is the rate of interest[1] (in our example, $r = 0.10$, so the discount factor is $1/1.1 = 0.9$); for payments in two years' time, it is $1/(1 + r)^2$ (in our example, it is $1/1.21$). The value *today* of $100 to be received two years in the future is thus $100/1.21 = $82.64. We then add up the value of what is to be received (or paid out) in each year of the project. The sum is called the **present discounted value (PDV)** of the project. If R_t is the *net* receipts from the project in period t, and r the rate of interest, then if the project lasts for N years, its PDV is given by

$$\text{PDV} = R_0 + \frac{R_1}{1 + r} + \frac{R_2}{(1 + r)^2} + \cdots + \frac{R_t}{(1 + r)^t} + \cdots + \frac{R_N}{(1 + r)^N}.$$

Table 11.1 provides an illustration of how this might be done for a hypothetical steel mill lasting five years. (Most steel mills last much longer than that; this makes the calculations more complicated, but the principle is the same.) For each year, we multiply the net receipts of that year by the discount factor for that year. Notice the large difference between undiscounted profits ($1000) and discounted profits ($169). This difference is likely to be particularly large for long-lived projects entailing large initial investments; the benefits for such projects occur later in time (and therefore have lower value) than the costs, which occur earlier in time.

---

[1] To see this, compare what the firm has at the end of the year if it receives $\$100 \times 1/(1 + r)$ today. It takes the $\$100 \times 1/(1 + r)$ and invests it, receiving a return of r. Thus, at the end of the year, it has

$$
\begin{aligned}
&\$100 \times 1/(1 + r) && \text{the original amount} \\
+\ &r \times \$100 \times 1/(1 + r) && \text{the interest} \\
\hline
&(1 + r) \times [\$100 \times 1/(1 + r)] = \$100.
\end{aligned}
$$

Therefore, the firm is indifferent between receiving $\$100 \times 1/(1 + r)$ today and $100 next year.

TABLE 11.1 HYPOTHETICAL CALCULATION OF PROFITABILITY FOR A FIVE-YEAR STEEL MILL

YEAR	BENEFITS (RECEIPTS)	COSTS	NET PROFITS	DISCOUNT FACTOR	NET DISCOUNTED PROFITS
1		3000	−3000	1	−3000
2	1200	200	1000	$1/1.1 = .909$	909
3	1200	200	1000	$1/(1.1)^2 = .826$	826
4	1200	200	1000	$1/(1.1)^3 = .751$	751
5	1200	200	1000	$1/(1.1)^4 = .683$	683
Total	4800	3800	1000		169

SOCIAL COST–BENEFIT ANALYSIS[2]

The government goes through basically the same procedures as the firm in evaluating a project. There are, however, two critical differences between social and private **cost–benefit analysis**.

First, whereas the only consequences of a project that are of concern to the firm are those that affect its profitability, the government may be concerned with a much broader range of consequences. For example, it may be concerned with the ecological effects of a dam and with the impact of the dam on the river's recreational uses.

Second, whereas the firm uses market prices to evaluate what it has to pay for its inputs and what it receives for its outputs, there are two instances in which the government might not use market prices in evaluating projects:

1. When the outputs and inputs are not sold on the market, market prices do not exist. Market prices do not exist, for instance, for clean air, for lives saved, or for the preservation of wilderness in its natural state.

2. When there is a market failure, market prices do not represent a project's true marginal social costs or benefits. The prices the government

---

[2] There is a substantial body of literature on social cost–benefit analysis. A collection of 32 seminal articles written from 1963 to 1999 are compiled by two of the leading authorities on this topic in A. Harberger and G. Jenkins, eds., *Cost–Benefit Analysis*, vol. 152, *International Library of Critical Writings in Economics* (Cheltenham, UK, and Northampton, MA: Edward Elgar, 2002). Four classic books are: E. Mishan, *Cost–Benefit Analysis: An Introduction* (New York: Praeger Publishers, 1971); I. Little and J. Mirrlees, *Project Appraisal and Planning for Developing Countries* (New York: Basic Books, 1974); L. Squire and H. van der Tak, *Economic Analysis of Projects* (Baltimore: Johns Hopkins University Press, 1975); and M. Roemer and J. Stern, *The Appraisal of Development Projects: A Practical Guide to Project Analysis with Case Studies and Solutions* (New York: Praeger Publishers, 1975). Three more recent works are: E. Gramlich, *A Guide to Benefit–Cost Analysis*, 2nd ed. (Long Grove, IL: Waveland Press, 1997); R. Zerbe and A. Bellas, *A Primer for Benefit–Cost Analysis* (Cheltenham, UK, and Northampton, MA: Edward Elgar, 2006); and A. Boardman et al., *Cost–Benefit Analysis: Concepts and Practice*, 4th ed. (Upper Saddle River, NJ: Pearson–Prentice Hall, 2010).

uses to evaluate its projects must reflect the market failure. (Recall from Chapter 3 that in the absence of market failures, market prices do reflect marginal social costs and benefits; accordingly, in the absence of market failure, the government should use market prices in evaluating its projects.)

Social cost–benefit analysis is concerned with developing systematic ways of analyzing costs and benefits when market prices do not reflect social costs and benefits. In the following sections, we will look at how the government values benefits that are typically not monetized—such as the value of the environment, or of lives—and how the government values marketed goods and services when there are reasons to believe that important market failures exist, such as massive unemployment, which result in market prices that do not reflect social benefits and costs.

CONSUMER SURPLUS AND THE DECISION TO UNDERTAKE A PROJECT

Before turning to these issues, however, there is one other set of situations in which cost–benefit analysis plays an important role. Even when the price system is working well, so prices reflect marginal benefits and costs, a project may not break even—and thus would not be provided by the market—and yet total benefits exceed costs, so the project *should* be undertaken. Typically, these are projects that have large fixed costs, such as a bridge, or, more generally, projects that are large enough to have an effect on prices. Thus, market prices can be used for valuing projects only when projects are sufficiently small that they have a negligible effect on prices. In the case of a bridge, although it may leave prices in general unchanged, the "price" of crossing the river at that particular place can be thought of as being reduced from infinite (the good simply is not available) to zero.

Figure 11.1 shows the demand curve for a bridge. Even at a price of zero, only a certain number of trips across the bridge will be taken—denoted by point E. Assume that the capacity of a minimal-size bridge is C, which exceeds E, and that the marginal cost of using the bridge is zero. Then efficient utilization of the bridge requires a zero toll (price); any
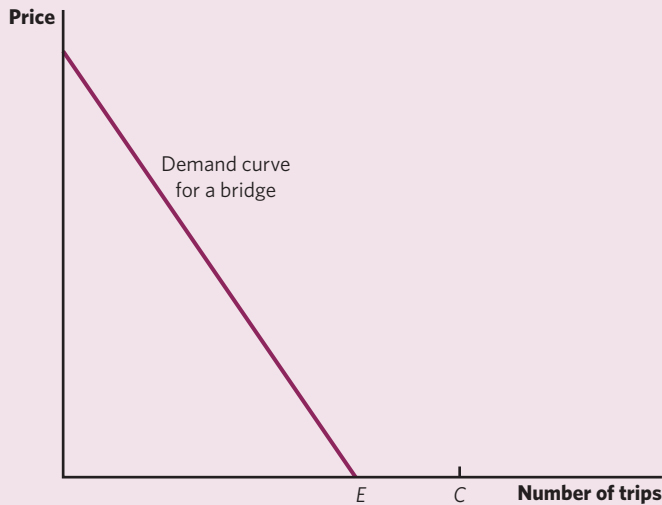
EFFICIENT UTILIZATION OF A BRIDGE

If the minimum scale capacity for a bridge, *C*, exceeds the demand at a zero price, *E*, then efficiency requires that no toll be charged, but it still may be worth constructing the bridge.

higher price will restrict usage, when the marginal cost of usage is zero. Clearly, though, at a zero price, no private firm would build the bridge.

Even though the marginal value of a trip is zero, however, the total value of the bridge is clearly positive. The question is: Is the total value large enough to offset the costs of the bridge? To find the total value of the bridge, we ask a simple question: How much, in total, would individuals be willing to pay to have the bridge (with a toll of, say, zero)? As we saw in Chapter 7, the total amount that individuals *would be willing to pay* in excess of *what they have to pay* is called the **consumer surplus**. There, we showed that consumer surplus is measured as the difference between the area under the compensated demand curve and what consumers actually have to pay for the good.[3]

---

[3] Recall that along the compensated demand curve, the individual's welfare (utility) is constant. The compensated demand curve tells us the quantity of the good that the individual demands at each price if, as the price is lowered, we take away just enough income to leave the individual no better off as a result of the price decrease. If the individual consumes relatively little of the good, then the amount we have to take away is relatively small. The difference between the compensated and uncompensated demand is the result of the "income effect"—the change in demand (here, for trips) from taking away this small amount of income. Accordingly, the difference between the compensated and uncompensated demand curves for an item like a bridge is typically small. See R. Willig, "Consumer's Surplus without Apology," *American Economic Review* 66 (1976): 589–597. Obviously, in other cases, such as the supply of labor (demand for leisure), the difference could be large. See J. Hausman, "Exact Consumer's Surplus and Deadweight Loss," *American Economic Review* 71 (1981): 662–676. Whether or not economists *should* ignore the income effect, in practice, they frequently do because of the difficulties in quantifying its magnitude. (As a matter of terminology, the area under the ordinary demand curve is often called the consumer surplus, as opposed to the *exact* consumer's surplus, which is the area under the compensated demand curve. The exact consumer's surplus is what is relevant for project evaluation.)

FIGURE 11.2

CALCULATION OF CONSUMER SURPLUS
The consumer surplus is the area under the (compensated) demand curve. If a zero toll is charged, the bridge should still be constructed if the consumer surplus exceeds the cost of the bridge. (If a toll of *P* is charged, then the consumer surplus is the area *AGB*, and the bridge should be constructed if the consumer surplus, which is now only *AGB*, plus the revenues raised, *FGBQ*, exceed the cost of the bridge.)
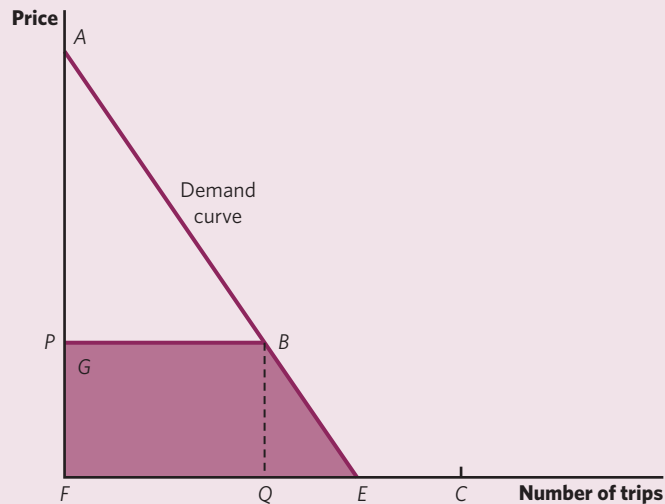
Figure 11.2 shows the compensated demand curve—everywhere along this line, each individual's welfare is the same as it would have been if no bridge had been constructed. If no tolls were charged, the consumer surplus would be the entire area under the demand curve, the area of the triangle AFE. This would measure the entire amount that all individuals would be willing to pay to have the bridge—say, the value of the savings in time and driving costs of using that bridge rather than using the bridge one mile downstream. If a toll P is charged, then the total people are willing to pay still exceeds the amount actually paid by the amount of the triangle AGB.

The decision to build the bridge is then a simple one: Do the total benefits (revenues *plus* consumer surplus) exceed the total costs, including any costs incurred in raising the revenue to finance the bridge?[4]

Sometimes economists look at the ratio of benefits to costs. The criterion for undertaking any project for which benefits, B, exceed costs, C,

$$\text{Undertake a project if } B > C,$$

can be rewritten: Undertake any project for which the ratio of benefits to costs exceeds unity:

$$\text{Undertake a project if } \frac{B}{C} > 1.$$

---

[4] There is one subtlety: the total costs should include not only the expenditures on the bridge, but also the *additional* costs associated with raising the required tax revenues to finance it.

Often, governments must choose *which* project among several to undertake. There may be a dam site on which only one dam can be constructed. Different dams may have different benefits and costs. In this particular case, the government should undertake the project in which the total *net benefit* from the project, the difference between the benefits and costs, is largest.

Choosing the project that maximizes the difference between benefits and costs is *not* the same as choosing the project that maximizes the cost–benefit ratio. A very small project with a small benefit and an even smaller cost could have a very high cost–benefit ratio, yet yield relatively small net benefits.[5]

COST–BENEFIT ANALYSIS: CRITERION FOR ACCEPTING PROJECTS

- A project should be undertaken if its total benefits exceed total costs, or if the cost–benefit ratio exceeds unity. Total benefits include the consumer surplus, the difference between what individuals would have been willing to pay and what they have to pay.

- If the government must choose one from among a set of projects (e.g., several alternative designs for a dam to construct), it should choose the project with the highest net benefits, not the highest cost–benefit ratio.

MEASURING NONMONETIZED COSTS AND BENEFITS

For many of the costs and benefits associated with government projects and regulations—such as lives saved, or environments protected—there are no market prices. Economists, therefore, have developed systematic procedures for estimating these values. In some cases, such as the value of time, we can make inferences about individuals' evaluations from market data and from their observed behavior in other contexts. In other cases, such as the value of the Grand Canyon, survey techniques have been employed. Many of these valuation techniques remain controversial.[6]

---

[5] The problem is that we have not included in our cost measure the opportunity cost of the dam site. If we correctly calculate the opportunity cost of the dam site, then there will be only one project for which the benefits exceed the total costs; this is, of course, the same project that we identified before as maximizing the net benefits from the dam site.

[6] For more on determining the benefits of environmental goods, see M. Cropper and W. Oates, "Environmental Economics: A Survey," *Journal of Economic Literature* 30 (June 1992): 675–740; "Symposium on Contingent Valuation," *Journal of Economic Perspectives* 8, no. 4 (Fall 1994): 3–64; "Economic Valuation of the Environment: A Special Issue," *Environmental Science & Technology* 34, no. 8 (April 2000): 1381–1461; and L. Wainger and M. Mazzotta, "Realizing the Potential of Ecosystem Services: A Framework for Relating Ecological Changes to Economic Benefits," *Environmental Management* 48 (October 2011): 710–733.

VALUING TIME

The old adage, "time is money," describes how most economists evaluate the savings in time resulting from an improved transportation system, such as a better subway system or road network. The typical approach is to attempt to ascertain the wage rate of those who use the transportation system; under certain ideal conditions, the wage provides a measure of an individual's evaluation of his or her own time. In simple economic models, an individual is pictured as making a choice between the amount of leisure and the amount of work undertaken. As a result of giving up one more hour of leisure, the individual gets an increase in consumption goods equal to his or her hourly wage. In equilibrium, the individual is indifferent when choosing whether to give up one more hour of leisure and increase consumption by an amount equal to his or her hourly wage, or to reduce work (increase leisure) by an hour and decrease consumption by an amount equal to his or her hourly wage. Thus, the individual's wage provides a monetary valuation of his or her time. If a faster subway reduces commuting time by twenty minutes, and the wage is $9 an hour, the value of the time saved is $3. We calculate the value of time saved by each individual and add the values together to obtain the total value of time saved.

Some claim that this overestimates the value of time: many individuals would like to work more at their wage rate but are unable to find additional employment at that wage; the job restricts the number of hours that they can work. The individual's valuation of his or her leisure is thus fairly low; the compensation that would be required for reducing an individual's leisure by one hour is, in this view, much less than the wage that the individual receives for the work that he or she is able to obtain.

Others claim that the wage may underestimate the value of leisure for some individuals and overestimate the value of leisure for other individuals. They point out that professors, for instance, have chosen a comparatively low-wage job relative to other options available to them because of the great non-monetary benefits associated with the job. The value of their leisure exceeds the wage they receive. On the other hand, the wage of the coal miner or the garbage collector includes some compensation for the unattractive features of those jobs and hence represents an overestimate of the value of leisure.

VALUING LIFE

Probably no subject in public cost–benefit analysis has engendered as much emotional discussion as economists' attempts to place a monetary value on life. As distasteful as such a calculation may seem, it is necessary,

CHILDREN, CAR SAFETY, AND THE VALUE OF LIFE

Even though life is priceless, economists have developed several alternative methodologies to put a dollar number on it (see the "Valuing Life" section of this chapter). Different studies produce different results, with a range of $1 to $25 million, but most studies show numbers at the lower end of that range, between $3 and $9 million (in 2012 dollars).*

The U.S. government has debated whether to use a single number for cost–benefit analyses in all agencies. So far, different agencies use different numbers—with the Environmental Protection Agency typically using numbers considerably larger than those of most outside studies to estimate the value of a statistical life (VSL).

Those who argue against using a single number suggest that a number of other factors should go into the analysis, such as whether the death resulted from an action voluntarily engaged in (such as driving).

One of the most difficult questions in valuing lives is whether the life of a child should be valued differently from the life of middle-aged adult, or of an 80-year-old. The issue comes up repeatedly: How should we allocate money between two cancer research programs, one targeting a cancer typically found in children and the other a form of cancer that typically only shows up among the elderly?

The Department of Transportation (DOT) is responsible for imposing regulations to ensure the safety of cars; in imposing these regulations, it looks at the costs and benefits. In the mid-1990s, DOT addressed whether to strengthen the standard for car frames that would largely affect deaths in the rear seat in side collisions. A disproportionate fraction of those saved would be children (as children more often ride in the rear seat). This raised the issue of whether to employ a different value of life instead of the one used elsewhere, when those saved were more typically adults. When a child's life is saved, more "life-years" are saved than when the life of an 80-year-old is saved.

DOT has continued to use the traditional method for valuing life, which treats all lives the same. Even so, an alternative methodology, which focuses on life-years saved rather than lives, was sanctioned in new federal guidelines established in 2003 with the issuance of Office of Management and Budget (OMB) Circular A-4, "Regulatory Analysis." OMB allows use of the value per statistical life-year (VSLY) approach for adjusting VSL estimates to reflect differences in remaining life expectancy. This entails calculating the value of each year of life extension, which adjusts VSL to reflect age differences.

---

*For surveys of the value of life, see pages 713–715 in M. Cropper and W. Oates, "Environmental Economics: A Survey," *Journal of Economic Literature* 30, no. 2 (June 1992): 675–740; Peter Dorman, *Markets and Mortality: Economics, Dangerous Work, and the Value of Human Life* (Cambridge, UK: Cambridge University Press, 1996); and W. Kip Viscusi, "The Value of Life: Estimates with Risks by Occupation and Industry," *Economic Inquiry* 42, no. 1 (January 2004): 29–48. For a detailed review of both the theory and application of VSL and VSLY, see L. Robinson, "How US Government Agencies Value Mortality Risk Reductions," *Review of Environmental Economics and Policy* 1 (Summer 2007): 283–299.

in a variety of circumstances, for governments to face up to this problem. There is virtually no limit to the amount that could be spent to reduce the likelihood of an accident on a road, or of death from some disease. At some point, however, a judgment must be made that the gain from further expenditures is sufficiently small that additional expenditures are not warranted. An individual may die as a result of this decision—yet we cannot spend 50 percent of our national income on transportation safety, or 50 percent of our national income on health.

Two methods have been used for estimating the value of life. The first is the *constructive method*—that is, we estimate what the individual would have earned had he or she remained alive (until the "normal" age of death). To do this, we extrapolate the individual's employment history, comparing it to that of those in similar positions.

This method, however, fails to distinguish between the value of life and the livelihood that goes with it. It thus suggests that after retirement, an individual's life has zero value, as there would be no loss of earnings. This seems clearly wrong. (It confuses means and ends: income is earned to provide consumption; producing income is not presumably the *object* of life, and therefore not the basis of valuing it.)[7]

There is an alternative, indirect method that does recognize the natural desire to live longer. In some occupations, there is a much higher chance of death than in others. For instance, the accident rates for coal miners are higher than those for college professors, and the death rates for workers in asbestos factories and jackhammer operators are much higher than those for clerical workers. Individuals who undertake riskier occupations normally require compensation for assuming these additional risks. By choosing the riskier occupation, they are saying that they are willing to face a higher chance of death for a higher income while they are alive.

The second method calculates the value of life by looking at how much income individuals are willing to expend to avoid potentially fatal risks, or need to compensate them for an increase in the chance of death. This is the approach used by the federal government, referred to as the "value of a statistical life" (VSL). For example, the Environmental Protection Agency uses an estimate of $7.4 million (in 2006 dollars), updated to the year of analysis, for how much people are willing to pay for small reductions in their risks of dying from adverse health conditions that may be caused by environmental pollution, regardless of the age, income, or other population characteristics of the affected population.[8] This is the same

---

[7] For an early critique of this method and one of the first developments of the second, indirect method, see T. Schelling, "The Life You Save May Be Your Own," reprinted in T. Schelling, *Choices and Consequences* (Cambridge, MA: Harvard University Press, 1984).

[8] For a detailed explanation of EPA's approach to mortality risk valuation, see http://yosemite.epa.gov/ee/epa/eed.nsf/pages/MortalityRiskValuation.html.

approach used by Joseph Stiglitz and Linda Bilmes in their analysis of the full budgetary and economic costs of the Iraq war—they assigned a VSL of $7.2 million (in 2007 dollars) to each death of a U.S. service member, and a prorated fraction of that amount to each serious injury.[9]

There is considerable controversy about this second method, however, just as there is about the first. Some believe that it grossly underestimates the value of life; they argue that individuals are not well informed concerning the risks they face.[10] Also, for well-known psychological reasons, individuals attempt to ignore the information they do have concerning the riskiness of their jobs.

As controversial as the estimates of the value of life may be, there appears to be no alternative to using them in the evaluation of projects that affect the likelihood of death.

> ## ALTERNATIVE METHODS OF VALUING LIFE
>
> 1. **Constructive method:** What would the individuals have earned had they remained alive?
> 2. **Revealed preference method:** How much extra income do individuals need to compensate them for an increase in the chance of death, as reflected in market wages for riskier jobs?

VALUING NATURAL RESOURCES

A question of increasing concern is how to value impacts on the environment. That issue was raised forcefully by the 1989 Exxon *Valdez* oil spill. Network news reports on the millions of dying otters, salmon, and birds brought into every American home the impact of the spill in a relatively remote area. If people had died, it is clear that Exxon would have owed the families huge amounts of money in compensation. Obviously, the relatives of the animals that had been killed had no standing in court to demand compensation. Many Americans, however, felt that Exxon should pay something, both to deter others from taking actions that might damage the environment and to compensate them for their perceived loss from the environmental injury. Using a relatively new technique called **contingent valuation**, courts valued the compensation that Exxon would have to pay at approximately $1 billion. This was compensation that went beyond the direct economic injury, for instance, to fishermen who lost their livelihood.

---

[9] J. Stiglitz and L. Bilmes, *The Three Trillion Dollar War: The True Cost of the Iraq Conflict* (New York: W. W. Norton & Company, 2008).

[10] Several studies have attempted to estimate the magnitude of workers' misperceptions and suggest that they may not be too large. See, for instance, W. K. Viscusi, *Risk by Choice: Regulating Health and Safety in the Workplace* (Cambridge, MA: Harvard University Press, 1983). For a general analysis of technical challenges in applying VSL, see O. Ashenfelter, "Measuring the Value of a Statistical Life: Problems and Prospects," Institute for the Study of Labor (IZA) Discussion Paper No. 1911, January 2006.

In contingent valuation, individuals are asked a series of questions intended to elicit how much they value the environmental damage or the preservation of some species. Many (but not all) individuals seem willing to pay something, for instance, to preserve whales or the spotted owl or other endangered species, or the Arctic National Wildlife area, even if they themselves do not directly come into contact with the species or do not visit the preservation area. These values are referred to as **existence values**. Even if each individual is willing to pay only a little, say, $5 or $10, when added up over all Americans, the values may be significant—in excess of $1 billion. This is what the court found in the *Valdez* case. Similar issues are now being litigated in respect to the 2010 BP oil spill in the Gulf of Mexico.

Although there is considerable controversy over the accuracy of these methods, a special panel set up by the National Oceanic and Atmospheric Administration, which included distinguished Nobel Prize winners Kenneth Arrow of Stanford and Robert Solow of MIT, recommended cautious use of the methodology by the government. In 1994, the government proposed new regulations for implementing contingent valuation, which are still being debated.[11]

SHADOW PRICES AND MARKET PRICES

Whenever there is a market failure, market prices may not reflect true marginal social costs or benefits. In such circumstances, economists attempt to calculate the true marginal social costs or benefits—for instance, of hiring an additional worker, or of importing or exporting additional goods. They call these "social prices" or "shadow prices." The term **shadow prices** reminds us that although these prices do not really exist in the market, they are the true social costs and benefits, reflected imperfectly in the market price.

---

[11] See M. Common, I. Reid, and R. Blaney, "Do Existence Values for Cost–Benefit Analysis Exist?" *Journal of Environmental and Resource Economics* 9, no. 2 (1997): 225–238; J. Duffield, "Nonmarket Valuation and the Courts: The Case of Exxon *Valdez*," *Contemporary Economic Policy* 15, no. 4 (October 1997): 98–110; K. Arrow et al., "Is There a Role for Benefit–Cost Analysis in Environmental, Health, and Safety Regulation?" *Environment and Development Economics* 2, no. 2 (May 1997): 196–201; P. Portney, "The Contingent Valuation Debate: Why Economists Should Care," *Journal of Economic Perspectives* 8, no. 4 (Fall 1994): 3–17; and three articles in *Journal of Economic Perspectives* 26 (Fall 2012): 3–56, namely, C. Kling et al., "From Exxon to BP: Has Some Number Become Better than No Number?" R. Carson, "Contingent Valuation: A Practical Alternative When Prices Aren't Available," and J. Hausman, "Contingent Valuation: From Dubious to Hopeless."

DIFFERENCES BETWEEN MARKET PRICES AND SHADOW PRICES

Shadow prices reflect true marginal social costs. When there are market failures, market prices may not fully reflect social costs.

   Examples are shown in the following table:

MARKET	DIFFERENCE BETWEEN MARKET AND SHADOW PRICES	EXPLANATION
Labor	Shadow wage is less than market wage when there is unemployment.	No loss in output elsewhere when an individual is hired; hence, marginal cost of hiring a worker is less than wage.
Capital	Shadow interest rate exceeds market interest when there is rationing in capital markets	Firm's expected return exceeds interest rate (firm would like to borrow more at given interest rate, but cannot). Thus, opportunity cost of funds is greater than the interest rate.
Steel	Shadow cost of production exceeds market cost.	Steel producer fails to value marginal social cost of pollution resulting from increased production.

In the absence of a market failure, the price of something equals its opportunity cost, what is forgone in alternative uses. In an economy in which there is massive unemployment, the market wage exceeds the opportunity cost—indeed, what is forgone is the individual's leisure—but when workers are unemployed involuntarily, the market wage exceeds the value of this forgone leisure, often by a considerable amount. The shadow price of labor when there is massive unemployment is the low value of the forgone leisure, not the market wage.

Similarly, in an economy in which capital markets work very imperfectly, and firms cannot raise additional capital at the "market rate of interest," the shadow cost of capital—what is forgone by using capital for one purpose rather than another—may exceed the market rate of interest by a considerable amount.

DISCOUNT RATE FOR SOCIAL COST–BENEFIT ANALYSIS

Our discussion of private cost–benefit analysis noted that a dollar next year or the year after is not worth as much as a dollar today. Hence, income to be received in the future or expenses to be incurred in the future must be discounted. In deciding whether to undertake a project, we look at its

present discounted value. The discount rate private firms use is *r*, the rate of interest the firm has to pay. What discount rate should the government use? The discount rate used by the government is sometimes called the **social discount rate**. The central question of concern is the relationship between this rate and the interest rate faced by consumers, on the one hand, and the rate faced by producers, on the other.

For evaluating long-lived projects, such as dams, the choice of the discount rate is crucial: a project that looks very favorable using a 3 percent interest rate may look very unattractive at a 10 percent rate. If markets worked perfectly, the market interest rate would reflect the opportunity cost of the resources used and the relative evaluation of income at different dates. However, there is a widespread belief that capital markets do not work well. Moreover, taxes may introduce large distortions, with large differences between before- and after-tax returns. Thus, it is not clear which of the various market rates of interest, if any, should be used: for instance, should it be the rate at which the government can borrow, or the rate at which the typical taxpayer can borrow?

If the individuals who benefit from the project are the same as those who pay the costs, we can simply use their *marginal rate of substitution*, how they are willing to trade off the reduction in current consumption for gains in future consumption. Because their marginal rate of substitution will be directly related to the rate of interest at which they can borrow and lend, in this case we can use that market rate of interest for evaluating costs and benefits in different periods. Often, though, the project has further ramifications—a public project may, for instance, displace a private project—and we then have to look at *all* the consequences, the *net* change in consumption.

If a public project displaces a private project of the same size, then the net reduction in consumption today from the project is zero. If both the public and private projects yield all their returns in the same period, then we can easily decide whether to undertake the public project: we should undertake it if its output exceeds that of the private project; or, equivalently, if its rate of return exceeds that of the private project. In this view, which, not surprisingly, is called the **opportunity cost view**—because the private project is the opportunity cost of the public project—it is the producer's rate of return that should be used in project evaluation.

Focusing on opportunity costs and focusing on consumers' marginal rates of substitution yield exactly the same result in economies in which there are no market failures, for then the marginal rate of substitution (which equals the rate of interest facing consumers) equals the rate of return on capital, or the producers' rate of interest (the opportunity cost). Problems arise when there are market failures or taxes and/or when

those who benefit from a project are different from those who pay for them.[12] Today, many economists argue that the appropriate rate of interest for government discounting may be *none* of the observed market rates of interest.

More generally, it is recognized that choosing the appropriate interest rate is an exceedingly complex matter, with very important consequences, especially when it comes to decisions that have effects long into the future. The effects of increases in greenhouse gases that most scientists believe will result in global warming, a rising of the sea level, and climate change will not be fully felt for decades. A discount rate of, say, 7 percent means that a dollar of cost 100 years ago is worth less than one cent now—we can essentially ignore it. So too, we will be dealing with the highly radioactive wastes from nuclear reactors for decades, or even centuries, but if we use a 7 percent discount rate, we would essentially ignore most of those future costs.[13]

In the more general case, there is no presumption that the ratio of the marginal valuation of an increase in consumption by one generation to that of another is related to *any* interest rate. One approach, in that case, is to use *social welfare functions*. We first introduced the concept of a social welfare function in Chapter 7 as a way of formalizing how consumption or income of different individuals could be compared. Exactly the same principles apply in comparing individuals over time as in comparing individuals at the same point of time, with one difference. In both cases, there is *diminishing marginal utility*, so if future generations have higher incomes than the current generation, the marginal valuation of a dollar of consumption to them is lower. But some economists believe that the welfare of individuals of future generations *at the same level of*

---

[12] In a few special cases, the fact that the benefits may accrue to different generations poses no problem. If the government has engaged in optimal intergenerational redistribution of income, then the marginal value of a dollar to every generation will be equal to the market rate of interest, and as long as the project is relatively small, we can evaluate the marginal benefits received by different generations using market rates of interest, just as we can when the impacts are felt by a single individual.

Similarly, if society consists of a set of family dynasties, in which each family optimally redistributes income from the current generation to succeeding generations, then the marginal value of a dollar to every generation should equal the consumer rate of interest. [See R. Barro, "Are Government Bonds Net Wealth?" *Journal of Political Economy* 82 (1974): 1095–1117.] The validity of this model has been strongly questioned. It implies, for instance, that when the government ran huge deficits in the 1980s, individuals increased their savings in a fully offsetting way. Thus, it implies that in the absence of these deficits, personal savings would have been negative. For an extensive discussion of these issues, see the symposium in the *Journal of Economic Perspectives* 3, no. 2 (Spring 1989); and K. Arrow et al., "How Should Benefits and Costs Be Discounted in an Intergenerational Context? The Views of an Expert Panel," Resources for the Future Discussion Paper No. 12-53, 2012.

[13] In a few special cases, there is a simple and clear solution—for instance, if the *only* imperfection in the market is *optimally chosen taxes* (in later chapters, we will describe in detail what is entailed by optimal taxes, but for now, we simply note that actual tax systems seldom comport even closely with optimal tax structures), then the producer's rate of return should be used in project evaluation. [See P. Diamond and J. Mirrlees, "Optimal Taxation and Public Production," *American Economic Review* 61 (1971): 261–278.] For a discussion showing how even slight changes in assumptions can lead to markedly different conclusions, see J. E. Stiglitz and P. Dasgupta, "Differential Taxation, Public Goods, and Economic Efficiency," *Review of Economic Studies* 39 (1971): 151–174.

CLIMATE CHANGE AND DISCOUNT RATES

As discussed in Chapter 6, one important policy issue facing the world over the coming decades will be how to respond to the threat of global warming resulting from increased concentrations of greenhouse gases such as carbon dioxide in the atmosphere. The effects go beyond just an increase in temperature: there are concerns about the rise in the level of seawater and increased weather variability.

In spite of international agreements to reduce emissions of greenhouse gases, within the United States there has been controversy about how much should be spent to reduce these emissions. The controversy arises in part because of discounting: most of the effects of global climate change will not be felt for a hundred years—and at a 7 percent discount rate, the value of $100 a hundred years from now is about 12 cents. It clearly will not be worth spending much today to avert even large costs in the future. On the other hand, at a 1 percent discount rate, $100 a hundred years from now would be worth more than $35.

A special working group of economists of the International Panel on Climate Change (IPCC), including Nobel Prize winners Kenneth Arrow and Joseph Stiglitz, argued that the appropriate methodology implied using a low interest rate for purposes of discounting for costs and benefits associated with climate change.* Future generations would be adversely affected if actions mitigating the pace of emissions of greenhouse gases were not undertaken, and there were no ethical grounds for valuing their welfare substantially less than the welfare of the current generation.

Critics of this view argued that future generations could be made whole "simply" by setting aside money today, investing it at the market rate of interest, and letting the amount accumulate to be used to address the costs of climate change.

The worry, however, was not only that estimates of future damage repair costs might be too low, and, more fundamentally, that there might be some damage that was irreparable at any cost, but also that countries would not set aside the funds. If they did not, then the appropriate trade-off analyzed by the IPCC committee was between consumption of the current generation and the welfare of future generations that would be adversely affected by climate change.

*J. Bruce, Hoesung Lee, and E. Haites, eds., *Climate Change, 1995: Economic and Social Dimensions of Climate Change,* Contribution of Working Group III to the Second Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge, UK, and New York: Cambridge University Press, 1996). More information on the IPCC can be found at http://www.usgcrp.gov/ipcc/.

income should be weighted less than the welfare of the current generation, simply because it exists in the future. The rate at which future generations' welfare should be discounted is referred to as the **pure discount rate**. Other economists, such as the distinguished Cambridge economist Frank Ramsey, argued that all generations should be given equal weight.

To see what is implied by this approach, assume that per capita income is increasing at the rate of 1.5 percent, and the *elasticity of marginal utility* is 1.

(The elasticity of marginal utility is the percentage decrease in marginal utility from a 1 percent increase in consumption. As we saw in Chapter 7, economists usually assume the elasticity of marginal utility is between 1 and 2.) Then if the pure discount rate is zero, the social rate of discount is 1.5 percent, roughly equal to the real interest rate on safe (government) securities, but considerably below the opportunity cost of capital.

More recently, the rate of increase in per capita income has been less than 1.5 percent. If it is 1 percent, then the discount rate is also smaller at just 1 percent. (This discussion assumes that we can ignore the distribution of income, but we should not do this. In recent years, the median family—the family such that half have an income that is higher and half lower—has seen its income essentially stagnate. If we value future income based on their marginal utility, the discount rate is close to zero.)

The question of the appropriate social rate of discount has become a hotly contested political issue. Those who are concerned about the environment and who see environmental impacts stretching out over decades, for example, believe strongly in low discount rates. For instance, in their view, simply because the effects of nuclear waste can be postponed for fifty or a hundred years is no reason to essentially ignore them—which a 10 percent discount rate effectively tells us to do.

Today, the federal government uses a real discount rate of 7 percent in cost–benefit analyses, except for the evaluation of projects whose impacts are felt over the very long run, like global warming, where lower discount rates are used. This partially reflects the opportunity cost view—the federal government uses this rate because it "approximates the marginal pretax rate of return on an average investment in the private sector in recent years."[14]

It is not surprising that the discount rate should be a subject of such *political* controversy. But why can't economists agree among themselves? Our discussion has highlighted several sources of disagreement about the economy and about the government, summarized in Table 11.2.

> ## THREE VIEWS ON THE SOCIAL DISCOUNT RATE
>
> 1. Reflects consumers' rate of time preference (the consumers' borrowing rate).
> 2. Reflects opportunity cost of capital (the producers' borrowing rate).
> 3. May reflect neither: in long-lived projects affecting different generations, when social marginal valuation of consumption of different generations may have nothing to do with observed interest rates, and in the absence of optimal intergenerational redistribution of income.

---

[14] This quotation is taken from Office of Management and Budget Circular No. A-94 Revised (October 29, 1992) which establishes guidelines and discount rates for cost–benefit analysis of federal programs still applicable today. Subsequent guidelines issued by the federal government in 1995 allowed for the use of lower discount rates for long-lived projects with impacts over many generations.

TABLE 11.2 SOURCES OF DISAGREEMENT IN DISCOUNT RATES

HIGH DISCOUNT RATE (OPPORTUNITY COST OF CAPITAL)	LOW DISCOUNT RATE (SOCIAL RATE OF DISCOUNT)
1. Government investment tends to displace private investment.	1. Assessing *net* impacts is typically far more complicated than just assuming that a dollar of public investment displaces a dollar of private investment.
2. Even in a world with distortions, everyone could be made better off if efficiency is maintained—this entails the rate of return on public projects' equaling that on private projects.	2. Assessing the desirability of a project must take into account *intergenerational distributional* effects as well as *efficiency* effects. (a) Programs' beneficiaries are often different from those who bear the costs. (b) Even if the government *could* in principle make everyone better off, the required compensations (for instance, to those who are adversely affected) are seldom made. Moreover, there may be high costs associated in providing compensation.
3. Even in the absence of government intergenerational redistributions, if parents leave bequests to their children, marginal valuations of consumption of different generations will be equalized (the dynastic model).	3. In the absence of optimal intergenerational redistribution, market rates of interest do not reflect marginal social valuations of dollars to different generations. Further, the dynastic model is implausible.
4. When market distortions are caused by optimal taxes, then efficiency is still desirable, so the government should use the opportunity cost of capital.	4. With market distortions, marginal rates of substitution (how individuals value a marginal dollar in different years) and marginal rates of transformation (the trade-offs facing firms) may differ markedly. With distortionary taxes, efficiency—as exemplified by using the private sector's opportunity cost of capital in the public sector—is desirable only under highly restrictive conditions.

THE EVALUATION OF RISK

The most common mistake in trying to cope with the uncertainties of the benefits and costs of a project is to argue that in the face of risk, the government should use a higher discount rate. Recall that the discount rate relates the value of a dollar at one date to its value at a later date. To see how increasing the discount rate may lead to absurd results, consider a

TABLE 11.3 EXAMPLE OF COST–BENEFIT ANALYSIS FOR RISKY INVESTMENT

YEAR	EXPECTED NET BENEFIT	CERTAINTY EQUIVALENT NET BENEFIT	TIME DISCOUNT FACTOR (10 PERCENT INTEREST RATE)	DISCOUNTED VALUE OF CERTAINTY EQUIVALENT NET BENEFIT
1	$−100	$−100	1	$−100
2	100	90	.91	81.90
3	100	80	.83	66.40
4	100	75	.75	56.25
5	−50	−75	.68	−51
Total	150	70		53.55

project that, at termination, requires an expenditure (say, an automobile must be towed to the junkyard). Assume that there is some uncertainty about the magnitude of that cost. We would normally think that this uncertainty would make the project less attractive than if we knew for sure what the termination costs were. However, consider what happens if we use a higher discount rate to offset the risk: the discount factor is lower, the present value of those costs is reduced, and the project looks more, not less, attractive. To use a higher discount rate confuses the evaluation of income at different dates with the evaluation of risk; these are two separate issues.

To evaluate risks, economists introduce the concept of **certainty equivalents**. Assume that there is some risky project. Next year, the output of the project may be worth $0 or $100; there is a fifty-fifty chance of each outcome. The *average* value is just $50 $\left(\frac{1}{2} \times \$100 + \frac{1}{2} \times \$0 = \$50 \right)$. If we dislike risk, however, we would clearly prefer a project whose return was a certain $50. In fact, we would prefer a project with a smaller average value, as long as the risk was smaller. If we would be indifferent in choosing between the risky project with an average value of $50 and a perfectly safe project with a value of $45, we would say that $45 is the certainty equivalent of the risky project with an average value of $50. To evaluate risky projects, then, we simply take the present discounted value of the certainty equivalents.[15]

Thus, to be acceptable, risky projects must earn a higher return than safe projects with the same certainty equivalent. The extra amount a risky project must earn to compensate is its **risk premium**.

We illustrate the procedure in Table 11.3, for a five-year project. We have assumed that the initial investment in the first period is certain. The benefits that accrue in years 2, 3, and 4 are increasingly uncertain,

---

[15] This methodology is not perfectly general. It requires that we be able to separate the analysis of risk at one date from that of other dates. For most practical purposes, however, it is sufficiently general.

reflected in the certainty equivalents. In the final year, the project is scrapped; there are large costs associated with the termination of the project. (Consider the problem of what to do with a nuclear power plant when its useful life has come to an end.) However, these costs are uncertain; hence, its certainty equivalent exceeds the $50 expected cost. (In contrast, had we employed a higher time discount rate to take account of risk, these uncertain scrapping costs would not have weighed very heavily in our cost–benefit calculation.)

To obtain the present discounted value of the certainty equivalent net benefit at any date, we multiply it by the time discount factor. To obtain the present discounted value, we add up the discounted certainty equivalent net benefits for the life of the project.

How should the government evaluate the risks associated with various projects? In some cases, such as the risks associated with the generation of electricity, it can look to how private markets value risks. However, for risks for which there is no comparable private project, matters are more difficult. Some, such as a flood control project, serve to *reduce* the risks individuals face, and for these projects, the risk premium is negative. Individuals are willing to pay something to reduce the risk of flood. Because the government can spread risks over the entire population, when the project neither serves an insurance function (reducing the risks individuals would otherwise face) nor provides a return that is correlated with income from other sources (that is, the return to the project is neither particularly high nor particularly low when the economy is, say, healthy), the government should employ no risk premium.

Sometimes, government projects essentially provide insurance; that is, they reduce the risks that society would otherwise face. A dam reduces the risk of floods. Individuals would have been willing to pay a high premium for the reduction of such risks. In this case, the risk premium is negative; that is, the project should be undertaken, even if the expected return is lower than the return on a perfectly safe project.

Expenditures to reduce the risk of global warming provide an important instance for which public expenditures should arguably be undertaken even if the average return is lower than on a safe project—that is, the government should use a discount rate lower than that for a safe project. One way of thinking about this is the following: such investments yield a high return when it turns out that the effects of (and costs of) global warming are high—when sea level rises more than we had expected, or the increase in temperature or the variability in climate is greater than we had anticipated. In those instances, our standards of living will be lowered, as we will have to spend a great deal of money dealing with the consequences of global warming and making investments to help us adapt to

these changes. However, that means that the marginal utility of a dollar will be higher. In short, when the efforts at reducing global warming pay off, we value those dollars especially highly.

RISK ASSESSMENT

An area of increasing scrutiny—and controversy—in risk analysis is the risks to health, safety, and indeed life, that are posed, for instance, by hazardous wastes, pesticides, and fungicides. Chemicals in the water and air increase the likelihood of cancer and a variety of other ailments, often life-threatening. About this there is little doubt. The debate has focused on **risk assessment**—on how the magnitude of these risks is assessed and how priorities for reducing these risks should be established.

For instance, many risks are related to exposure. A chemical in dirt that is sealed under a thick layer of concrete is unlikely to impose significant risks; there would be a much higher risk if that same dirt were ingested directly by a child. In assessing the overall risk, one must take into account the probability of different levels of exposure, as well as the risks associated with each level of exposure.

The Environmental Protection Agency, in setting its priorities and its standards—for instance, for cleaning up hazardous wastes—has been criticized on several grounds. Rather than determining "compounding of probabilities" effects the way students are typically taught in modern statistics courses, the EPA uses a "worst case scenario analysis," which looks at the risks associated with the worst possible case. For example, what would be the risk assuming that the concrete seal around the dirt cracked, and a child wandered into the site? There have been some famous stories of the EPA insisting on cleanups to the standard that a child could eat the dirt for a six-week period without having any significant increase in health risk.[16] In setting priorities, there has been concern that the government has not gone after the highest risks, but rather the risks that have the most "popular appeal." The risks addressed by the EPA are often far lower than the risks that individuals take in their day-to-day lives—for instance, from drinking alcohol moderately, let alone from smoking. There is, however, one critical distinction: the risks on which the EPA focuses are those (like air and water pollution) over which individuals have no choice; they are incurred involuntarily, as opposed to the risks associated with smoking and drinking. Still, the fact that individuals seem willing to incur certain risks reveals information about their

---

[16] See Stephen G. Breyer, *Breaking the Vicious Circle: Toward Effective Risk Regulation* (Cambridge, MA: Harvard University Press, 1993), p. 12.

valuation of the risks, a fact that government should presumably take into account when adopting environmental risk standards. Recent government regulations have put greater emphasis on assessments of comparative risks; there is a reluctance to impose costly regulations to reduce risks that are of the size that individuals seem willing to accept in ordinary circumstances.

There has been increasing concern that environmental risks are borne disproportionately by the poor, who often live in industrial areas with heavier pollution. This is not surprising, as land in such areas typically is less valuable, so lower-income individuals can obtain housing at lower costs. President Clinton signed an executive order on **environmental justice** in 1994, instructing agencies to ascertain the distributional impact of environmental measures they might undertake. After more than a decade of federal government inertia in advancing environmental justice, in 2011 President Obama strengthened efforts to implement the 1994 executive order by issuing a memorandum of understanding signed by the heads of seventeen federal agencies commiting each agency to "identify and address, as appropriate, any disproportionately high and adverse human health or environmental effects of its programs, policies and activities on minority populations and low-income populations."[17]

DISTRIBUTIONAL CONSIDERATIONS

The benefits of any given public project are not uniformly distributed across the population. Some projects, such as a dam, have benefits that are limited geographically. Other projects, such as bilingual education programs and jobs retraining programs, are directed mainly at particular groups (e.g., immigrants and the unemployed). The government is clearly concerned about the impact of its programs on the distribution of income.

Should these distributional effects be taken into account in cost–benefit analysis? If so, how can they be quantified?

The issue of whether government should take distributional effects into account is analogous to the issue of whether, in choosing a social discount rate, the government needs to be concerned with the impact on different generations. If the "social" marginal value of a dollar to all individuals is the same, we can simply add up the dollar value of the impacts on consumption of different individuals. However, there is a widespread presumption that the social marginal value of a dollar to a poor individual is greater than it is to a rich individual.

---

[17] Memorandum of Understanding on Environmental Justice and Executive Order 12898.

The first step in any distributional analysis is to ascertain as precisely as possible how the program affects individuals in different circumstances. Typically, the focus is placed on individuals of different incomes, although regional impacts are also frequently taken into account. Of two programs with similar overall impact, the one in which more of the benefits and fewer of the costs accrue to poor individuals would presumably be preferred, if society cares about distribution.

Often, however, there is a desire to go beyond simply enumerating the impacts on different groups, to obtain a broader picture. This is done in two different ways. The first uses the social welfare function approach referred to earlier. It recognizes that the marginal valuation of a dollar is greater to a poor person than to a rich person, and uses the concept of the elasticity of marginal utility to quantify the extent to which this is so. For example, using an elasticity of unity (1), and giving a weight of unity to those at median income, impacts on those with half the median income receive a weight of 2, whereas impacts on those with twice the median income receive a weight of $\frac{1}{2}$. Using these weights, the total "weighted benefit" is calculated, and of two programs with the same costs, the one with the highest weighted benefit is undertaken.

The second approach looks at the impact on the overall distribution of income or wealth.[18] However, this approach lends itself only to major programs with the capacity for substantial distributional effects, such as changing the welfare system or the tax system. Most projects undertaken by the government are smaller in scale.

> ## KEY ISSUES IN MEASURING A PROJECT'S BENEFITS
>
> 1. Measuring consumer surplus
> 2. Measuring nonpecuniary benefits:
>   > Valuing time
>   > Valuing life
>   > Valuing the environment
> 3. Valuing marketed goods in the presence of market failure:
>   > Using shadow prices to measure marginal social costs when market prices do not accurately measure it
> 4. Valuing consumption (output) at different dates:
>   > Choosing the right discount rate
> 5. Valuing risk
> 6. Valuing distributional considerations:
>   > How are impacts on different groups to be compared?

COST EFFECTIVENESS

In some cases, it is difficult to compare costs and benefits. The benefits may be improved health; the costs are dollars expended. Although we have emphasized the necessity of making hard—monetary—judgments concerning life and health, the political process often tries to avoid making such

---

[18] This approach relies on measures like the Gini coefficient, discussed in the appendix to Chapter 7.

judgments, when possible. **Cost effectiveness (CE)** analysis provides a way of doing this by looking at programs with the same (or similar) benefits, and asking which produces those benefits at the least cost.

Assume that we wish to avoid the problems associated with valuing lives while helping the government assess a variety of ways of reducing highway deaths. We could calculate the costs associated with each of several methods of accomplishing the same goal. Or we might simply show the marginal costs associated with incremental reductions in the death rate under each method, and leave it to the legislators to determine which point along the curve should be chosen (and, therefore, what method of improving traffic safety should be chosen).

When the Occupational Safety and Health Administration considered standards for noise pollution, it did a CE study, calculating how many extra workers would be protected from hearing loss as a result of alternative standards. It then calculated the cost associated with each standard. From this information, it calculated the marginal gross and net costs (taking into account the fact that hearing losses reduce productivity) associated with different levels of protection, as depicted in Figure 11.3.

FIGURE 11.3

COMPARISON OF ALTERNATIVE STANDARDS FOR OCCUPATIONAL NOISE EXPOSURE

Higher standards cost more and protect more workers from hearing loss.
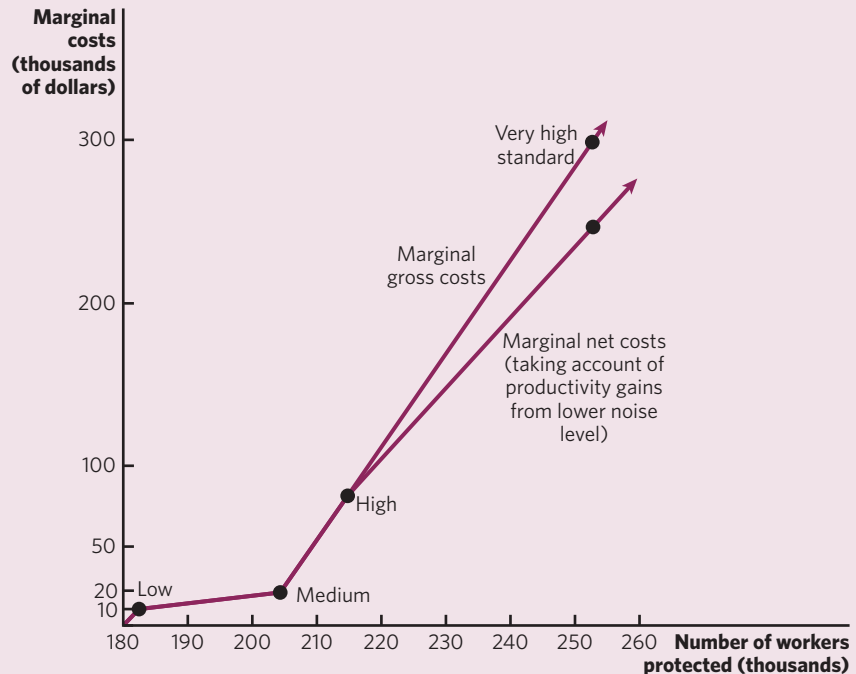
SOURCE: J. R. Morrall III, "Exposure to Occupational Noise," in *Benefit–Cost Analyses of Social Regulation*, ed. James C. Miller III and Bruce Yandle (Washington, DC: American Enterprise Institute for Public Policy Research, 1979).

The curve shows that there are significant extra costs of trying to protect additional individuals from hearing loss. On the basis of this, one study concluded, "an effectively administered hearing-protector program could provide most of the benefits at much lower cost in comparison with an industrywide engineering-only noise standard. . . . [A]n 85-decibel hearing-protector standard [has] the relatively reasonable marginal cost of about $23,000 per hearing impairment avoided." In ordinary English, the study recommended the use of earplugs rather than the drastic changes in plants and equipment that would be required to implement the same level of hearing protection.

Table 11.4 shows another example of CE studies, this time comparing the cost effectiveness of alternative medical interventions. There is an enormous range of cost effectiveness ratios, from $2158 per life-year saved for administering a low dose of the drug lovastatin to reduce cholesterol for males between ages 55 and 64 who had prevalent coronary heart disease, to $41,000 for annual breast examination and mammography for females aged 55 to 65, to $88,000 for a coronary artery bypass graft for someone with a single-vessel disease with moderate heart weakness, to $335,000 for the use of an exercise cardiogram as a screening test for heart disease for 40-year-old females.

The use of CE analysis to guide resource allocation in health care has grown substantially in Europe over the past decade, but there has been considerable resistence to explicit consideration of costs in the policy coverage criteria of major U.S. health schemes such as Medicare. More attention, though, is now being given in the United States to the effectiveness component of CE analysis through support of comparative effectiveness research (CER), exemplified by the provision of $1.1 billion under the American Recovery and Reinvestment Act of 2009 (ARRA) for CER. However, the primary objective of CER is to improve the quality of health care by providing information on which medical interventions work best for whom under what circumstances, without direct reference to relative costs.[19]

Although CE analysis is simpler than cost–benefit analysis because it avoids all the problems of measuring and valuing benefits, most of the issues discussed in measuring and valuing benefits remain, scaled down, for measuring and valuing costs. For instance, shadow prices for inputs may differ from market prices; a social discount rate must be used to value costs incurred at different dates; and there is considerable uncertainty—for example, we might be unsure of the exact degree to which hearing loss hurts productivity, or how much it will cost to bring a new weapon to completion, or how much it will cost firms to comply with stricter environmental standards.

---

[19] For an update on the study cited in Table 11.4, see Alan M. Garber and Charles E. Phelps, "Future Costs and the Future of Cost-Effectiveness Analysis," *Journal of Health Economics* 27 no. 4 (2008): 819–821.

TABLE 11.4 ESTIMATED COST EFFECTIVENESS OF COMMONLY USED MEDICAL INTERVENTIONS (ALL INTERVENTIONS COMPARED TO "USUAL CARE" UNLESS OTHERWISE NOTED)

INTERVENTION	COST/LIFE-YEAR ($1993)
Low-dose lovastatin for coronary heart disease prevention[a]	
Males, age 55–64, cholesterol level ≥ 250	2,158
Males, age 55–64, cholesterol level < 250	22,929
Female nonsmokers, age 35–44	2,023,440
Exercise electrocardiogram as screening test[b]	
40-year-old males	124,374
40-year-old females	335,217
Hypertension screening[c]	
40-year-old males	27,519
40-year-old females	42,222
Breast cancer screening[d]	
Annual breast examination and mammography, females age 55–65	41,008
Physician advice about smoking cessation[e]	
1% quit rate, males age 45–50	3,777
Pap smear starting at age 20, continuing to 74[f]	
Every 3 years, versus not screening	24,011
Coronary artery bypass graft[g]	
Left main coronary artery disease	8,768
Single-vessel disease with moderate angina	88,087
Neonatal intensive care units[h]	
Infants 1000–1500 g	10,927
Infants 500–999 g	77,161

[a] L. Goldman et al., "Cost-Effectiveness of HMG-GoA Reductase Inhibition for Primary and Secondary Prevention of Coronary Heart Disease," *Journal of the American Medical Association* 265 (1991): 1145–51.
[b] H. C. Sox Jr. et al., "The Role of Exercise Testing in Screening for Coronary Artery Disease," *Annals of Internal Medicine* 110 (1989): 456–69.
[c] B. Littenberg et al., "Screening for Hypertension," *Annals of Internal Medicine* 112 (1990): 192–202.
[d] D. M. Eddy, "Screening for Cervical Cancer," *Annals of Internal Medicine* 113 (1990): 214–26.
[e] S. R. Cummings et al., "The Cost-Effectiveness of Counseling Smokers to Quit," *Journal of the American Medical Association* 261 (1989): 75–79.
[f] D. M. Eddy, "Screening for Breast Cancer," *Annals of Internal Medicine* 111 (1989): 389–99.
[g] M. C. Weinstein, "Economic Assessment of Medical Practices and Technologies," *Medical Decision Making* 1 (1981): 309–30.
[h] M. H. Boyle, G. W. Torrance, J. C. Sinclair, and S. P. Horwood, "Economic Evaluation of Neonatal Intensive Care of Very-Low-Birth-Weight Infants," *New England Journal of Medicine* 308 (1983): 1330–37.

SOURCE: Alan M. Garber and Charles Phelps, "Economic Foundations of Cost-Effectiveness Analysis," *Journal of Health Economics* 16, no. 1 (1997): 1–31.

Cost–benefit and CE analysis are important tools used by policy makers throughout the world. They provide discipline to the decision-making process. Although critics complain that they reduce everything to cold calculations, these tools can be used to bring systematically into the analysis not only economic costs and benefits, but also concerns about the

environment, health, and distribution. Although there never will be complete precision, especially in these hard-to-quantify areas, judgments will be made weighing these various considerations, and quantification can be a helpful step in resolving the complicated trade-offs that must be faced.

POST-EXPENDITURE EVALUATION: ASSESSING AND IMPROVING GOVERNMENT PERFORMANCE

Thus far, this chapter has been devoted to pre-expenditure project appraisal. It has explored the use of cost–benefit analysis to compare expected future benefits with anticipated costs to produce these benefits, as well as the use of CE analysis to estimate the cost effectiveness of alternative potential future interventions.

However, it is also important to undertake post-expenditure evaluation to assess and improve government performance: Did the government spend its funds wisely? Did the public get good value for its tax payments? How can expenditure policies and practices be improved?

In addition to evaluating the economic efficiency and distributional effects of public expenditures (see Chapter 10), we can also evaluate public expenditures in terms of meeting policy objectives. The broadest form of this type of evaluation is a national **public expenditure review (PER)**, commonly conducted by the World Bank. PERs provide an overview of the level, composition, and results of public expenditures, and usually entail examination of (1) aggregate spending levels in the context of a country's macroeconomic policies; (2) the impact of sectoral, geographic, and demographic allocation of spending on poverty reduction; (3) the respective roles of government and the private sector in the financing, production, and provision of public infrastructure and services, with a focus on the need for public sector mitigation of market failures while not crowding out the private sector; (4) the balance between capital investments to increase future capacity and recurrent expenditures to operate and maintain existing infrastructure; and (5) quality of the expenditure process in terms of transparency and accountability. PERs are expensive and complex endeavors, so they are usually undertaken periodicially (about every five years) rather than annually.[20]

---

[20] For a detailed description of PERs, see Sanjay Pradhan, "Evaluating Public Spending: A Framework for Public Expenditure Reviews," World Bank Discussion Paper 323, Washington, DC, 1996, and the World Bank public expenditure website www.worldbank.org.

TAKING A BITE OUT OF CRIME IN THE BIG APPLE

Although traditional indicators of police performance, such as response times and arrest rates, measure operational efficiency, they do not reveal whether police departments are actually making their communities any safer. Thus, New York City Police Commissioner William Bratton and Deputy Commissioner Jack Maple created CompStat in 1994 as a management tool to both measure and improve performance in crime reduction.

The centerpiece of CompStat is frequent crime control strategy meetings between all precinct commanders and New York Police Department (NYPD) top management to discuss the latest crime statistics. The key is to creatively compile and analyze crime data, and then make these data accessible in a timely manner to front-line police officers so they can use it to achieve NYPD policy objectives. CompStat generates a considerable amount of information over time and by police precinct, so

all can see the latest trends, try to understand the underlying causes of these trends, assess results of follow-up responses to previously identified problems, and devise strategies for coping with performance shortcomings. CompStat also produces "commander profile reports," so senior NYPD executives can evaluate the internal management performance of their field leaders.

CompStat is widely credited with helping to dramatically reduce crime in New York City, and has been adopted by police forces throughout the United States and abroad—for example, in Australia. CompStat has also been adapted to other public agencies and even entire jurisdictions, giving rise to the term "PerformanceStat"—for example, JobStat (New York City), EdStat (Rhode Island), FEMAStat (Federal Emergency Management Agency), CitiStat (Baltimore, Somerville, Atlanta), and StateStat (Maryland).

SOURCES: Eli Silverman, "CompStat's Innovation," in *Police Innovation: Contrasting Perspectives*, ed. David Weisburd and Anthony Braga (Cambridge, UK: Cambridge University Press, 2006); and Robert D. Behn, "Designing PerformanceStat," *Public Performance and Management Review* 32, no. 2 (December 2008): 206–235.

We can also evaluate public expenditures made to deliver specific goods and services. The most common methodology for doing this is **comparative performance measurement (CPM)**, which is a tool to determine expenditure efficiency by assessing the total, unit, or relative costs of inputs and measure expenditure effectiveness by evaluating the quantity and quality of both short-term outputs (intermediate results) and long-term outcomes (achievement of policy objectives). The key to CPM's usefulness is to place these measurements in a comparative perspective—that is, to track them over time, as well as between sectors, institutions, or jurisdictions. This allows a unit of government to see how well it is performing by comparing itself to similar agencies or localities; communicate past performance to its constituents and prioritize future expenditure needs in consultation with these constituents;

and revise its policies, management, and operations to improve future performance.[21]

For example, if a city wants to evaluate its expenditures on secondary education, it can analyze how much it spent on this in total for the past five years (aggregate costs), how much it spent per teacher or per student (unit costs), or how much it spent versus primary education or public health expenditures (relative costs), all compared with expenditures of similar cities. Although such an analysis does not indicate how well this money is being spent (value for money), it can reveal whether production costs seem to be in line with established norms—that is, with what is going on elsewhere.[22]

A city could go further and also measure expenditure effectiveness. Again using secondary education as an example, it can measure, in a comparative time series, short-term results such as the number of students who attend school or who graduate (aggregate outputs), as well as teacher qualifications or student–teacher ratios (which are supposed to be proxies for quality). It can also try to determine whether students are actually getting a good education, the primary policy objective of secondary education expenditures, by using performance standards or benchmarks such as standardized test scores or college enrollment rates. The difficulty, however, is that even these measures do not really answer the question in which we are really interested: Have our schools made their students more productive, with higher incomes, more creative, better citizens, contributing to rising living standards? It typically takes decades before we know the answers to such questions, and then, there are many other factors affecting outcomes. There is also considerable debate on whether measuring basic skills distracts from investments in more important outcomes, such as cognitive development. Still, there is a broad concensus that these indirect indicators, as imperfect as they are, provide some insight into the extent to which schools are succeeding in their mission.

CPM is a simple but powerful tool for public institutions to generate and analyze data in a form that facilitates learning from both their own experience and the experiences of others with similar missions and challenges.

---

[21] For a detailed explanation of CPM and examples of the application of CPM, see Elaine Morley, Scott P. Bryant, and Harry P. Hatry, *Comparative Performance Measurement* (Washington, DC: The Urban Institute Press, 2001).

[22] A more sophisticated application of CPM to assess expenditure efficiency is data envelope analysis (DEA). This is used most extensively to assess efficiency in the provision of municipal services, such as street repair, building maintenance, fleet management, solid waste collection, water and sewerage provision, and public transit. DEA uses a nonparametric measurement of technical efficiency, defined as production of the most outputs from the fewest inputs. For a detailed explanation of DEA and examples of the application of DEA, see Adrian Moore, James Nolan, and Geoffrey F. Segal, "Putting Out the Trash: Measuring Municipal Service Efficiency in U.S. Cities," *Urban Affairs Review* 41, no. 2 (2005): 237–259. For an international adaptation of DEA, see António Afonso, Ludger Schuknecht, and Vito Tanzi, "Public Sector Efficiency: An International Comparison," European Central Bank Working Paper No. 242, Frankfurt am Main, 2003.

Considerable progress has been made around the world over the past two decades in the quantitative measurement of public sector performance under the general rubric of "new public management" as governments continue to search for ways to do things "better, faster, cheaper."[23] For example, widespread adoption of performance-based budgeting (PBB) reflects the desire to allocate resources based on documented results rather than previous levels of input—PBB allocates resources based on the achievement of specific, measurable short-term outputs and long-term outcomes. The growth of PerformanceStat initiatives takes the concept further, as they incorporate performance *measurement* into integrated performance *management*—leadership uses data on how well people are doing their jobs as part of an integrated management information system to improve overall institutional performance (see case study, "Taking a Bite Out of Crime in the Big Apple"). The hope is that a public leadership strategy based on the utilization of objective and transparent performance measurement will make government more accountable to the public for the way it spends tax revenue, and provide the public sector with incentives to replicate private sector efficiency and effectiveness in service of the public interest.

---

[23] This is the name of a popular website for state and local governments that provides "innovative ideas for government that deliver public value and lower the cost of government services," http://www.governing.com/blogs/bfc.

REVIEW AND PRACTICE

SUMMARY

1. Cost–benefit analysis provides a systematic set of procedures by which a firm or government can assess whether to undertake a project or program and, when there is a choice among mutually exclusive projects or programs, which one to undertake.

2. Private cost–benefit analysis entails determining the consequences (inputs and outputs) associated with a project, evaluating these using market prices to calculate the net profit in each year, and finally, discounting profits in future years to calculate the present discounted value of profits.

3. Social cost–benefit analysis involves the same procedures as private cost–benefit analysis, except that a broader range of consequences is taken into account, and the prices at which inputs and outputs are evaluated may not be market prices, either because the inputs and outputs are not marketed (so market prices do not exist) or because market prices do not accurately reflect marginal social costs and benefits, due to a market failure.

4. When the government makes available a good or service that was not previously available (e.g., constructs a bridge across a river), the value of the project is measured by the consumer surplus it generates; this is the area under the (compensated) demand curve.

5. The government has to make inferences (based on market data or observed behavior) concerning the valuation of nonmarketed consequences, such as lives and time saved or impacts on the environment.

6. The rate of discount used by the government to evaluate projects may differ from that used by private firms.

7. To evaluate risky projects, the certainty equivalent of the benefits and costs needs to be calculated.

8. Distributional considerations may be introduced into evaluations, either by weighting the benefits accruing to different groups differently or by assessing the impact of the project on some measure of inequality.

9. It is also important to undertake post-expenditure evaluation to assess and improve government performance.

KEY CONCEPTS

Certainty equivalents

Comparative performance measurement (CPM)

Consumer surplus

Contingent valuation

Cost–benefit analysis

Cost effectiveness (CE)

Discount factor

Environmental justice

Existence values

Opportunity costs

Opportunity cost view

Present discounted value (PDV)

Public expenditure review (PER)

Pure discount rate

Risk assessment

Risk premium

Shadow prices

Social discount rate

QUESTIONS AND PROBLEMS

1. Consider a project that costs $100,000 and yields a return of $30,000 for five years. At the end of the fifth year, there is a cost of $20,000 to dispose of the waste from the project. Should the project be undertaken if the discount rate is 0? 10 percent? 15 percent? The interest rate at which the net present discounted value of the project is zero is referred to as the *internal rate of return* of the project.

2. Assume there is uncertainty about the costs of disposing of the waste: there is a fifty-fifty chance that they will be $10,000 or $30,000. Discuss how this uncertainty affects the cost–benefit calculation, if the government is risk neutral (i.e., that it requires no risk premium to compensate it for bearing risk) or very risk averse (i.e., that it requires a large risk premium to compensate it for bearing risk).

3. Assume now that there are two groups in the population. Each contributes equally to the cost of the project, but two-thirds of the benefits accrue to the richer group. Discuss how this alters the cost–benefit calculation. Under what circumstances will the decision to undertake the project be altered?

4. Assume that the government now has a choice between undertaking the project described in problem 1 and undertaking a larger project. If it spends an additional $100,000, returns will be increased by $25,000 per year and disposal costs in the final year will increase by $20,000. Which project should be undertaken if the discount rate is 0? 10 percent? 15 percent? If there are two groups in the population, how are your answers affected if two-thirds of the incremental benefits go to the poor (with the incremental costs being shared equally, as before)?

5. Discuss why, under each of these circumstances, a social cost–benefit analysis might differ from a private cost–benefit analysis:

   a. The unemployment rate is 10 percent.

   b. The government has imposed a tariff on the importation of textiles.

   c. The government has imposed a quota on the importation of oil.

   d. The government has imposed a tax on interest income.

   e. The government has imposed price controls on natural gas.

   f. The government has regulated airlines, so prices exceed the competitive levels.

6. For each of the following projects, what benefits or costs might be included in a social cost–benefit analysis that might be excluded from a private cost–benefit analysis:

   a. A hydroelectric project

   b. A steel mill

   c. A chemical plant

   d. A project to improve car safety

   e. A training program to improve the skills of minority workers in a firm

   How might your answers be affected by changes in legislation (e.g., concerning manufacturers' liabilities for automobile accidents, legislation imposing fines on polluters)?

7. How might the techniques used to analyze the distributional consequences in cost–benefit analysis be employed to ensure that concerns about environmental justice are incorporated into the analysis?

8. How are issues of incidence analysis and capitalization (discussed in the previous chapter) incorporated into cost–benefit analysis? For each case below, does it make a difference to your answer whether poor individuals own or rent their houses? In particular, if you wished to incorporate the distributional consequences of the following policies and programs, how might you do so?

   a. A government regulation that reduces the allowable level of noise for aircraft. (Assume that those who live in the neighborhood of the airport are relatively poor.)

   b. A subway line intended to make it less expensive for those in low-income neighborhoods to get to jobs in the center city.

   c. The Superfund program is intended to clean up toxic waste sites. Currently, a disproportionate number of poor people live near toxic waste sites.

9. The government is debating whether to spend $100 billion to reduce global warming damage 100 years from now. It is estimated that $800 billion of damage will be averted. A critic of the expenditure says that it would be far better to take the $100 billion, invest it in the stock market, earning an average return of 6 percent per year, and use the proceeds of the investment in 100 years to repair the damage. Should the project be undertaken?

12

DEFENSE, RESEARCH, AND TECHNOLOGY

FOCUS QUESTIONS

1. How do we go about deciding how much to spend on national defense? What is the role of marginal analysis?

2. What are the current key issues concerning defense strategy and their implications for the level and allocation of defense expenditures?

3. What are some of the key ways by which the Defense Department is currently attempting to increase its efficiency?

4. What is the rationale for government actions to promote research and new technologies?

5. What are intellectual property rights? How well do they address the problem of underinvestment in research?

6. What are the other ways by which government encourages the private production of knowledge? Why does government provide direct support to research?

The end of the Cold War, marked by the breakup of the Soviet Union in 1991, had profound consequences. At the time, defense had long been the largest single item of public expenditures in the United States. In 1953, the last year of the Korean War, defense spending reached 14.7 percent of GDP. After the Vietnam War, defense expenditures fell to under 6 percent of GDP, but during the 1980s, as part of the Cold War, they increased again, peaking at 7.4 percent of GDP in 1986 and 1987.

After the collapse of the Soviet empire, there was excitement in the United States about a **"peace dividend."** Defense expenditures dropped to half of their 1980s share of GDP—by 2000 they were just 3.7 percent of GDP. These spending reductions were critical to the successful effort to reduce the federal budget deficit during the Clinton administration.

The end of the Cold War did not bring global peace, however. There were conflicts in the Balkans and sub-Saharan Africa. There were also continuing worries about nuclear proliferation. Then, with the attack on the World Trade Center on September 11, 2001, the United States became engaged in a new kind of conflict, called the global war on terrorism. Although it centered on Iraq and Afghanistan, it was a war that knew no

clear boundaries and in which there was no clear victory. By 2010, defense expenditures had risen to 5.6 percent of GDP, and at $717 billion (in 2005 dollars), they were at their highest point in constant dollars since World War II (see Figure 12.1).

In 2010, U.S. defense expenditures comprised the greatest share of GDP and the greatest share of total central government expenditures of any of the OECD countries, and were almost twice the OECD average for both these public expenditure measures. Although no low-income countries had similar spending figures in 2010, several middle income countries spent comparable amounts on national defense in terms of share of GDP and share of central government expenditures (see Table 12.1).

U.S. expenditures amounted to nearly half of all global military expenditures, raising questions about whether spending had become excessive. Expenditures were forecast to remain high even as the United States exited from the wars in Afghanistan and Iraq, although budgetary pressure limited the growth in these expenditures beginning in 2013.

Another consequence of the end of the Cold War related to the federal government's expenditures on research and development (R&D). For decades, defense spending and R&D spending had been closely intertwined. Indeed, in 2010, 57 percent of the federal government's expenditures on R&D were defense related. There had been great spin-offs to the commercial sector—from advances in computer technology to advances in ceramics—but U.S. commercial interests almost surely did not benefit from these expenditures to the same extent that they would have if more of government R&D expenditures had been directed at commercial uses. Nondefense R&D expenditures lagged behind those of the country's competitors, and in the long run this could have significant adverse effects on U.S. competitiveness. There was thus a need for a new U.S. research and technology policy.

This chapter discusses these two important areas of government expenditure.

DEFENSE EXPENDITURES

Even though defense has long been the major item of expenditure at the federal level, it has traditionally received little attention from economists. Instead, military experts have led the way in determining how to achieve the country's defense objectives. However, defense spending is fundamentally a question of resource allocation—the country wishes to receive the best defense for the money spent—and therefore hinges on economic reasoning.

As the defense budget is being formulated, the fundamental question is: How much? Conventional economic analysis—focusing on marginal
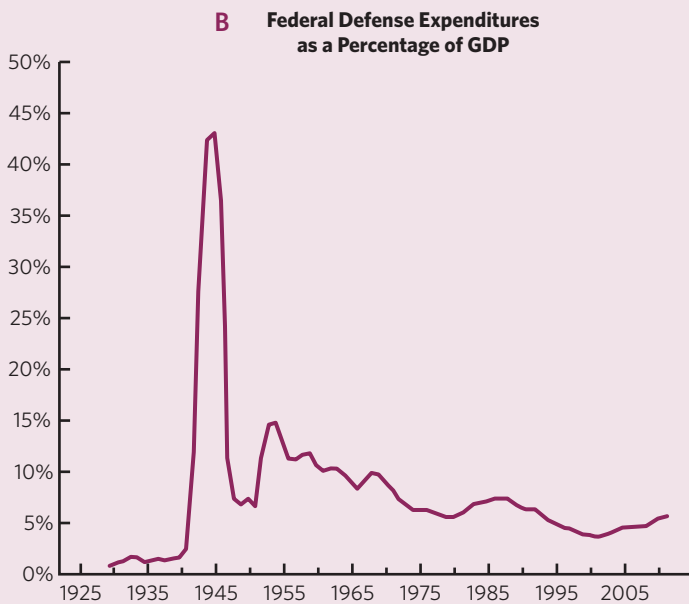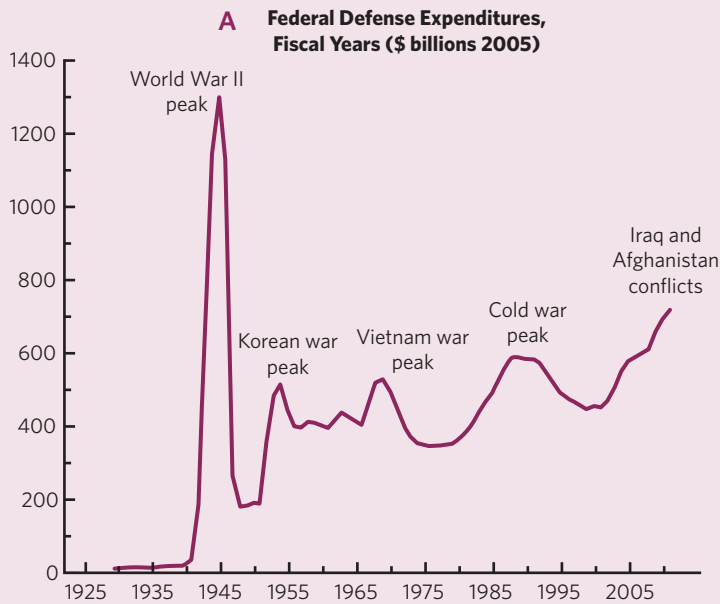
A   **Federal Defense Expenditures, Fiscal Years ($ billions 2005)**

World War II peak

Korean war peak

Vietnam war peak

Cold war peak

Iraq and Afghanistan conflicts

B   **Federal Defense Expenditures as a Percentage of GDP**

TABLE 12.1 DEFENSE EXPENDITURES IN COMPARATIVE PERSPECTIVE

COUNTRY/COUNTRY GROUP	% OF GDP (2010)	% OF EXPENDITURES* (2009)
HIGH-INCOME COUNTRIES		
Australia	1.9**	7.3
Canada	1.4	7.6
France	2.3	5.3
Germany	1.4	4.5
Greece	3.1	6.3
Japan	1.0	na
South Korea	2.7	13.4
Saudi Arabia	10.4	na
United Kingdom	2.7	5.7
United States	4.8	18.0
High Income, OECD	2.8	10.0
High Income, Non-OECD	5.5	na
MIDDLE-INCOME COUNTRIES		
Brazil	1.6	6.3
Chile	3.2	15.4
China	2.0	16.1***
Colombia	3.7	18.9
Ghana	0.4	2.4
India	2.4	16.4
Indonesia	1.0	5.5
Pakistan	3.2	19.3
Russian Federation	4.0	14.0
South Africa	1.2	4.1
Thailand	1.5	9.5
Turkey	2.4	9.7
LOW-INCOME COUNTRIES		
Bangladesh	1.2	10.2
Burkina Faso	1.5	10.4
Kenya	1.9	9.1
Middle and Low Income	2.0	na
Lower Middle Income	2.1	12.0

*Share of central government expenditure
**2009
***2008

SOURCES: World Bank, *World Development Indicators*; and Stockholm International Peace Research Institute (SIPRI), *Yearbook: Armaments, Disarmament and International Security*.

benefits and costs—is essential to answering this question, yet the most difficult questions go beyond simple economics. How do we assess the magnitude of the benefits from defense expenditures, and how should an appropriate military strategy be selected? The next two sections take up these questions. Later sections consider how to increase the efficiency of our defense expenditures and how to address the problems associated with downsizing the military.

THE VALUE OF MARGINAL ANALYSIS

In allocating a given defense budget, one needs to consider the effect of the expenditures on various defense objectives. In evaluating whether we should spend more on defense, we similarly need to know how much extra "protection" we get from an extra expenditure of $1 billion.

The following example, provided by Charles Hitch, who was assistant secretary of defense in the Kennedy–Johnson administration, illustrates the role of marginal analysis.[1] Assume that each missile has a 50 percent probability of success in killing its target. We have 100 targets that we would like to destroy. One hundred missiles sent at the targets would "achieve an expectancy of 50 kills, 200 missiles—75 kills, 300 missiles—87 kills," as depicted in Figure 12.2. Clearly, there are very strong diminishing returns.

---

[1] From C. J. Hitch, *Decision Making for Defense* (Berkeley: University of California Press, 1966), pp. 50–51.
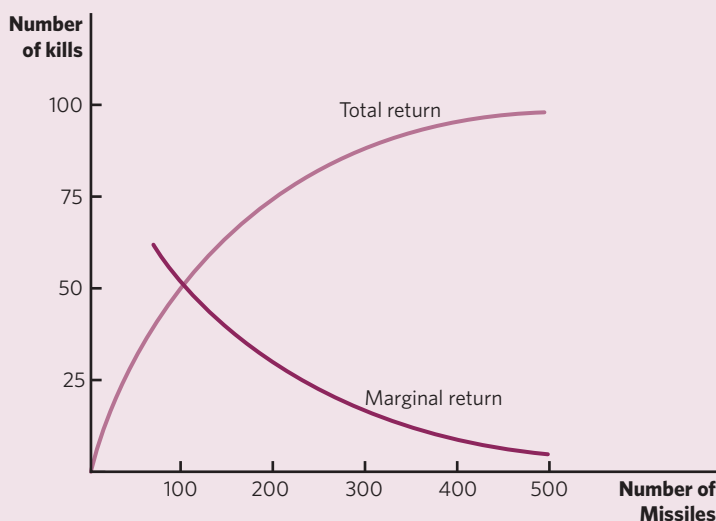
FIGURE 12.2

THE ROLE OF MARGINAL ANALYSIS IN DEFENSE
The relevant question is not whether we should have 500 missiles or no missiles, but how many extra kills we get from each additional missile. There may be sharply diminishing returns.

Each target can be destroyed only once, and some of the additional missiles will land on a target that had already been destroyed. Whereas the first 100 missiles give us 50 kills, increasing the number of missiles from 400 to 500 increases the number of kills by only 3. We need to ask ourselves not whether it is worth the cost of 500 missiles to get 97 kills but whether it is worth the cost of 100 additional missiles to get 3 extra kills.

This kind of analysis is not easy. By relating expenditures to objectives and showing what one gets from additional expenditures, though, one can hope to make more rational decisions concerning how much is enough.

In making these assessments, however, one set of considerations is particularly hard to evaluate: deterrence.

DEFENSE STRATEGY

In the 1960s and 1970s, defense focused on **deterrence**—establishing a strong enough force that no one would contemplate attacking. Under President Ronald Reagan, another approach was advocated: the **Strategic Defense Initiative (SDI)**, which attempted to provide a failsafe defense capability. Today, the focus of debate is on several other issues: the capability of fighting two wars simultaneously, nuclear proliferation, terrorism, and chemical and biological warfare.

DETERRENCE   The existence of strong capacities for retaliating serves to deter others from "misbehaving." This thinking was central to the defense strategy of the United States and the Soviet Union after World War II. It was believed that with a large enough capacity to destroy the opponent, there would be little incentive for each country to attack. The capability of retaliating depended on the number of missiles that survived the first strike, known as the **second strike capability**. Because the number that survived depended on the size of the first strike, each side had a strong incentive to create a larger arsenal; thus, the arms race was born.

However, with the end of the Cold War and the dismantling of much of Russia's nuclear capabilities, no country had the capacity to maintain a sustained attack against the United States. This led many to question the relevance of the deterrence strategy, and, if it was still needed, whether the size of the U.S. deterrence capacity could be smaller. The United States reduced its nuclear arsenal, but critics suggested that there was scope for significant further reductions.

TWO-THEATRE CAPABILITY   With the threat of massive nuclear war subsiding, attention has focused on regional wars—a conflict in the

Persian Gulf or on the Korean peninsula. There is a widespread view among military experts that we should have the capability of fighting at least two such localized wars simultaneously.

The problem is analogous to larger communities facing how large a fire department to have. Assume that it takes twenty men to fight a moderate-sized fire. Most of the time, there will be no fires. Occasionally, there will be a fire, and it is surely worthwhile to have a fire department ready for that emergency. For simplicity, we assume it takes two hours to contain a fire. Assume that the community has 10,000 buildings, and the chance of any building catching fire in any two-hour span is one in a thousand. There is a small probability that in any two-hour span more than one fire will occur. The question is: Is it worth having a second fire department—an additional twenty firefighters—on reserve for this small-probability event? It may or may not be. If it is, we then need to ask, is it worth having a third fire department on reserve for the very slight chance that there will be three or more fires? The probability of using each additional fire department—and thus the expected benefit—diminishes markedly.

In one respect, the fire analogy is inappropriate. Typically, we do not worry about an arsonist waiting to light a fire until the fire department is busy putting out a fire elsewhere. In the area of defense, however, this is a real concern: potential enemies or troublemakers may wait until the United States is occupied with fighting in one theatre to start trouble in another. Many worry that without a two-theatre capacity, the United States would be inviting trouble. From this perspective, the major role of the **two-theatre capability** is its role as a deterrent.

Currently, the dominant view—reaffirmed in the 2010 *Quadrennial Defense Review Report*[2]—is that the United States should have the capacity to fight two localized wars of moderate scale simultaneously. It is the ability to fight the second war that is compromised when the defense budget is cut.

Of course, matters are seldom clear-cut. One of the main bottlenecks in fighting a war is logistics—in particular, getting troops and equipment to the scene of battle. These bottlenecks are likely to be particularly binding in the early stages of a war, and at times when troops and equipment must be deployed rapidly. Given a span of a few months, ships being used for other purposes can be converted into transports for military equipment and personnel. This can be done even when troops need to be deployed quickly, provided there is willingness to force the domestic economy to undertake greater short-run costs: the military could, for instance, order civilian aircraft to transport troops. Thus, greater expenditures

---

[2] U.S. Department of Defense, *Quadrennial Defense Review Report*, February 2010, p. vi.

GAME THEORY, THE ARMS RACE, AND THE THEORY OF DETERRENCE

The 1994 Nobel Prize in economics was awarded to three economist–mathematicians who pioneered in the development of game theory: John Nash of Princeton, John Harsanyi of the University of California at Berkeley, and Reinhold Selten of Rheinische Friedrich-Wilhelms-Universität in Bonn, Germany.

Most of game theory is based on the postulate of rationality. For example, Bradley not only plays to win but also believes Boris, his opponent, plays to win; and believes Boris believes that he is playing to win, and so forth. Thus, each player, in deciding what to do, puts himself in his rival's shoes: If I do X, what will my rival do? But in putting himself in Boris's shoes, Bradley reasons that Boris will be thinking about what *he* would do, assuming Boris is rational.

Much of the development of game theory was supported by the Department of Defense, in an attempt to understand better how the Soviet Union would respond to what the United States did. Game theory was used both as the basis of the theory of deterrence, which underlay American defense strategy, and to explain the arms race.

Although there is some debate about the adequacy of the theory of deterrence and the arms race that game theory provided, it has subsequently found wide use throughout economics. For instance, game theory has been used extensively in the analysis of patent races, to understand when and whether patents actually spur innovation. Some models suggest, for example, that in some circumstances, there may be little impetus for research, as the lead firm gets sufficiently ahead of its rivals and then rests on its laurels. The firm's rivals know that should they accelerate their research, the firm will respond, and thus, their prospects of winning the patent race are minimal.

for defense readiness reduce the costs should a crisis occur. The question then is: How much of an "insurance premium" are we willing to pay?

The issue, then, is not (as it is sometimes put) simply whether the United States could fight on two fronts simultaneously, but rather, what would be the additional costs and risks on the second front? Almost all the serious conflagrations that the United States has faced in the past fifty years have lasted months, if not years. The Iraq and Afghanistan wars have, indeed, been the longest-lasting wars in America's history. If this is the likely pattern in the future, then the need for a full capacity for rapid deployment on a second front may not be a high imperative.

However, increased threats from nonconventional warfare—whether from forces representing failed states or nonstate entities, or from new weapons such as those of mass destruction and those operating in cyberspace—have required the Defense Department to enhance capacity elsewhere. New priorities include unmanned aircraft systems for intelligence, surveillance, and reconnaissance; rapid deployment special operations forces; and counterinsurgency and counterterrorism. Skills and

technology required for success in these arenas are quite different from those that have marked the military of the past. A current concern is that expenditures are unbalanced—too much directed at dealing with the "old" threats, too little on the new.

ARMS PROLIFERATION   After the fall of the Soviet empire, attention focused on denuclearizing Ukraine and Kazakhstan (the two former Soviet republics, besides Russia, that inherited nuclear weapons), reducing the number of nuclear warheads in Russia, and countering arms proliferation in general (see case study, "Converting Swords into Plowshares").

The arms industry is profitable, and it is one of the few industries in which Russia is globally competitive. With many small countries interested in buying weapons, there is real danger that the arms will feed the ethnic and political struggles that seem to be proliferating around the world. As the United States has itself been a major arms supplier to the rest of the world, it is difficult for this country to criticize others for doing as it does, even though it claims to be more responsible in its arms sales.[3]

The United States and seventy-nine other countries signed a nonproliferation treaty in 1968, attempting to restrict access to nuclear weapons and the methods of delivering them.[4] The United States has used the threat of economic sanctions to enforce the agreement since it passed its first nonproliferation sanctions law in 1974, targeted at India for what India called its "peaceful nuclear explosion." Nonproliferation laws impose sanctions against individuals, private entities, and governments that engage in proliferation activities.

For example, in 1993 and again in 1996, in retaliation for exporting missiles to Pakistan, China was threatened with losing its most favored nation status—under which its exports to the United States faced the lowest tariffs that the United States imposes on any country (other than Canada and Mexico). In May 1998, India and Pakistan both successfully tested nuclear bombs, joining the exclusive nuclear club. In response, the United States and other countries imposed sanctions on both countries, cutting off nonhumanitarian assistance and restricting loans to the Indian and Pakistani governments.

---

[3] From 2003 to 2010, the United States exported $170.8 billion worth of weapons (in constant 2010 dollars), which was higher than any other country and accounted for 39 percent of the world's total. The second largest arms exporter for the same period was Russia, whose $81.1 billion worth of weapons sales was 18 percent of the world's total. (Source: Richard F. Grimmett, *Conventional Arms Transfers to Developing Nations, 2003–2010*, Congressional Research Service, September 2011, Table 31.)
[4] By 2010, 190 countries were members of the nonproliferation treaty. [Source: United Nations Office for Disarmament Affairs, "Treaty on the Non-Proliferation of Nuclear Weapons (NPT)," http://www.un.org/disarmament/WMD/Nuclear/NPT.shtml.]

The alleged reason that the United States went to war in Iraq was over concerns that the country was building nuclear weapons. Even though UN inspectors had contended that this was not the case, the United States argued otherwise. After the invasion, it turned out that the UN inspectors had been correct. Reports concerning Iraq's nuclear capacities were based on flawed and fabricated evidence, throwing into question the competency of America's intelligence services and the credibility of its public assurances.

One of the most heavily invoked nonproliferation laws at present is the Iran, North Korea, and Syria Nonproliferation Act of 2006. Military confrontation between North Korea and the United States has been avoided periodically over the past two decades when standoffs between the two countries have ended with the signing of agreements committing the United States to provide aid in return for North Korea's dismantling of its nuclear potential, although these agreements have invariably broken down amid mutual recriminations of noncompliance.

Subsequently, the major concern over nuclear proliferation has been with Iran. Iran claims that the nuclear enrichment facilities it wishes to

CONVERTING SWORDS INTO PLOWSHARES

The end of the Soviet Union posed a challenge: how to make sure that the highly enriched uranium in the nuclear warheads in the missiles in Russia, Ukraine, and Kazakstan did not fall into the wrong hands. The United States proposed an innovative approach: it would buy the nuclear material, bring it to the United States, and de-enrich it so that it could be used in nuclear power plants.

There was, however, one obstacle: the United States was in the process of privatizing the government agency responsible for enriching uranium. The U.S. Enrichment Corporation (as it was called) would make most of its money from selling enriched uranium to power plants, and it did not want competition. Opponents of privatization argued that this was just one of the reasons that one should not privatize the company. The same processes that are used for making enriched uranium for nuclear power plants can further enrich it, to the level required for nuclear weapons. It seemed reckless to turn over responsibility for making enriched uranium to a private firm, whose incentives were to sell the enriched uranium to the highest bidder.

In the end, the privatization went ahead. It slowed the process of bringing the enriched uranium into the United States, but it could not stop it. On the other hand, the alleged benefits of privatization—greater efficiency—never materialized. The firm has remained perilously close to bankruptcy, with occasional proposals being discussed in Congress to renationalize it. This is one privatization that no one considers a real success.

SOURCES: The early story of the battles over privatization of USEC are told in J. Stiglitz, "Unfair Trade Laws and Other Mischief," *in Globalization and Its Discontents* (New York: W. W. Norton & Company, 2002), pp. 166–179. The saga is updated in D. Koplow, "Subsidies to Uranium Enrichment," in *Nuclear Power: Still Not Viable without Subsidies* (Cambridge, MA: Union of Concerned Scientists, February 2011), pp. 62–72.

construct are for peaceful purposes (nuclear energy), but others are not convinced, setting the stage for a number of confrontations between Iran and other countries committed to limiting nuclear proliferation.

CHEMICAL AND BIOLOGICAL WEAPONS AND TERRORISM

Nuclear nonproliferation sanctions have now been expanded to prevent the proliferation of chemical and biological weapons as well, collectively referred to as **weapons of mass destruction (WMD)**.

The 1991 Gulf War focused attention more broadly on the threat of rogue states, and their use of terrorism and chemical and biological weapons. That a small vial containing a deadly virus could cause massive death is a frightening prospect. Even if Iraq had not engaged in biological warfare, the numerous maladies plaguing those returning from the Gulf War were a constant reminder of the threat of chemical weapons (although it subsequently turned out that many of these maladies may have been related to spent uranium used in many U.S. military weapons). A bipartisan effort under both President Bush and President Clinton led to an international agreement to ban chemical weapons.

Establishing enforceable international agreements and cooperation to combat biological and chemical weapons and terrorism is one part of the U.S. strategy for dealing with these threats. Research aimed, for instance, at building the capacity to develop antidotes and antibodies quickly, is another. Given the magnitude of the danger, there is concern that the United States is spending too little on such research.

INCREASING THE EFFICIENCY OF THE DEFENSE DEPARTMENT

No matter how much is spent on defense, it should be spent well. Recent attention has been focused on improving the Defense Department's procurement policies and on reorganization.

DEFENSE PROCUREMENT

During the height of the Cold War, thousands of firms focused on supplying the research and weapons required for the expanding military. Producing for the military was different in many ways from producing for the civilian sector. Civilian aircraft had many potential buyers. The various aircraft manufacturers produced the plane that they thought best

met the aviation needs of the world, and then tried to persuade each of the airlines that their aircraft was better than their rivals'. If an aircraft manufacturer produced an advanced military plane, though, there was usually only one customer: the Department of Defense. Usually, the Department of Defense had only one supplier of that particular plane; it was simply too costly to have two or more firms producing exactly the same plane. Many of the procurement problems described in the following sections apply across all branches of government, but some are particularly acute in the Defense Department because of the unique nature of its purchases.

STANDARD PROCEDURES   To ensure that it obtains the best price, the government usually resorts to competitive bidding: different contractors tell the government the price at which they are willing to deliver, say, one thousand tanks of a given specification, and the government purchases the tanks from the lowest bidder. Frequently, however, there are major **cost overruns**—that is, the costs exceed the producer's original estimate. Sometimes the contract calls for these costs to be shared by the government and the private contractor; such contracts are called **cost-sharing contracts**. Even when the contract does not explicitly call for sharing of cost overruns, the government may absorb all or a significant fraction of the additional costs. The contractor may claim that the cost overruns are a result of changes in the design specification; such changes almost always accompany the development of a new weapon, particularly when the development occurs over a period of several years, and it is often difficult to ascertain to what extent the cost overruns are, in fact, a result of the design changes. In other cases, the private contractor simply says that it cannot complete the contract without further funds; the government then has the choice of losing all that it has already spent or negotiating a settlement with the contractor. Even if the government were to sue the contractor for breach of contract, the delays in the development and deployment of the weapons could be very costly.

The prevalence of cost overruns means that the public seldom has an accurate view of the cost of a ship, a defense system, a tank, and so on at the time a commitment is made to purchase them. It also means that the government seldom knows whether it has, in fact, let out the contract to the lowest-cost producer[5]; all that it knows is that it let out the contract to the lowest bidder.

What are the reasons for these cost overruns? In the case of new weapons, errors in estimating costs are common—but why should there

---

[5] Notice that at the time the cost overruns occur, what limited competition there was before the contract was let out no longer exists: it would, in general, be costly or impossible for the government at that point to turn to other potential suppliers.

be a bias in these errors? That is, why should there be a systematic tendency to underestimate the costs? Part of the reason has to do with the competitive bidding process: potential contractors know that they have to produce a low bid to win. The system of cost sharing (implicit or explicit) means that there is relatively little penalty associated with bidding too low. There is, however, a penalty associated with bidding too high, particularly when other firms are bidding low (using, say, their most optimistic estimates of costs): the high-bidding firm will fail to get the contract.

The system of cost sharing has a further disadvantage in addition to reducing the penalty for underbidding: the winner of the contract has little incentive to be efficient. **Cost-plus contracts** further reduce the incentive to be efficient; for example, a contract where the government pays whatever it costs to develop weapons plus, say, 10 percent, actually provides incentives to be inefficient—the more the firm spends, the more it gets from the government.

Why does the government engage in cost sharing, with all its obvious disadvantages? Part of the reason for cost sharing is the uncertainty inherent in the development of a new weapons system. The best that a firm can do, as we have said, is to provide an estimate of these costs. If there were **fixed-fee contracts** (contracts under which the contractor was paid a fixed amount, regardless of the eventual cost), the contractor would have to bear considerable risk; even if it were very efficient, there is some chance that it would encounter difficulties in the development of the system that would increase its costs way beyond the fixed fee it would receive, in which case it might incur an enormous loss. If firms (or their managers) are risk averse and insist on being compensated for bearing risks, they will all put in high bids, representing their estimate of the actual costs plus a fee for bearing the risk. The government is in a better position to bear the risk. By agreeing to a cost-sharing contract, it absorbs much of the risk, but at the same time reduces the incentives for efficiency.

Even though this provides an important rationale for cost-sharing contracts, some critics of the Pentagon argue that other forces are at work when the government agrees to pay all or part of a cost overrun. A large number of military officers, upon their retirement from the armed forces, take up positions in private industry—in particular, with defense contractors. Critics say that this provides an incentive for these officers to be accommodating to the requests of defense contractors.

Supporters of the current system, though admitting that it is far from perfect, point out that there is a healthy level of competition among defense contractors and that this competition provides at least some limit to the extent of inefficiency. Any contractor that performed consistently worse than other firms would find itself having difficulty obtaining contracts.

EXCESSIVE MONITORING AND PROCUREMENT REFORM   Many believe that the most important problem with defense contracting is not cost overruns, but the detailed procedures that are instituted to ensure that the government is not cheated. These require excessive monitoring and record keeping, and reduce the flexibility that firms need to produce in a cost-effective manner. Indeed, when the government buys exactly the same equipment that the private sector buys, such as a jet engine, as a result of government procurement procedures the government winds up paying substantially more—because it costs the firm substantially more to produce the item. In 1994, the Department of Defense reformed its procurement procedures, which had added substantially to its cost of doing business. It sought to make its procedures conform more closely to those that were standard in the private sector. Because, with military downsizing, fewer firms would be able to produce exclusively for the military, fewer firms would be willing to go through the hassle required to conform to government procurement rules.

There were several aspects to procurement reform. One was the promotion of **dual-use technologies**—that is, technologies that could be used by both civilian and military customers. This would have two advantages. For many of the items, the larger market would reduce the cost of production and enhance the degree of competition. Another aspect to procurement reform was to emphasize "off-the-shelf" products—products that are available commercially. Thus, rather than buying made-to-order T-shirts, the Defense Department might choose among T-shirts already in the market. There is no compelling reason why defense T-shirts need to be designed differently from T-shirts used elsewhere.

Although this reform improved general defense procurement efficiency, the Defense Department's weapon acquisition programs have remained on the list maintained by the Government Accountability Office (GAO; formerly the General Accounting Office) of areas in which the government is at high risk for overpaying since 1990. For example, the GAO's seventh annual assessment of selected weapon programs (FY 2008) found that research and development costs were 42 percent higher than originally estimated, total acquisition cost was $296 billion over budget (FY 2009 dollars), and the average delay in delivering initial capabilities had increased to 22 months.[6] In response, the Senate and the House of Representatives unanimously passed the Weapon Systems Acquisition Reform Act of 2009. This law creates a new Office of Cost Assessment and Program Evaluation (CAPE) at the Defense Department, reporting directly to the Secretary of Defense.

---

[6]U.S. Government Accountability Office, *Defense Acquisitions: Assessments of Selected Weapon Programs*, Report to Congressional Committees, March 2009, p. 6.

It also calls for more extensive testing of new weapons before they enter production to ensure that they perform as designed, and gives field commanders a greater voice in formulating the technical specifications for new weapons.

DEFENSE CONVERSION

The end of the Cold War brought with it both a downsizing of the military and a restructuring. This process was referred to as **defense conversion**. Any change of this magnitude—a redeployment of even a couple of percent of GDP—puts strains on the economy, and strains on the economy get translated quickly into political pressures. Although the major rationale for defense expenditures is protection of the United States, political support for the Department of Defense also arises from the fact that defense expenditures generate so many jobs. A sound macroeconomic policy should be able to maintain the economy at or near full employment, but individuals in defense-related jobs worry that defense cutbacks will cost them their jobs, and that even if they could get other jobs, they would not be as good as their current jobs.

These concerns are exemplified by the controversy over base closures. As the size of the military forces was reduced, fewer bases were required. However, closing a base means a loss of jobs. Congressional representatives see maintaining jobs in their districts as one of their key responsibilities.

The politics of base closure were so intense that in 1988, Congress created the Commission on Base Realignment and Closure (BRAC) to recommend which bases should be realigned or closed. The commission depoliticized the process (and took the heat off Congress), because although Congress could veto the commission's entire recommendation, it could not remove individual bases from the list. This process was repeated in 1991, 1993, and 1995 under the auspices of the Defense Base Realignment and Closure Act of 1990. The first four BRAC commissions recommended the closing or realignment of approximately 450 installations, eliminating 20 percent of the Defense Department's 1988 pre-BRAC capacity. This included closing 97 out of 495 major domestic installations. All recommendations

DOWNSIZING DEFENSE

- Defense conversion entails redeploying resources. There may be significant costs to redeployment.

- It is important that resources, such as bases, be redeployed efficiently.

- Even though, in the long run, society benefits from redeployment, particular individuals and communities may be adversely affected. Markets fail to provide insurance against economic dislocation, and accordingly there may be grounds for government assistance.

were accepted by Congress. These four BRAC rounds produced net savings of approximately $16.7 billion through 2001, including the cost of environmental cleanup, with recurring savings beyond 2001 estimated to be about $6.6 billion annually.[7] In 2005, a fifth BRAC commission recommended an additional 182 closures or realignments, including twenty-two major closures and thirty-three major realignments, estimated to save taxpayers $15 billion over twenty years. Again, all recommendations were accepted by Congress.[8]

Defense downsizing led to high levels of unemployment and other serious adverse economic impacts in many communities that relied heavily on defense expenditures. Some have argued that the federal government has special responsibility to aid these communities. There is a market failure rationale for such aid: the inability to obtain insurance against events. In effect, by providing transitional help in the face of large shocks, society is providing a form of insurance not available through the market. Critics of such aid argue that this is simply part of capitalism: communities are buffeted by all sorts of shifts in demand. They ask, why should one kind of shift in demand—that arising from a change in the demand by government for defense services—warrant preferential treatment?

In practice, the government has tried to soften the adverse effects by slowing down the pace of transition and giving substantial advance warning. Disengagement from the conflicts in Iraq and Afghanistan, coupled with large budget deficits and a rapidly mounting public debt, means that the United States is facing several more rounds of defense downsizing. As a first step, in 2009, the Secretary of Defense proposed canceling or significantly curtailing weapons programs with a projected cost of at least $126 billion.

ACCOUNTING AND THE DEFENSE DEPARTMENT

The Iraq and Afghanistan wars brought out a new set of problems in the conduct of our defense: the process of budgeting and appropriation. When the Bush administration went to war, it estimated that the costs would be in the range of $60 billion. The eventual tab to the government will turn out to be somewhere in excess of $2 trillion.

For a long time, it was difficult even to ascertain how much the government was spending, partly because some of the spending was not fully

[7] U.S. Army Base Realignment and Closure Division, "BRAC Frequently Asked Questions," http://www.hqda.army.mil/acsim/brac/faq.htm.
[8] Defense Base Closure and Realignment Commission, *2005 Defense Base Closure and Realignment Commission Report*, pp. iii–viii.

disclosed (in effect, it was partially hidden) and partly because the money for the war was dribbled out in a series of small "emergency" appropriations that became almost regular in nature.

Part of the problem, too, arose because of the way that the government does accounting, which is different from that used by most private sector firms. It is based on cash receipts and outlays, rather than *accrual*—that is, if a firm undertakes an obligation today for which it pays tomorrow, it must account for the obligation *today*. For instance, many firms pay a part of a worker's salary in the form of deferred compensation—that is, pay that the worker will receive some time later. However, the firm must treat those future costs as part of today's production costs. This is not so for the government, which does not have to account for those future costs until they actually occur.

In the case of the military, these costs can be enormous. For instance, close to half of the troops returning from the Afghanistan and Iraq wars have some form of disability, and many have multiple disabilities. Future government expenditures for disability pay and health care costs are conservatively estimated at around a trillion dollars.

Worse still, misleading accounting systems can give rise to distorted incentives. Most important, if true total costs are understated, the level of current spending may be beyond what citizens would choose if they knew the true costs.

RESEARCH AND TECHNOLOGY

The United States government (like governments in other countries) has long played an important role in basic research and the diffusion and development of new technologies.

For example, a key factor in the remarkable increase in agricultural productivity in the last century was establishment of land-grant colleges by the Morrill Land–Grant Acts of 1862 and 1890, followed by the creation of the Cooperative Agricultural Extension Service in 1914 by the Smith–Lever Act to disseminate knowledge generated by the land-grant colleges. The nuclear age was a direct outgrowth of government research, largely conducted in government-run laboratories. Other marvels of the twentieth century, such as the jet engine and supercomputers, were largely by-products of defense R&D expenditures. The Internet and the biotech revolution are based on government funded research as well.

The federal government continues to invest heavily in R&D—in 2010, it spent $141 billion on research and development. While the government is currently increasing expenditures on nondefense R&D, defense R&D still accounts for 57 percent of total R&D expenditures—about the same as the average for 1960–2010, but down from the peak of 89 percent in 1954 (see Figure 12.3). Moreover, although the United States accounts for one-third of global R&D spending (see Table 12.2), when public and private expenditures are combined as a percentage of GDP, the United States spends substantially less on nondefense R&D (2.3 percent) than Japan (3.3 percent), South Korea (3.2 percent), or Taiwan (2.9 percent).[9] It is perhaps not surprising that the United States has not been doing as well at innovating, as evidenced, for instance, by patent applications.

---

[9] National Science Foundation, *Science and Engineering Indicators, 2012*, Table 4-20. (The percentages noted are for 2009, except for South Korea, which is for 2008.)

TABLE 12.2 RESEARCH AND DEVELOPMENT EXPENDITURES IN COMPARATIVE PERSPECTIVE (2010)

COUNTRY	% OF GDP (PPP, $US BILLIONS)	GERD* (PPP, $US BILLIONS)	SHARE OF TOTAL GLOBAL R&D SPENDING (%)
TOP 10			
United States	2.8	415.1	33.2
China	1.5	149.3	11.9
Japan	3.4	148.3	11.8
Germany	2.8	82.9	6.6
South Korea	3.4	49.0	3.9
France	2.2	47.4	3.8
United Kingdom	1.8	39.3	3.1
India	0.8	32.5	2.6
Canada	2.0	25.9	2.1
Brazil	1.1	23.9	1.9
SELECTED OTHERS			
Russian Federation	1.0	22.9	1.8
Taiwan	2.3	18.9	1.5
Sweden	3.6	12.9	1.0
Israel	4.3	9.4	0.8
Singapore	2.5	7.4	0.6
Finland	3.9	7.2	0.6

*GERD: Gross expenditures on research and development, includes public and private, current and capital expenditures.

SOURCE: Battelle, *2012 Global R&D Funding Forecast*, December 2011.

In both Japan and South Korea, the rate of patent applications, adjusted for population, is far higher than in the United States.[10]

Considerable concern has been expressed in recent years at the slow rate of growth in the United States—productivity growth from 1973 to 1995 averaged less than 0.5 percent, and although it rebounded to 1.4 percent from 2000 to 2007, it subsequently dropped to 0.7 percent from 2008 to 2010. This is in contrast to the two decades after World War II, when it averaged slightly less than 3 percent.

The three sources of growth in productivity are increases in capital, improvements in human capital (the quality of the labor force) through education and experience, and technological change. The major source of the slowdown in productivity growth since 1973 has been the slowdown in the pace of technological change—hence the concern with the underinvestment in R&D.

[10] World Intellectual Property Organization (WIPO), *World Intellectual Property Indicators, 2011 edition*, Table P1.

MARKET FAILURES

The slowdown by itself might not be a rationale for government intervention, at least from an economist's perspective. The question is: Is there a market failure—a reason why the market, left to its own, would underinvest? The answer is yes: knowledge is, in many respects, like a public good, and as we saw in Chapter 5, the private sector will underinvest in public goods.

Recall from that discussion that there are two critical properties of public goods: the undesirability of exclusion (the zero marginal cost of providing the good to an additional individual)[11] and the impossibility of exclusion. Research and development (or more accurately, knowledge, the product of research) has the first property, and often has the second as well. Giving information to additional individuals does not detract from the total amount of knowledge available.[12] However, if knowledge were provided freely, it would not pay for anyone to produce it. Thus, either the government must provide for the production of knowledge, through direct support of R&D, or it must ensure that individuals or firms that produce knowledge are compensated for doing so.

There are many ways that innovators receive rewards for their investments in research. With the "first mover advantage," the first firm in the market with a new product can often get loyal customers for itself, and is in a better position to make further advances. In many sectors, much of the knowledge associated with innovation is "tacit," hard to convey from one firm to another. In sectors like metallurgy, little reliance is placed on intellectual property. In fact, to get a patent, one must disclose a considerable amount of information, and many firms would rather maintain their advantage through trade secrets. For instance, the formula for Coca-Cola is not patented: its discoverers chose instead to keep the formula in a bank vault.

INTELLECTUAL PROPERTY RIGHTS   In some sectors, however, to enable individuals to reap rewards from their knowledge-creating activities, others must be legally excluded from using it, or at least from using it without compensating the creator. The government does this by creating "property rights" in knowledge; that is, it grants a **patent**, which gives the discoverer of the knowledge exclusive use of the knowledge, including the right to license others to use it, for a limited period of time.

---

[11] The observation that knowledge was like a candle, which, even as it lights another candle, does not have its own light diminished, was made forcefully by Thomas Jefferson.

[12] This fact should not be confused with the fact that the *return* that an individual can obtain from a particular piece of knowledge may depend on how many other people know that piece of information. A monopolist of a piece of information may be able to obtain a return that would not be possible if the information were made freely available.

A **copyright** gives the author of a work the exclusive right to use and market his or her own writings. Patents and copyrights establish **intellectual property rights**, which ensure that inventors and authors can appropriate some of the fruits of their labor. The framers of the U.S. Constitution recognized the importance of intellectual property rights when they gave the newly established federal government the right to grant copyrights and patents to encourage creative activity.

Not all ideas and discoveries are patentable—mathematical theorems are not, whereas algorithms may be—and even when a particular discovery is patentable, it is often possible to invent around the patent. Thus, whereas it is possible to patent a drug, it is often easy to devise a slight modification of the same drug with the same medicinal properties.

In determining the life of a patent, the government faces a trade-off. By extending the life of the patent, it provides greater incentives for private firms to engage in R&D; on the other hand, the knowledge produced will not be used efficiently for a longer period of time. Assume, for instance, that a firm has discovered a new, less expensive way of making a product—so much less expensive that the firm can undercut all its rivals. By patenting the discovery, the firm will be in a monopoly position; therefore, less of the product will be produced than if the knowledge were freely disseminated.

The loss resulting from the patent can be seen in Figure 12.4, in which we have drawn the market demand curve for a medicine. The cost of production
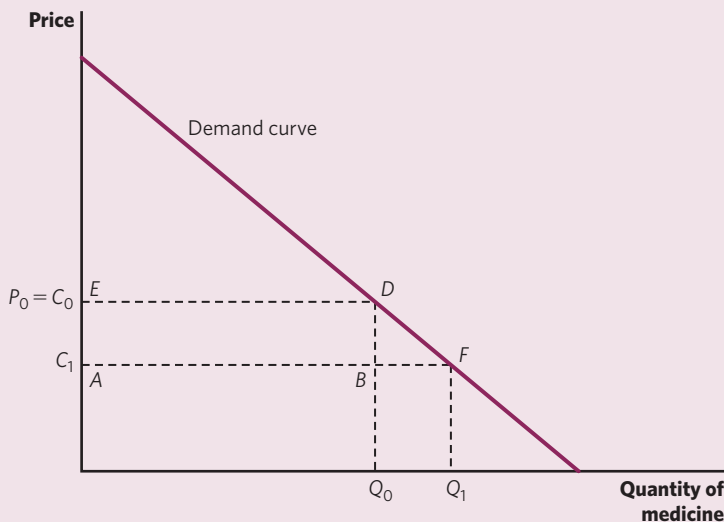
FIGURE 12.4

EFFECTS OF PATENTS
A patent system results in a lower output than with free dissemination of knowledge. The deadweight loss is given by the triangle *BDF*.

prior to the invention was C_0, and the competitive equilibrium was thus (P_0, Q_0) at point D. The firm that makes the small invention has a lower cost of production, C_1. It charges a price just below P_0, getting the entire market; its profits are thus $ABDE$, and its sales are Q_0.[13] If the information about the new innovation were made freely available, the price would fall to C_1 and the quantity would rise to Q_1. Giving the firm with the patent a monopoly on its knowledge has resulted in output being smaller than it otherwise would be, and a deadweight loss of BDF. When the patent expires, the price will fall to C_1, but the return to the innovator will drop to zero. Thus, the longer the life of the patent, the greater the deadweight loss associated with the inefficiency of giving the firm a monopoly over the use of the information—but the greater the return to the innovators and hence the greater the incentives for innovation.

[13] The firm is not, of course, free to offer any price and quantity, because at any price on or above C_0, the firm loses its monopoly position. In other words, the firm maximizes its profits when subject to the constraint that $P_1 \leq C_0$.

THE SCOPE OF THE PATENT: CAN THE HUMAN BODY BE PATENTED?

Intellectual property rights enable inventors to appropriate some part of the fruits of their innovative activity. However, there are important issues in defining intellectual property. For instance, the traditional life of a patent is seventeen years. The longer the life, the greater the returns to the innovator, but the longer the period over which the innovator exercises monopoly power, the longer the period during which the production of the innovation is restricted.

The *life* of the patent is not the only issue in defining intellectual property. An equally important issue is *scope*: How broad should the patent be? Consider recent advances in genetic engineering. Should the first person who develops a genetically engineered tomato be granted a patent for (a) all genetically engineered plants and animals, (b) all genetically engineered plants, (c) all genetically engineered tomatoes, or (d) this particular variety of tomato?

Controversy over the scope of a patent has recently been brought home forcefully by research on human genes. Decoding the millions of genes that make a human being is an immensely important task; it should, for instance, enable the development of drugs that can address innumerable diseases and health conditions. For these purposes, though, it is not necessary to decode the entire gene; most of the relevant "information" is contained in only a fraction of the gene. Some firms claim that if they can decode that information, they should be granted a patent on the gene. Others find the entire idea of patenting a human gene repugnant.

Most of the important decisions on intellectual property are made in the courts, on a case-by-case basis. Prior to the *Diamond v. Chakrabarty* case, 447 U.S. 303 (1980), life forms were considered a part of nature and were not patentable. Since then, genetically engineered bacteria have been deemed patentable because they do not occur naturally.

The Patent Office has also issued patents on whole genes whose function is known, and has developed guidelines on dealing with DNA sequences of less than a whole gene ("fragments"), as they do not have a known function.

In 2009, the American Civil Liberties Union (ACLU), on behalf of medical associations, doctors, and patients, sued Myriad Genetics, an American molecular diagnostic company, before the U.S. District Court for the Southern District of New York, to declare the patents on two genes related to breast cancer invalid (*Association For Molecular Pathology et al. v. United States Patent and Trademark Office*). Women who have the two genes have more than 50 percent chance of getting cancer; knowledge that they have the gene is important to maintain close monitoring to prevent the development of the disease. In this case, the social value of the patent was especially questionable, as there was a successful global project to decode the entire human genome; in the absence of Myriad's research, it would have been discovered shortly after. The cost of the patent, however, has been enormous, as the patent holder has insisted on remaining the monopolist for the supply of tests—even when better tests were available at lower costs elsewhere. Patients who were given false assurances that they did not have the genes because they had the less accurate tests have subsequently faced deaths that might have been avoided or delayed. In 2013, the Supreme Court upheld that naturally occurring genes cannot be patented.

Knowledge is one of the essential inputs in the production of further knowledge. Broadening the scope of patents can increase the cost of the "knowledge" input into follow-on research, and thus actually reduce the overall pace of innovation. The academic community, in particular, has been concerned that excessively broad patents could stifle their research.*

---

*The private sector has argued (following on the ideas of Coase, discussed in Chapter 6) that property rights can never be excessively broad. It would always pay a firm to grant a license to a follow-on inventor who might make use of its innovation. As we noted there, however, Coase's analysis assumed no transactions costs, perfect information, and competitive markets. In practice, these assumptions are not satisfied, and excessively broad patents can stifle innovation.

SOURCE: "Genetics and Patenting," *Human Genome Project Information*, http://www.ornl.gov/sci/techresources/Human_Genome/elsi/patents.shtml.

Intellectual property rights *may* enhance the private provision of knowledge, but they typically interfere with its efficient utilization. Giving a firm a monopoly on the production of a new product will result in too small a level of production—this is true even if there is a rationale in providing the monopoly. There is a trade-off between "dynamic efficiency," which provides firms with an incentive to innovate, and "static efficiency," which requires that firms produce up to the point at which price equals marginal cost.

More recently, there has been a concern that intellectual property rights, especially if they are not well designed, may actually impede innovation. The most important input into any research is knowledge; by making access to prior knowledge more difficult or expensive, follow-up research is discouraged. Moreover, in certain industries, innovators have been faced with a patent thicket, a complex web of claims on intellectual property; the innovator faces the threat of a suit from any one of these who might claim that his or her intellectual property rights have been infringed.

Complex products require putting together a multitude of ideas, and America's intellectual property regime puts impediments to doing this. Even worse, some firms specialize in getting patents that they can use to "hold up" others, demanding large compensation for the use of an even minor patent, while knowing that the cost and delay in inventing around the patent may be very expensive. BlackBerry was subject to such a holdup, and had to pay more than $600 million, lest it be shut down. A small company threatened similarly to shut down the imports of most of Intel's advanced microprocessors—and the Apple and HP computers that used them. The Supreme Court has recently demonstrated concern that America's intellectual property rights regime, including the way that it has been enforced, is unbalanced and may be impeding innovation.

As we have noted, many discoveries are not protected by intellectual property rights, or are protected only imperfectly. The key lesson is that intellectual property rights, as important as they are, only partially solve the market failures associated with the production of knowledge. Other government interventions are required.

OTHER MECHANISMS FOR ENCOURAGING PRIVATE PRODUCTION OF KNOWLEDGE   Governments also encourage the private production of knowledge by subsidizing a critical input into R&D, scientific manpower, through its support of education programs (see Chapter 14), and through a tax credit, the **incremental research and experimentation tax credit (R&E credit)**.[14] Firms that increase their research expenditure are allowed to deduct 20 percent of their incremental expenditures from their tax; the government pays, in effect, 20 percent of the marginal costs. The reason the credit is on incremental expenditures is to lower the *marginal* cost of research—the government does not want to subsidize research that would otherwise have been undertaken anyway—and thus to maximize the "bang for the buck." The advantage of a tax credit over a direct subsidy is that the government does not have to choose among applicants: those who believe that their projects are worth undertaking put up their own money, matched by the government (on a one-to-four basis) at the margin. The disadvantage of a tax credit is that it does not distinguish research projects which have large externalities from those which have none. Indeed, there is even some question about whether some of what gets supported is real research—overhead and marketing research may be passed off as R&D expenditures.

---

[14] The credit has not been a permanent feature of the tax code. It was originally introduced as part of the Economic Recovery Act of 1981, and has been renewed almost continuously over the past three decades. There is concern that the incentive effects have been attenuated because companies cannot rely on the credit—it was temporarily extended for the fourteenth time in December 2010 (after it expired on December 31, 2009), which also marked the ninth retroactive extension in a row.

GOVERNMENT DIRECT SUPPORT

Government supports research not only indirectly, such as through the patent system and tax credits, but also directly. Such direct support is particularly important for *basic research*. Basic research adds to our underlying store of fundamental knowledge (as opposed to *applied research*, which is intended to result more immediately in a new product or manufacturing technique). The results of basic research are more likely to be inappropriable; even if they are appropriable, the social cost of restricting the utilization of this knowledge may be particularly high, as basic research is such an important input into the production of other ideas. As a result, it is generally agreed that some form of direct support for basic research is required if there is to be an efficient allocation of resources to it.

Far more controversial, however, is government support to applied research. Sometimes, governments pick out particular industries to encourage through supporting research in those areas. Such funding or other policies, such as protection from foreign imports, aimed at promoting particular industries are referred to as **industrial policies**. Advocates of these policies argue that there is no clear line between basic and applied research; that much of applied research has huge spillovers—externalities to others that are not appropriated by the innovator—and that, hence, there will be underinvestment. The transistor was invented in Bell Labs, for example, and was of immediate benefit to AT&T, which paid for the research, but the benefits to the world clearly extended far beyond.

Critics have argued that the government has a terrible record of picking winners, because of its lack of "profit motive," and that the government should not be in the business of directing the economy. Michael Boskin, chairman of the Council of Economic Advisers under the first President Bush, was quoted as saying, "It makes no difference whether the economy produces potato chips or computer chips."

Advocates of government support for technology reply that the government actually has a remarkably good record of picking winners.[15] More importantly, they argue that the objective of government support is not to push the economy in particular directions, but simply to identify areas

---

[15] Government attempts at coal gasification and other synthetic fuel projects, as well as renewable energy projects, are usually cited as examples of these failures. Interest in industrial policies grew in the 1980s when Japan's economic success was often contrasted with the problems facing the U.S. economy. Japan had long used industrial policies, for instance, to encourage the development of the computer chip industry. The U.S. government provided support for the computer chip industry: it helped establish a consortium of U.S. producers, called Sematech, which it then subsidized. The industry has enjoyed a resurgence in the United States to the point at which the country once again dominates the world market. Government subsidies to Sematech have now ceased.

KNOWLEDGE, PUBLIC GOODS, AND EXTERNALITIES

Knowledge is, to a large extent, a public good. Through patent protection (or other forms of intellectual property rights), inventors can appropriate some of the returns to their inventive activity. Still, there are likely to be externalities. Moreover, the appropriation interferes with the efficient diffusion and utilization of knowledge. This provides the rationale for government support. It is particularly cogent for basic research.

in which there appear to be large spillovers, so that without government support, there will be underinvestment. Several examples could be cited. For instance, the telecommunications industry owes much to the federal government, from the first telegraph line between Baltimore and Washington in 1842 to the development of the Internet. In the nineteenth and twentieth centuries, government support for agriculture research and its dissemination was key to the enormous increase in productivity in that sector—which used to be the core of the economy—whereas in the twentieth century, government was central in high-technology industries, from computers to biomedicine.

In its recent technology programs, the government has tried to learn from its past mistakes. To improve incentives, it has focused on partnerships, in which the private sector puts up at least 50 percent of the costs; it has provided grants on a competitive basis, with evaluations by outside experts, and with projects in a wide variety of areas competing against each other; and it has instituted a system of ongoing review to terminate unsuccessful projects. (In the past, the reluctance to terminate such projects has been a major problem; congressional representatives from districts in which projects are located typically argued that success was just around the corner.) Ironically, although these reforms seem to have worked, they have also undermined support for technology research programs: special interests no longer see them as cash cows and therefore devote few resources to lobbying for them.

REVIEW AND PRACTICE

SUMMARY

1. In determining "how much is enough" in defense expenditures, marginal analysis—looking at the incremental benefit from increased expenditures—is essential.

2. Alternative views of overall defense strategy are important determinants of defense needs. During the Cold War, attention was focused on deterrence; President Reagan initiated the Strategic Defense Initiative; more recently, debate has focused on the desirability of having

a two-theatre capability. There is increasing concern over nuclear proliferation and cyberspace threats, and over the use by rogue states and non-state entities of chemical and biological warfare and terrorism.

3. The Defense Department has attempted to increase its efficiency by reforming its procurement policy, with greater reliance on performance standards and off-the-shelf purchases and on the development of dual-use technologies. Traditional procurement systems, including cost-plus contracts, may have contributed significantly to high costs by providing attenuated or perverse incentives. Despite procurement reforms, most major weapon acquisition programs are still well over budget and behind schedule.

4. Base closures and other reductions in defense spending will cause economic hardship in some areas. There is disagreement on what and how much the government should do to help affected communities. The absence of insurance markets insuring against these economic risks provides a rationale for government action.

5. The government has long had a role in the support of research and the development of technology. Knowledge has both of the properties of a public good, so without some government intervention, there will be underinvestment in research.

6. The government encourages innovation by establishing intellectual property rights (through patents). Although patents thus allow innovators to appropriate some of the returns to their innovation, and thus provide incentives for the *production* of knowledge, they interfere with its efficient *use*. But the patent system may actually impede innovation.

7. The government also provides direct support for research, especially basic research, and indirect support through tax credits and the support of education—producing the scientists who are the critical input into research. Government programs in support of technology are aimed at identifying areas in which there are large spillovers. Although such industrial policies have been controversial, the government has a credible record

of picking winners. Recent government policies emphasize public–private partnerships, with the private sector providing a significant fraction of the costs.

KEY CONCEPTS

Copyright
Cost overruns
Cost-plus contracts
Cost-sharing contracts
Defense conversion
Deterrence
Dual-use technologies
Fixed-fee contracts
Incremental research and experimentation tax credit (R&E credit)
Industrial policies
Intellectual property rights
Patent
Peace dividend
Second strike capability
Strategic Defense Initiative (SDI)
Two-theatre capability
Weapons of mass destruction (WMD)

QUESTIONS AND PROBLEMS

1. Assume the government has decided to install a missile system designed to provide a second strike capability with 100 missiles. It is now considering whether to increase the number to 110. Assume you are on the congressional committee that must approve the increased expenditure. List some of the questions you might ask to ascertain whether the increased expenditures are desirable.

2. Should military officers and Defense Department officials be proscribed from working for private defense contractors for a period of several years after termination of their government service?

3. In what ways is the purchase of a hammer or of labor services by the military different from the purchase of an MX missile system? How does this affect government procurement policies in these two areas?

4. Consider the following proposed system of bidding: the contract would be awarded to the lowest bidder at the price bid by the second-lowest bid. Why does such a system encourage bidders to bid their true costs of production honestly?

5. In some cases, the government runs duplicative projects, particularly at early stages of development. What do you think are the advantages of doing this? The disadvantages?

6. Consider the following proposal for reducing cost overruns. Two contractors would be given a contract to produce a tank of a given specification. Producer A would be reimbursed for the *actual* costs incurred by producer B, and vice versa. Explain why this system might induce each firm to produce efficiently. If the two firms were essentially identical, what risk premium would they require in bidding on the contract? Under what conditions might the risk premium be large? What are other possible pitfalls in this system?

7. Imagine that Congress is considering a bill to reduce the current seventeen-year life of patents to eight years. What negative effects might this change have on the rate of innovation? What positive effect might it have for the economy?

8. Suppose that many years ago, one inventor received a patent for orange juice, and then another inventor came forward and requested a patent for lemonade. The first inventor maintained that the orange juice patent should be interpreted to cover all fruit juices, whereas the second inventor argued that the original patent included only one particular method of producing one kind of juice. What trade-offs does society face in setting rules for deciding cases such as these?

9. Why might a company invest in R&D even if it does not believe it will be able to patent its discovery?

10. Some public goods are "local"—that is, they provide services only to those living in a particular locality. What might be meant by a "global" public good? Why might knowledge be thought of as a global public good? What implications does this have for government policy in a small country? Can a country be a "free rider" in basic research?

11. In what sense is international security, including efforts to stymie nuclear proliferation, a global public good?

13

HEALTH CARE

Three separate, and somewhat conflicting, concerns about the U.S. health care system continue to inform the debate over health care: excessive costs, limited insurance coverage, and the fiscal strains that providing health care imposes on government.

Health expenditures as a percentage of GDP are higher in the United States than in any other country, amounting to $2.6 trillion in 2010—about 18 percent of GDP, or $8400 per capita. Even so, health status, as recorded by such measures as life expectancy and infant mortality, is actually lower here than in many countries that spend considerably less. Furthermore, expenditures are rising rapidly, partly due to increased quantity of services, and partly due to prices increasing faster than the price level in general, as indicated in Figure 13.1.[1]

---

[1] Changes in the quality of health care services pose serious measurement problems. Most economists agree that conventional price measures overstate the rate of inflation in health care costs, because they do not adequately reflect improvements in what is being provided, although there is little consensus over the magnitude of the measurement errors.

SOARING HEALTH CARE COSTS AND EXPENDITURES, 1960–2010

(A) Health care expenditures have been rising rapidly— faster than GDP. (B) This is partly because the volume of services has increased, and partly because health care prices (measured by the medical consumer price index [medical CPI]) have been rising faster than prices in general (measured by the CPI).
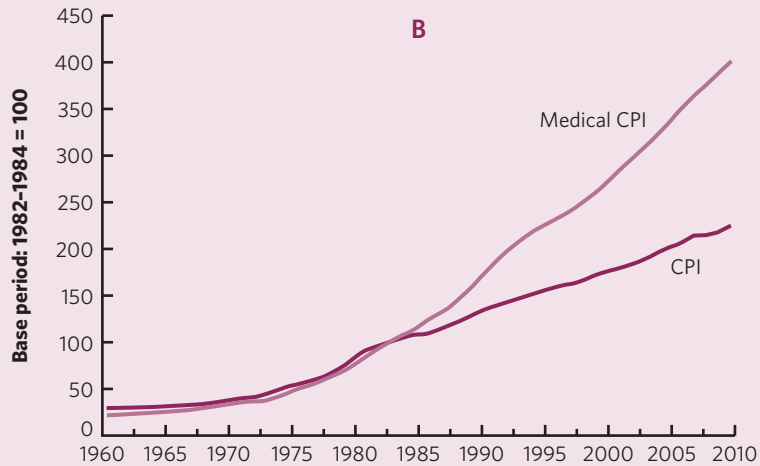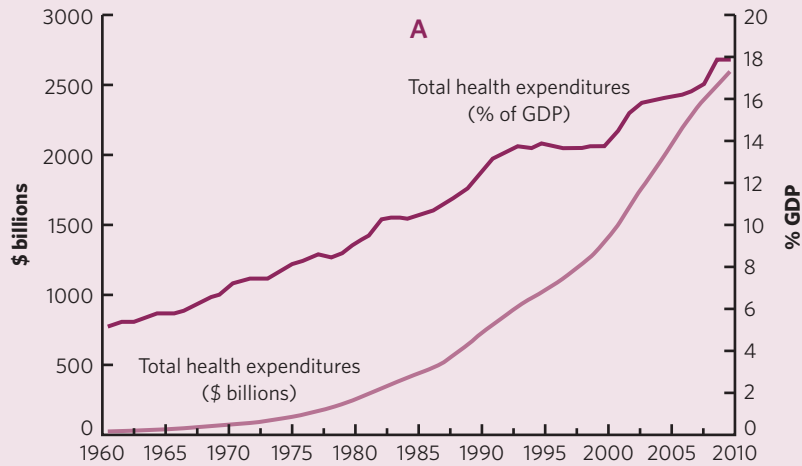
SOURCES: Centers for Medicare and Medicaid Services, Office of the Actuary, National Health Statistics Group, *National Health Expenditure Accounts, 1960–2010*; and Bureau of Labor Statistics, Department of Labor, http://www.bls.gov/data.

A second problem is that, at least before the Patient Protection and Affordable Care Act (PPACA) of 2010, many individuals lacked health insurance. Health insurance coverage declined in the United States in the years before the pasage of the PPACA. Slightly more than one of eight people were uninsured in 1987, and that number had risen to one in six before passage of the new health care law. Lack of coverage has become a major political issue because of the anxiety that it imposes, especially on middle-class workers. Because most health care insurance is provided by employers, workers who lose their jobs lose their insurance. The poor are usually covered by public programs, such as Medicaid. It is typically workers in low-paying jobs, workers without regular employment, and the self-employed who are left

uncovered. Although those without coverage can get access to emergency health care services—under the Emergency Medical Treatment and Labor Act (EMTALA) of 1986, hospitals must provide emergency services regardless of ability to pay—these individuals often delay getting adequate treatment until their condition qualifies as an emergency, thereby raising the total costs of the treatment. Because those without coverage do eventually get treatment, for which they may not pay, lack of coverage also results in **cost shifting**, a shift of the costs of unpaid bills to others.

Soaring health costs put a strain on government budgets, pushing up government expenditures for **Medicare** and **Medicaid**. As the federal budget deficit has soared, these rising government expenditures for health care, illustrated in Figure 13.1A, and for the two principal government programs, Medicare and Medicaid, in Figure 13.2, have become an increasing source of concern. Federal government expenditures on Medicare and Medicaid, which in 2010 accounted for 5.5 percent of GDP, have been projected to almost triple as a share of the economy over the next sixty years, growing to 14.9 percent of GDP by 2070. This growth is due to an increase in the elderly population, greater quantities of services being used by the aged, and health care prices that are anticipated to rise faster than the rate of inflation. Combined with Social Security, total federal expenditures on these programs have been projected to grow to 23.6 percent of GDP by 2085.[2] (To put this into perspective, *total* federal

---

[2] See Congressional Budget Office, *Historical Budget Data—January 2012 Baseline*, Publication 42911, and *CBO's 2011 Long-Term Budget Outlook*, Publication 41486. The health expenditure figures after 2010 include Medicare, Medicaid, Children's Health Insurance Program (CHIP), and exchange subsidies.

expenditures have historically amounted to about 20 percent of GDP.) If such expenditures do grow unchecked, then either taxes will have to be raised, the deficit will soar, or other government programs will have to be cut back. However, from 2010 to 2012, health spending grew at an annual rate of just 1.1 percent in real per capita terms, which is the lowest rate in the fifty years such data have been collected and much lower than the 6 percent growth rate from 2000 to 2010. Much of the slowdown in health care expenditure growth is attributed to PPACA, signed into law in March 2010, and this encouraging trend has dramatic ramifications for the nation's long run fiscal soundness.

This chapter provides an overview of the U.S. health care system. It examines how different market failures have shaped the role of government in the system; tracks the sources of the problems of excessive costs, limited insurance coverage, and fiscal strains; and discusses the different efforts to reform the system in recent years.

THE HEALTH CARE SYSTEM IN THE UNITED STATES

Governments may be involved in the health care system in a variety of ways: directly paying for health care, subsidizing individual purchases of health care and health insurance, providing health care services, financing and conducting research, preventing the spread of communicable diseases, and regulating drugs and medical devices. In the United States, the government is involved in each of these areas, but to a lesser extent than in many other countries. For instance, in Great Britain, the major health care delivery system is run by the government.

The U.S. health care system, with its mix of public, private, and nonprofit providers, is also one of the most complex systems in the world. Most health care consumers in the United States fall into one of four groups: the poor, who receive medical care through the federal government's Medicaid program; the aged, whose basic medical costs are paid by Medicare, another program of the federal government; employed individuals covered by employer-provided health insurance; and the uninsured, who sometimes purchase health care services directly, but often receive **uncompensated care**, that is, the providers of the services are not compensated and the cost is passed on to others. In addition, a small number of those who do not enjoy employer-provided insurance purchase their own insurance.

For virtually everyone, a substantial part of health care is paid not by the individual receiving the treatment, but by a third party—the government

1960

Other
(1.6%)

Government
(21.4%)

Private
health
(21.1%)

Out-of-pocket
(55.9%)

1980

Other
(4.3%)

Out-of-
pocket
(26.9%)

Government
(40.5%)

Private
health
(28.3%)

2000

Other
(5.3%)

Out-of-
pocket
(17.3%)

Government
(42.5%)

Private
health
(34.9%)

2010

Other
(4.3%)

Out-of-pocket
(13.7%)

Government
(47.9%)

Private
health
(34.1%)
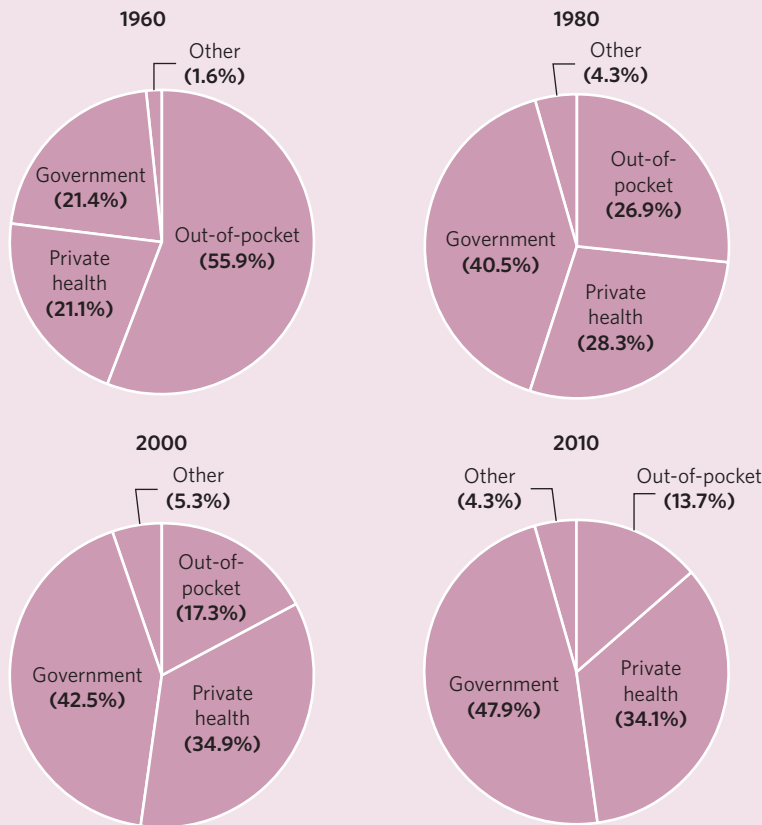
FIGURE 13.3

**SOURCES OF FUNDS
FOR PERSONAL
HEALTH CARE
EXPENDITURES,
SELECTED YEARS**

Americans pay a relatively small
percentage of health care costs
out of pocket, and the share of
such payments has been falling
in recent decades. "Other"
includes spending by charitable
organizations and industrial
on-site health services.

SOURCE: Centers for Medicare and
Medicaid Services, Office of the
Actuary, National Health Statistics
Group, *National Health Expenditure
Accounts, 1960–2010.*

or a health insurance provider. As Figure 13.3 shows, consumers prepay
(directly or through their employers) 34 percent of the cost of personal
health care in the form of private insurance premiums. Another 48 percent
of health care costs are paid by the government. Only 14 percent of the
money spent on medical care comes from direct payments by consumers.[3]
The percentage of total expenditures which individuals bear themselves has
declined steadily over the past half century. Total **third-party payments**

[3] These figures on the overall sources of funds disguise the fact that the share of out-of-pocket costs
varies a great deal among consumers (depending on their insurance plan) and among kinds of health
expenditures. For example, in 2010, only 3.2 percent of hospital costs were paid directly by consumers,
compared to 9.6 percent of physician and clinical expenditures and 28.3 percent of nursing home and
continuing care retirement community costs.
    What matters, of course, is not just the average amounts paid by consumers, but also their mar-
ginal costs. At the margin, what fraction of incremental costs do consumers bear? On average, the share
of *marginal* costs borne by individuals is likely to be less than these numbers indicate. See Centers
for Medicare and Medicaid Services, Office of the Actuary, National Health Statistics Group, *National
Health Expenditure Accounts, 1960–2010.*

(by government and health insurance) in 1960 amounted to 43 percent of total personal health care expenditures; by 2010, they were 82 percent.[4]

As indicated in Table 13.1, U.S. health expenditures are a third higher than the OECD average as a share of GDP, and are the highest among OECD countries as a share of total government expenditures; in contrast, out-of-pocket expenditures as a percentage of total health expenditures in the United States is among the lowest of OECD countries. As a share of GDP and share of total government expenditures, middle- and low-income countries spend, on average, half and one-third of the amount spent by OECD countries on health, respectively; out-of-pocket expenditures in lower–middle-income and low-income countries account for half of total health expenditures.

Not only are U.S. health care expenditures larger (both as a share of GDP and per capita) than in any other advanced industrial country, but the numbers have been increasing as well. The complaint, though, is not so much with how much is spent, but with how poor the outcomes have been—that is, with the effectiveness of the health care system. With the aging of the population and advances in technology, it might be natural for the United States to increase its expenditures on medicine, but the concern is that outcomes—measured, for instance, by life expectancy or infant mortality rates—are far poorer than those in other advanced countries that spend but a fraction of the amount that the United States spends. In some respects, U.S. outcomes are comparable to those in poor developing countries, particularly when disaggregated by location (high-income versus low-income communities) and demographics (age, race, ethnicity).[5]

Some of this is attributable to lifestyle decisions—contributing, for instance, to the high incidence of obesity or the existence in parts of the country of "food deserts," where it is difficult to get access to fresh fruits and vegetables. Some of this is attributable to the fact that the United States has a higher level of inequality than other countries, and has not provided easy access to health care for those who cannot afford it.

There are two implications of these observations: (a) There is ample room for improvement in both the efficiency and effectiveness of America's health care sector. (b) Given the concern about health, it makes sense to focus not just on the delivery of medical services, but also on other aspects of our economy and society that affect health, such as smoking and drinking, general eating habits, and exercise regimes.

---

[4] Centers for Medicare and Medicaid Services, Office of the Actuary, National Health Statistics Group, *National Health Expenditure Accounts, 1960–2010.*
[5] National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, *Health, United States, 2011 with Special Feature on Socioeconomic Status and Health*, May 2011; and *Summary Health Statistics for the U.S. Population: National Health Interview Survey, 2011*, December 2012.

TABLE 13.1 HEALTH EXPENDITURES IN COMPARATIVE PERSPECTIVE (2009)

COUNTRY/COUNTRY GROUP	PUBLIC AND PRIVATE HEALTH EXPENDITURES (% OF GDP)	PUBLIC HEALTH EXPENDITURES (% OF GOVERNMENT EXPENDITURES*)	OUT-OF-POCKET EXPENDITURES (% OF TOTAL HEALTH EXPENDITURES)
HIGH-INCOME COUNTRIES			
Australia	8.5	17.1	18.6
Bahrain	4.5	10.9	18.1
Canada	10.9	17.0	15.5
France	11.7	16.0	7.2
Germany	11.3	18.0	11.4
Greece	10.6	13.0	35.3
Japan	8.3	17.9	14.9
South Korea	6.5	12.3	34.8
United Kingdom	9.3	15.1	10.4
United States	16.2	18.7	12.4
High Income, OECD	12.1	16.6	13.7
MIDDLE-INCOME COUNTRIES			
Brazil	9.0	6.1	31.0
Chile	8.2	15.6	34.4
China	4.6	10.3	41.2
Columbia	6.4	17.9	7.9
Ghana	6.9	9.2	43.2
India	4.2	4.1	50.0
Indonesia	2.4	6.9	35.2
Pakistan	2.6	3.6	56.8
Russian Federation	5.4	8.5	28.8
South Africa	8.5	9.3	17.7
Thailand	4.3	14.0	16.5
Turkey	6.7	12.8	16.0
Middle Income	5.6	na	36.5
Lower–Middle Income	4.3	5.8	49.6
LOW-INCOME COUNTRIES			
Bangladesh	3.4	7.5	65.9
Burkina Faso	6.4	16.3	35.6
Kenya	4.3	5.4	51.2
Low Income	5.1	na	50.1

*Share of central and local government expenditures

SOURCES: World Bank, *World Development Indicators*; and World Health Organization, *National Health Account Database*.

THE PRIVATE SECTOR

Private insurance providers play a major role in the U.S. health care system. There are two dominant forms of private insurance.

In one, called **fee-for-service plans**, doctors are reimbursed on the basis of the service provided—this is, the traditional insurance plan. Typically, insurance reimburses a fixed amount of the total bill, up to some limit. Often, the insurance company pays only an amount in excess of a certain level, called the **deductible**; the fraction of the excess over this deductible that the individual pays is called the **co-payment**. Insurance policies also have caps on what they will pay for certain services.

In the other form of private insurance, called **managed care**, a third-party payer (such as an insurance company, a corporation, or the government) mediates between physicians and patients, specifying the treatments that they will cover and negotiating fees for these medical services. The most common managed care institution is called **health maintenance organization (HMO)**, which oversee a system under which the patient pays a fixed annual amount, called a **capitation fee**, to cover medical costs. Members of an HMO must go to the doctors in that HMO. In addition, to see a specialist, they must be referred by their HMO family doctor.[6] HMOs have grown rapidly in the past few years, especially among those who are covered by employer-provided health plans, mainly because they have helped contain cost increases through a combination of expenditure efficiency and an emphasis on the prevention and early detection of disease.

THE ROLE OF GOVERNMENT

We have already noted the large role of government in the U.S. health care system, and the huge growth over the past two decades of the two most important government programs, Medicare and Medicaid. Both programs were motivated largely by a concern that two major segments of the population, the poor and the aged, did not have health insurance coverage (and, in many cases, could not have obtained it, or at least not at premiums that could generally be regarded as affordable). The larger of the two programs is Medicare, which provides medical care for everyone over 65 and for certain disabled persons. Medicare has three components—hospital insurance (Part A) and supplementary medical insurance, which pays for

---

[6] There are also hybrids of the two systems. An example is preferred provider plans that pay for all the costs of visits to its preferred providers. If the member goes to another provider, the plan pays for only a fraction of the costs. Similarly, many HMOs allow patients to go to other doctors, but reimburse only a fraction of the costs.

physicians' services (Part B) and prescription drugs (Part D).[7] Recipients must make some contributions to supplementary insurance, but their contribution covers only about a quarter of the costs of even this part of Medicare. The rest comes from general tax revenues. Hospital costs are covered by a 2.9 percent payroll tax (paid by working people as part of the Social Security tax, although half is nominally paid by employers). The money from the 2.9 percent payroll tax goes into a trust fund, which has been near exhaustion a number of times. Each time, quick fixes have enabled the fund's life to be extended.

Medicaid provides medical care for certain low-income families with dependent children, and for most poor aged, blind, and disabled persons. Unlike Medicare, Medicaid is administered by the states. The eligibility standards are set by each state within federal guidelines, and the states are required to pay for between 17 and 50 percent of the benefits, depending on the per capita income of the state. The federal portion is paid out of general tax revenues. States also pay 50 percent of administrative costs.

The rapid growth in the costs faced by states under Medicaid has imposed serious problems for states, just as the rapid costs for health care services under Medicare and Medicaid poses one of the most serious fiscal problems for the federal government. It is worth noting, though, that reform proposals that simply shift the burden to individuals will not improve standards of living: they simply change how services are paid for. By contrast, improvements in the efficiency of the health care delivery system could result in lower costs (whoever pays) without adverse effects on health. By some estimates, if the United States had as efficient a health care system as several of the European countries that have *better* outcomes, the government's fiscal deficit would be eliminated entirely.

OTHER EXPENDITURE PROGRAMS

Three other categories of direct expenditures should be mentioned: the Public Health Service, VA hospitals, and expenditures on medical research and teaching. The Public Health Service, which traces its roots to 1798 with the establishment of the Marine Health Service, was created in part to address important externalities associated with communicable diseases. (If individuals have a contagious disease, they impose an externality on others by increasing the likelihood that the disease will spread.) The Public Health Service played a pivotal role in reducing or eliminating contagious diseases such as smallpox and measles.

---

[7] There is also a Medicare Part C, which is a combination of Parts A and B, provided by Medicare-approved private insurance companies.

Another major government medical program is the VA hospitals, run by the Veterans Administration. Dating back to the establishment of the Veterans Administration in 1930, the VA hospitals provide medical care for those injured while serving in the armed forces and for other veterans. As those who fought in World War II have grown older, requiring more medical care, increasing demands have been imposed on the system. VA medical costs reached $46 billion in 2010, almost half of total VA expenditures.

Expenditures for medical research and teaching are another important category of expenditures. The government runs some very successful research laboratories (the National Institutes of Health) and supports extensive research in universities and medical schools. Medical research expenditures amounted to $49.3 billion in 2010.[8]

TAX EXPENDITURES

For more than four decades, there has been concern about the uninsured. Without insurance, a serious illness or accident can cause a huge financial burden on an individual. Accordingly, the government has encouraged employers to provide insurance by giving their employees a significant tax advantage: employees do not have to pay taxes on the value of their employer-provided health benefits. In 2011, the value of this tax break was estimated to be $163 billion.

Government provides a second substantial tax break for health expenditures: medical expenses in excess of 10 percent of income are deductible from income. The estimated value of this tax break in 2011 was $8.3 billion.[9] The rationale for this provision is that taxes should be based on some measure of ability to pay, and that large medical expenses—to the extent that they are not voluntary—reduce an individual's ability to pay.

These **tax expenditures** encourage both health insurance purchases and medical expenditures. They effectively lower the price an individual must pay for insurance, and health insurance lowers the price individuals have to pay for medical care.

There is concern not only about the efficiency consequences of our tax expenditures—that is, the excessive consumption of medical services that is induced—but also about their equity. Tax benefits are clearly larger for those at higher marginal tax rates (wealthier individuals). The unemployed and those at low-paying jobs with few or no benefits

---

[8] Centers for Medicare and Medicaid Services, Office of the Actuary, National Health Statistics Group, *National Health Expenditure Accounts, 1960–2010*; and National Center for Veterans Analysis and Statistics, Department of Veterans Affairs.
[9] Office of Management and Budget, *Analytical Perspectives, Fiscal Year 2013 Budget*, Table 17-1.

obviously do not enjoy the tax expenditures associated with employer-provided medical insurance.

Prior to 1981, these tax expenditures were so regressive that total federal expenditure, per capita on high-income individuals, including tax expenditures, actually exceeded those on middle-income individuals, even though direct expenditures through Medicare and Medicaid were considerably lower on a per capita basis for upper-income groups. In fact, expenditures per capita on the middle-income group were lower than those on any other group. Changes to the tax code in the 1980s reduced marginal tax rates and increased the minimum expenditure required for tax deductibility of medical expenses. These changes made health expenditures less regressive.[10]

RATIONALE FOR A ROLE OF GOVERNMENT IN THE HEALTH CARE SECTOR

At the beginning of this chapter we noted that rising costs are a major concern about the U.S. health care system. Typically, an increase in the price of some commodity is not by itself taken as grounds for government intervention. Market prices change in response to demand and supply, resulting in changes in scarcity.[11] Similarly, the fact that Americans spend more on health care (as a percent of GDP) than other countries may simply be a reflection of preferences. (We would not infer from the fact that Americans prefer larger cars that there is something wrong with the U.S. automobile market requiring government intervention.) The health care market, though, is rife with imperfections, and some of these do lead to excessive expenditures. In this section we begin with general theory, and then focus on two special aspects of the U.S. health care market.

We will first discuss four market failures. The first two, imperfect information and limited competition, were discussed in general terms in

---

[10] See G. Wilensky, "Government and the Financing of Health Care," *American Economic Review* 72, no. 2 (May 1982): 205; J. Gruber, "The Tax Exclusion for Employer-Sponsored Health Insurance," Working Paper 15766, National Bureau of Economic Research, February 2010; and M. Pauly, T. McGuire, and P. Barros, eds., *Handbook of Health Economics*, vol. 2 (Waltham, MA: North Holland, 2012).

[11] In competitive markets, increases in expenditures can result from a shift in the demand curve along a given supply curve and/or from an upward shift in the supply curve along an inelastic demand curve. Both of these may well have occurred, with demand shifting as Americans have become more health conscious, and the supply curve shifting as health care costs have been rising faster than the costs of goods and services in general. As we comment below, however, the concern is that the increases in health care expenditures do not reflect the normal workings of competitive markets.

Chapter 4. The two others relate to the large roles of nonprofit institutions and of the insurance industry in the health care sector.

Even if there were no market failures, there might be a role for government: some people might have such a low income that they could not afford or would choose not to get adequate health care. The final section addresses these concerns about the provision of health care to the poor.

IMPERFECT INFORMATION

When consumers go to a doctor, in large measure they are buying the doctor's knowledge and/or information. As a patient, the consumer must rely on the doctor's judgment as to what medicine is required or whether an operation or other procedure is advisable. Because they lack medical expertise, patients generally cannot effectively assess and evaluate their doctors' advice. They may not even be able to tell whether a doctor is qualified. This explains why government has long taken a role in licensing doctors and regulating the drugs they can administer to their patients. [12]

These information problems are far more severe than those faced by consumers in other areas. In the case of repeat purchases, like groceries, consumers either are able to judge the quality of the products themselves or come to rely on a grocery store (say, for the freshness of its vegetables). Typically, however, individuals do not have repeat purchases of medical procedures, such as kidney transplants, heart bypass surgery, or even ulcer treatments. In the case of products like cars, independent rating agencies, such as Consumers Union (which publishes *Consumer Reports*), test the product and describe its strengths and weaknesses. However, there are simply too many doctors and hospitals and too many procedures for that to be feasible in the health care sector: a hospital may be good at one procedure and weak at another. Success may depend on subjective factors like a doctor's personal manner, which may work well with some individuals and not with others. Even "report cards"—say, the fraction of heart bypass patients who survived one year—may not be fully informative, because hospitals in one region may have an intake of sicker patients than hospitals in another.

Insurance companies also encounter information problems relating to doctors and patients. Like patients, they must rely largely on doctors

---

[12] Kenneth Arrow has emphasized the importance of imperfect information in medical markets. See K. J. Arrow, "Uncertainty and Welfare Economics of Medical Care," *American Economic Review* 53 (1963): 941–973.

to determine what procedures are necessary and useful. Imperfect information about patients creates problems in the market for insurance, as discussed later (see the section on adverse selection).

LIMITED COMPETITION

Imperfect information decreases the effective degree of competition.[13] A firm selling a standard commodity, like a Sony television, knows that it can attract customers away from other stores by lowering its price. Customers can easily ascertain where they are getting the best value for their money.

By contrast, potential patients who see a doctor with lower prices than competitors may infer that this doctor is not in great demand and is therefore trying to attract more customers; but the apparent lack of demand for the doctor's services may also suggest to them that he or she is not a good doctor.

By the same token, the heterogeneity of medical services makes price and quality comparisons difficult and thus inhibits the effective dissemination of information. My neighbor may be pleased with the medical treatment that he obtained from his doctor, but if his medical problems are different from mine, his satisfaction is no assurance that I will be pleased if I go to the same doctor. And if I hear that one doctor charges more than another doctor, to evaluate whether one is a better buy, I would have to know precisely what services were performed.

The practices of the medical profession may compound the inevitable limitations of competition resulting from imperfect information. In the past, doctors were not allowed to advertise. In other contexts, restrictions on advertising have been shown to raise prices (because they inhibit competition). Thus, for instance, several states now allow advertising for eyeglasses, and in those states, there has been a dramatic decrease in the price of eyeglasses. In many states, doctors and hospitals have been allowed to advertise, but the effects on competition are not yet clear. In particular, much of advertising is not informative. Given the potentially huge variations in quality, simply knowing about price is only of limited value.

Probably of more value will be government assessments of hospitals, for example, in terms of the numbers of particular procedures they perform and the outcomes, and accessible information about the relationship between certain procedures and health outcomes. Some costly procedures have been shown to have little statistical relationship to health outcomes.

---

[13] See, for instance, S. Salop, "Information and Monopolistic Competition," *American Economic Review* (May 1976): 240–245.

One of the objectives of the PPACA was to undertake more assessments of the benefits and costs of different medical practices to help increase the efficiency of the health care system.

There are other measures by which doctors can attempt to restrict price competition. It has been suggested, for instance, that "lowering fees might provoke one's colleagues to deny a surgeon hospital privileges or seek to damage his reputation."[14]

Finally, in many small communities, there are few doctors from whom to choose. The grouping together of doctors into HMOs and preferred provider plans may actually be significantly lessening competition in these areas.

Moreover, there is also limited competition among hospitals. Most smaller communities have at most only a few hospitals. In the event of an emergency, an individual seldom is in a position to choose from among many hospitals. Even when there is time to make a choice, the choice is made not by the individual but by his or her doctor.

ABSENCE OF PROFIT MOTIVE

Another important difference between medical markets and standard competitive markets is the large role of not-for-profit organizations in the provision of health care. For a long time, not-for-profit hospitals vastly outnumbered their for-profit counterparts; even today, the majority of hospitals in the United States are not-for-profit institutions. Such institutions do not view their objective as simply minimizing the cost of delivering medical care, or maximizing profits.

In contrast, in *theory*, for-profit hospital chains respond more clearly to incentives for efficiency, and at times, some of the large for-profit hospital chains seemed both efficient and innovative.[15] More recently, the concern has shifted to perverse incentives for lowering quality, especially in circumstances under which quality is hard to judge, and much of the money the for-profit institutions collect goes to pay high salaries of administrators and dividends, not to provide better health care. For-profit blood providers, for instance, earned a reputation for gathering blood from drug addicts and others whose blood was more likely to carry disease. As a result, that market has been dominated by not-for-profit firms.

Incentive problems encountered with both for-profit and not-for-profit providers arise in large measure from imperfect information. If consumers

---

[14] For a detailed examination of these issues, see M. Gaynor and R. Town, "Competition in Health Care Markets," Chapter 9 in *Handbook of Health Economics*, vol. 2, ed., M. Pauly, T. McGuire, and P. Barros (Waltham, MA: North Holland, 2012), pp. 499–638.

[15] In 1986, Humana, a large private health care chain, received considerable attention for the successful implanting of an artificial heart in its Louisville, Kentucky, hospital.

could easily ascertain the quality of what they purchased, both types of providers would have stronger incentives to achieve efficiency and high quality.

SPECIAL CHARACTERISTICS OF THE U.S. MARKET

Two characteristics of the U.S. health sector exacerbate the problems identified thus far: the prevalence of third-party payments and the fee-for-service system.

THIRD-PARTY PAYMENT   For an ordinary commodity, for which the consumer directly pays the full price, it can be taken for granted that the consumer believes that the benefits of the commodity are at least as great as its cost. Health care differs from an ordinary commodity in that consumers are insulated from cost considerations at the point of consumption, partly through private insurance and partly through government programs. As we have noted, individuals pay for less than 14 percent of all personal health care costs, an even smaller fraction of doctor costs, and a negligible fraction of hospital costs (around 3 percent). Because so much of medical expenditures are paid for by third parties, consumers have little incentive to be cost conscious. The force of the price system is greatly diminished.

FEE FOR SERVICE   The fee-for-service system that is prevalent in the United States may also exacerbate the problems arising from imperfectly informed consumers. The patient goes to the doctor for advice about what medical services are necessary and appropriate. However, the doctor, like any purveyor of services, has a vested interest in selling more services: the more services the doctor provides, the higher his or her income. The threat of **malpractice suits** (see case study, "Medical Malpractice") also encourages doctors to be very cautious. In addition, patients tend to be risk averse and demand any test or treatment that might protect their health. As a result, there is a tendency for doctors to recommend, and for patients to accept, high levels of service. These problems are exacerbated by the fact that a third party (the insurance company or the government) pays the costs.

HMOs may resolve some of these problems by altering the incentives faced by doctors and patients. Evidence shows that there are important differences in the care provided and utilized under the fee-for-service and HMO systems. Generally, utilization of health care services is reduced when doctors are compensated with salaries or capitation systems, as is common in HMOs.[16]

---

[16] L. DeBrock and R. J. Arnould, "Utilization Control in HMOs," *Quarterly Review of Economics and Finance* 32, no. 3 (Autumn 1992): 31–53. See also Laurence Baker, "HMOs and Fee-for-Service Health Care Expenditures: Evidence from Medicare," NBER Working Paper 5360, November 1995, p. 27; and David M. Cutler, "A Guide to Health Care Reform," *Journal of Economic Perspectives* 8, no. 3 (Summer 1994): 13–29.

One of the concerns with HMOs, though, is that their incentives may lead to underprovision of services. This concern has grown as competition among HMOs has intensified and profit margins have been cut.

THE ROLE OF THE HEALTH INSURANCE INDUSTRY

Individuals are risk averse; that is why they buy insurance. They would rather pay a certain amount every year to the health insurance company than go one year with few expenditures because they are lucky and have no illness or accident, and another year with high expenditures when they are less lucky. This is all the more so because when expenditures are very high, income may be particularly low because those with health problems may not be able to work. Indeed, if, on average, individuals expect to pay $4000 in medical bills, most individuals are willing to pay an insurance company more—a "risk premium"—to avoid the risk. Thus, even if there

MEDICAL MALPRACTICE

Consumers are frequently disappointed with what they purchase. In the context of health care, people who believe they have received poor medical treatment often file malpractice suits against physicians and health care providers.

Mixed trends over the past two decades have drawn considerable attention to this contentious issue: between 1992 and 2001, median trial awards increased 70 percent, but the number of medical malpractice jury trials remained stable; furthermore, the cost of medical malpractice insurance rose from $5.4 billion in 1991 to $11.2 billion in 2008, although total losses incurred fell 45 percent from 2003 to 2008 and most medical malpractice claims were closed without any compensation provided to those claiming a medical injury.

Some are concerned that malpractice claims have led to unreasonably high malpractice insurance premiums, which have in turn driven up health care costs and caused doctors to abandon high-risk specialties such as obstetrics. There is also alarm at the rise of "defensive medicine," whereby doctors order unnecessary diagnostic tests and prescribe ineffective medication to avoid claims of medical negligence. Defensive medicine is practiced because a doctor can be sued for failing to administer some test, even if the costs relative to the information it yields are high, or for failing to prescribe some drug, even if the probability of its having a beneficial effect is low. Because third parties bear most of the costs of such drugs and tests, doctors and their patients have every incentive to make use of them. The problem is further exacerbated as the defensive practices become widely used, and thus become the standard of practice against which malpractice is judged.

What trade-offs would arise from limiting malpractice suits? On the one hand, economic efficiency would probably be increased through a reduction in needless expenditures. Excessive legal expenditures, which now equal the amount received by plaintiffs in such cases, could be reduced, as might excessive expenditures designed to forestall

malpractice suits. Limiting malpractice liability costs could also reduce physician shortages in high-demand medical specialties that are also considered high-risk from a legal perspective—many physicians are especially unwilling to go into fields in which there is a higher chance of a malpractice suit.

On the other hand, a system that limits malpractice suits (or the magnitude of rewards) might result in some individuals not being adequately compensated for their injuries. The Institute of Medicine estimates that between 44,000 and 98,000 people die in hospitals each year as a result of medical errors that could have been prevented, and that preventable medical errors in hospitals cost between $17 billion and $29 billion annually as a result of additional care necessitated by the errors, as well as lost patient income and productivity. Limits might also fail to significantly reduce medical expenditures.

There are, however, ways of providing compensation for those who suffer as a result of inappropriate health care (whether an honest mistake or the negligence of a doctor or a hospital) other than through malpractice. New Zealand, for instance, has instituted such a scheme.

The evidence suggests that increases in malpractice payments are not a major factor driving either increases in insurance premiums or overall health care expenditures, although the threat of malpractice lawsuits has contributed to increased expenditures on diagnostic procedures. Moreover, although many states have amended their laws to discourage liability suits, states with tort reform (as such changes are referred to) do not have lower medical malpractice premiums or defensive medicine than states without such reforms.

SOURCES: National Association of Insurance Commissioners, *Countrywide Summary of Medical Malpractice Insurance, Calendar Years 1991–2008* and *Statistical Compilation of Annual Statement Information for Health Insurance Companies in 2009*; T. H. Cohen and K. A. Hughes, *Medical Malpractice Insurance Claims in Seven States, 2000–2004*, NCJ Report 216339, Bureau of Justice Statistics, Department of Justice, March 2007; Daniel Kessler and Mark McClellan, "Do Doctors Practice Defensive Medicine?" *Quarterly Journal of Economics* 111, no. 2 (1996); Institute of Medicine, *To Err Is Human: Building a Safer Health System*, November 1999; U.S. Government Accountability Office, *Medical Malpractice Insurance: Multiple Factors Have Contributed to Increased Premium Rates*, GAO-03-702, June 2003; and K. Baicker and A. Chandra, "Defensive Medicine and Disappearing Doctors," *Regulation* (Fall 2005).

are substantial transactions costs (the costs of filing and processing insurance claims), insurance may be desirable.

There has been considerable dissatisfaction with the insurance provided by the market. The complaints are that (a) some people buy too much insurance, and insurance induces excessive expenditure on health care; (b) many people cannot obtain insurance (there is too little coverage) or can obtain it only at an excessive cost; (c) transactions costs, including profits, are excessive; (d) competition among the health insurance providers is too limited; and (e) insurance firms have restricted consumer choice (e.g., in their choice of doctors) and, in their attempt to keep costs down, have squeezed doctors and often denied individuals access to the care that they need and that their doctors think is appropriate. The problems interact: lack of competition leads to higher profits, and the high profits provide the insurance companies with the resources and incentives to lobby Congress against many reforms that would increase the overall efficiency of the health care system.

INSURANCE AND EXCESSIVE EXPENDITURES ON HEALTH CARE

When individuals buy insurance, they no longer pay the full costs of health care. If illnesses simply happened, unaffected by the actions of the individual, and if doctors faced no choices in how much to spend to deal with the illness, then there would be no problem. However, individuals can often affect their likelihood of needing health care by taking preventive actions. (As a result of insurance their incentives to do so are reduced, although this effect may not be too significant, as most of the "costs" of an illness are probably the discomfort costs and risks borne by the individual, not the doctors' and hospitals' medical bills.)

More important, however, is the fact that there is some discretion in treatment. There may be some benefit from staying in a hospital an extra day, but the benefit may be far less than the cost. A very expensive drug may represent a very slight improvement over a much cheaper drug. A patient might not be willing to pay the cost, but would not turn it down if offered the more expensive drug for nothing, or for pennies.

The demand curve for health services, like the demand curve for any other commodity, is downward sloping—and because it is downward sloping, lowering the *marginal* cost paid by the patient increases the utilization. With many medical plans, patients pay only 20 percent of the true marginal costs; with some, they pay essentially none of the costs. (Demand is still limited, because there is a time and discomfort cost to spending time in the hospital or going to the doctor.) Figure 13.4 shows a downward-sloping demand curve, and demonstrates how consumption is increased from what it would be if individuals paid the full marginal cost. The magnitude of the increase depends on the elasticity of demand. One estimate puts the demand elasticity for medical care at 0.14, so that with a 20 percent co-payment (the individual pays 20 percent of the marginal cost) demand is increased by 11.2 percent (price is lowered by 80 percent, so demand is increased by 0.14×0.80).[17]

The issue is not so much the increase in the health care expenditures, but the fact that *at the margin* the social benefits of the extra expenditures are less than the costs. Indeed, at the margin, the marginal benefit (with a 20 percent co-payment) is only equal to 20 percent of the marginal cost.

The fact that, with insurance, individuals spend too much on health care is called the **moral hazard** problem. The insurance industry has long worried that insurance would increase the likelihood of whatever was insured against; it thought of this as a *moral* problem—with excessive fire insurance, say, an owner of a property might even be induced to burn down his or

---

[17] Joseph P. Newhouse, *Free for All? Lessons from the RAND Health Insurance Experiment* (Cambridge and London: Harvard University Press, 1993), p. 120.
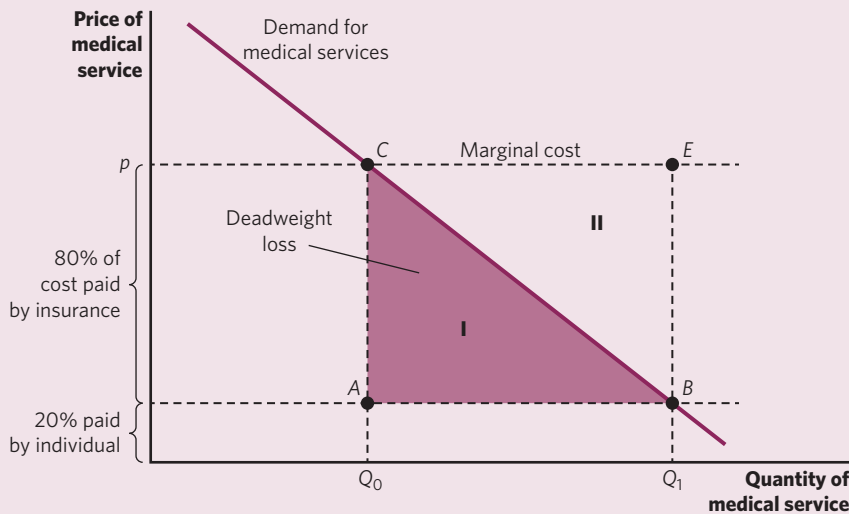
INSURANCE AND THE UTILIZATION OF HEALTH CARE SERVICES

Insurance lowers the price paid by the individual and thus increases the quantity of health care services consumed, from Q_0 to Q_1. For these additional expenditures, marginal costs (reflected in the market price, p) exceed the marginal benefit to the individual (reflected in willingness to pay, as evidenced by the individual's demand curve). The area ABC measures the deadweight loss from excessive consumption.

her building. Economists, however, view the issue simply as one of *incentives*. With insurance, incentives to maintain health and control health expenditures are attenuated. There is a trade-off: the more insurance, the less risk the individual faces, but the weaker the incentives and thus the greater the overall costs. Optimal insurance balances these trade-offs.[18]

The significance of the moral hazard effect depends, of course, on the particular issue being addressed. Individuals are not likely to chose to have quadruple heart bypass surgery simply because it is less expensive if they have insurance. They will do what the doctor recommends. Furthermore, because a very large fraction of overall medical expenditures is for such "catastrophic" events, including the treatment of cancer, accidents, heart attacks, and a variety of diseases of old age, many question the relevance of moral hazard for these expenditures.

There is a concern that government policies have led to excessive insurance, and because insurance is excessive, expenditures on health care are excessive. The tax system subsidizes insurance. Employer-provided insurance premiums are essentially a tax-free form of compensation. The marginal cost of providing health care through employer-provided insurance is thus markedly lower than if the individual buys the health care services directly (or buys insurance directly). These tendencies for excessive insurance were

---

[18] That is, with optimal insurance, individuals will spend more on health care than they would with no insurance, but this is the "price" that must be borne to reduce the risks they face.

exacerbated by the way that employers used to allow employees to choose among insurance plans. The employer would pay a fraction of the additional costs of the more expensive plan (which traditionally provided more services). Thus, although the marginal benefit of the extra insurance to the individual could be less than the marginal cost, the individual would still choose the more expensive plans. As many firms have worried about the soaring costs of their health programs, they have required individuals to pay the full marginal costs if they choose a plan costing more (and typically providing more extensive coverage) than the basic plan offered by the firm.

ADVERSE SELECTION   Another problem that arises in many insurance markets is **adverse selection**. Consider a simple situation in which the insurance firm could tell nothing about an individual, other than that he or she was willing to purchase an insurance policy at the premium being offered. At higher premiums, those who are least likely to need medical care—say, healthy young individuals—decide it is not worth paying that premium, or they may decide to buy a policy that covers only very large medical expenses. Thus, as premiums increase, there is an adverse selection effect: the best risks decide not to purchase the policy. This is illustrated in Figure 13.5A. With the best risks dropping out of the market, however, the average cost per policy issued increases. Figure 13.5B illustrates the competitive market equilibrium, which occurs when the premium equals the expected payout per policy. Because there are few highly risk-averse individuals who are willing to pay a large amount in excess of their expected costs, a small fraction of the population obtains insurance. These individuals pay a high price for *not* having insurance. Individuals buy insurance because they are *risk averse*; that is, they do not like the volatility of what they would have left over to spend on other things if they faced a big health care bill. The result of a market economy in which individuals feel that premiums are too high to buy insurance is that these individuals must bear the costs of the risks themselves.

Insurance firms do not, of course, sit idly by, accepting anyone who applies. They actively attempt to increase the "quality" of those they insure. They do this in several ways, each of which has adverse effects on the effectiveness of insurance. Two of the most common tactics are now banned by PPACA. First, they often did not provide insurance for "pre-existing conditions"—for instance, illnesses whose onset occurred before the purchase of insurance. In some cases, insurance firms simply refused to insure those with a bad health history. These provisions had a particularly adverse effect on the ability of individuals to switch jobs. If anyone in the worker's family had an illness, the worker could become locked into his or her current job, because if the worker moved, the insurance at the new company often did not cover the medical costs of that illness. Second, insurers imposed limits on the extent of coverage. Although insurance is supposed to cover big losses—those losses which individuals are
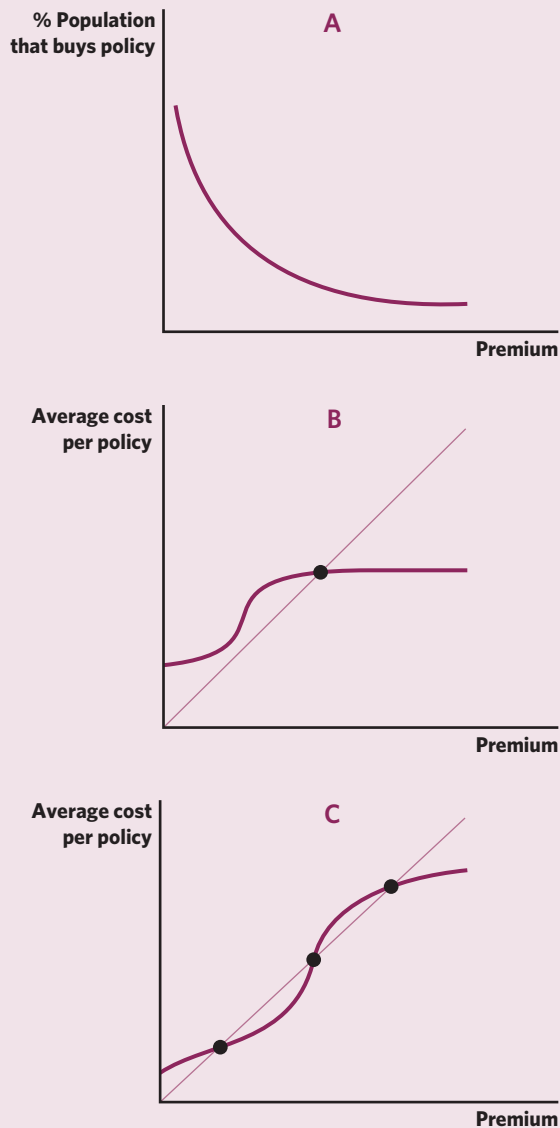
ADVERSE SELECTION
(A) As the premium rises, the percentage of the population purchasing insurance decreases, with the best risks—those with the least chance of needing insurance—dropping out of the market first. As a result, the average payout increases as the premium increases. (B) Shows the market equilibrium, at which the premium equals the average payout. In equilibrium, either relatively few or relatively many individuals may remain uncovered. (C) Shows that indeed there may be multiple equilibria: in the high-price equilibrium, relatively few people are covered; in the low price, most people are covered.

least able to bear—in fact, insurance markets simply failed to cover them. Even insurance designed to cover major medical losses typically put caps on claims—for instance, at $1 million. However, even under PPACA, insurers may concentrate their selling efforts in low-risk communities. Health status varies greatly across different communities; the incidence of malnutrition, violence, drugs, and alcoholism—all of which can lead to high medical costs—is much higher among some socioeconomic groups than others.

These attempts by insurance companies to improve the mix of those covered are sometimes referred to as *cherry picking* or *cream skimming*. The important point is that insurance firms may be able to increase their profits more by engaging in such activities than by increasing the efficiency with which they provide insurance services. Although the drive to increase profits by reducing costs is clearly socially beneficial, the drive to increase profits by cream skimming is more problematic: the gains to one firm occur largely at the expense of other firms, which find that they have a worse mix of insured individuals.[19]

Concern about the lack of coverage for certain high-risk groups in the population—in particular, the elderly—has formed one of the strongest motivations for the expansion of the government's role in health care.

The costs that insurance companies bear to try to make sure that they insure the best risks contributes to the inefficiency of America's private health insurance industry. Moreover, there is a moral issue: Should someone who has the misfortune of having been born with a heart condition, which will shorten his or her life, also be asked to pay more for insurance? Most other advanced industrial countries have answered that question in the negative, and have provided some system of *universal health insurance*, sometimes run by the government, sometimes run by private or not-for-profit entities.

TRANSACTIONS COSTS   The costs of buying and selling goods, including the cost of running markets, are referred to as transactions costs. These are the costs associated with making the economic system work.

The costs of running the insurance system appear to be high. In a competitive world without transactions costs, all the money going into an insurance company in the form of premiums would be paid out in the form of benefits. However, there are costs of selling insurance policies and paying the benefits. For instance, insurance firms spend large amounts in an effort to identify good and bad risks. These transactions costs are relatively low for policies provided through large firms—administrative expenses amount to about 5.5 percent of the claims paid—but small firms face a much heavier burden, with administrative expenses amounting to 40 percent of the claims paid. Private health insurance companies spent an average of 12 percent on administrative expenses in 2010.[20]

These direct transactions costs, however, are not the entire transactions costs of the system. Doctors and hospitals must fill out vast numbers

---

[19] The issue is somewhat more complicated because there can be efficiency gains from sorting individuals according to their risks. For instance, it may be useful to ascertain that some individuals are bad drivers in order to discourage them from driving. Charging higher premiums for health insurance to those who smoke or drink may similarly have beneficial incentive effects. However, this is not the case for most of the sorting that occurs in health insurance markets.

[20] Centers for Medicare and Medicaid Services, Office of the Actuary, National Health Statistics Group, "National Health Expenditure Accounts, 1960–2010."

of forms. By one estimate, 20 percent of expenditures on hospitals actually goes to administrative costs. Still more costs are borne by the insured as they try to collect on their insurance. Critics of the current system argue that standardization of forms and other reforms of the health care system could substantially reduce those costs.

Even with standardization, however, transactions costs would likely be substantial. Insurance firms need to monitor doctors and hospitals to make sure that claims are legitimate. Critics of government insurance programs argue that in spite of their high transactions costs, they have done an ineffective job of monitoring. The estimated cost of fraud and abuse in public and private health care spending is more than $100 billion annually.[21]

Because of the incentive effects and large transactions costs typically associated with insurance, most economists believe that insurance markets should focus on large losses, those that the individual would find difficult to bear. Unfortunately, many individuals feel that they are not getting their "money's worth" from insurance unless they regularly get at least something back; insurance companies have thus found that customers are attracted to covering small losses, such as, in automobile insurance, towing costs.

Modern behavioral economics has helped explain why individuals act in such seemingly irrational ways—and how insurance companies have learned how to exploit these irrationalities to induce individuals to buy insurance products for low-probability events at prices well above those based on actuarial odds.

> ## PROBLEMS WITH MARKET-PROVIDED HEALTH INSURANCE
>
> - Moral hazard—reduced incentive to avoid insured-against event.
>   - Trade-off between risk and incentives.
>   - Insurance firms try to limit through use of co-payments and deductibles.
> - Adverse selection—those who choose a particular policy may have risk characteristics that differ from those of the population as a whole.
>   - As price of insurance increases, best risks may decide not to apply.
>   - Insurance firms have incentive to try to insure only low-risk individuals—"cherry picking" or "cream skimming."
> - Transactions costs.

CONSEQUENCES OF INEFFICIENCIES IN HEALTH CARE MARKETS

Failures in health care markets lead to inefficiencies and may contribute to the rising costs that have fueled recent public policy debates. Two important examples of such inefficiencies are the excessive provision of health care services and the provision of inappropriate care.

---

[21] For a discussion of these estimates, see M. K. Sparrow, Testimony at "Criminal Prosecution as Deterrent to Health Care Fraud" before Senate Committee on Judiciary, Subcomittee on Crime and Drugs, May 20, 2009; and W. J. Rudman et al., "Healthcare Fraud and Abuse," *Perspectives in Health Information Management* 6 (Fall 2009).

SUPPLY CREATES ITS OWN DEMAND   Conventional competition theory predicts that an increase in the supply of doctors will lower the price of medical services. However, during the 1970s, when the number of doctors almost doubled—expansion far greater than the increase in population or income—prices for medical services did not fall. One explanation for this was discussed earlier: patients judge quality partly on the basis of price because of their limited information, so doctors are reluctant to cut prices, lest they get a reputation as being second-rate.

For a conventional commodity, if supply increases and price does not fall, there will be *excess supply*. There is some evidence of this occurring in medical practice; one detailed study in a suburban New York community showed surgeons having a workload only one-third of what experts viewed as a full schedule. These results were consistent with similar findings based on aggregate statistics of the number of surgeons and operations.

In health care, though, there is another possible response: doctors may increase the demand for their services, because patients do not know what care is necessary and appropriate and pay little of the cost of their care. *Supply creates its own demand*. There is some evidence, in fact, that increasing the number of surgeons leads to an increase in the number of operations even when prices do not change at all.[22]

Indeed, there is even evidence of what might be viewed as a backward-bending supply curve. In Canada, as the government has attempted to ratchet health care costs down, it has reduced the fees paid to doctors for each service performed, to which doctors responded by performing more services. (To some extent, the doctor may not actually be performing more services, but rather will be billing more of the services actually performed. Thus, a doctor who might have treated a patient for two maladies at the same time, and billed for one office visit, might ask the patient to come back to take a more careful look at the second malady.)

Questions of whether an operation is advisable or necessary are, of course, debatable. Most doctors do not recommend operations simply to increase their own income. In making a professional judgment about whether an operation is desirable, however, time constraints and the demands of other patients have important effects—and patients with limited information are likely to have surgery if their surgeon recommends it.[23]

---

[22] See V. Fuchs, *Who Shall Live?*; W. McClure, "Buying Right: The Consequences of Glut," *Business and Health* (September 1985): 43–46; V. R. Fuchs, "The Supply of Surgeons and the Demand for Operations," in *The Health Economy* (Cambridge, MA: Harvard University Press, 1986), pp. 126–147; and M. Gaynor and R. Town, "Competition in Health Care Markets," Chapter 9 in *Handbook of Health Economics*, vol. 2, ed. M. Pauly, T. McGuire, and P. Barros (Waltham, MA: North Holland, 2012), pp. 499–638.

[23] Although seeking second opinions before undergoing surgery may improve matters, it is far from a solution. In a community in which there is an increased supply of physicians, diagnostic criteria may change, so a second opinion would merely reflect the same general inclination to recommend surgery more frequently.

INAPPROPRIATE CARE   There is considerable evidence that much of medical care expenditures may, in fact, be inappropriate. Perhaps the most telling are the data comparing practices in different locations. For instance, the average length of a community hospital stay in South Dakota is more than twice that in Oregon. This large difference cannot be accounted for by differences in the populations.[24] Huge practice differences occur across states as well as countries. For example, although hospital stays are relatively infrequent and short in the United States when compared with those in other OECD countries, hospital spending per discharge is much greater than in other OECD countries, at almost triple the OECD median.[25]

These variations lead to huge cost differences. Total personal health care spending per capita in Massachusetts is almost double that in Utah.[26] There is also a wide variation in long-term cost trends. For example, in 1992, Boston, San Francisco, and eastern Long Island (New York) had almost identical per capita Medicare spending, but their expenditures grew at very different annual rates over the next fourteen years: 2.4, 3.0, and 4.0 percent, respectively. By 2006, per capita spending in eastern Long Island was $2500 more than in San Francisco, equal to about $1 billion in additional annual Medicare spending. It is generally agreed that differences in age and other demographic characteristics of the population fail to explain most of the variation in practices.[27]

One of the objectives of the PPACA was to improve the overall performance of the health care system by systematically studying these differences in practices, and ensuring that the best practices are used everywhere.

POVERTY, INCOMPLETE COVERAGE, AND THE ROLE OF GOVERNMENT

Most of our discussion so far has focused on market failures, such as inefficiencies in the market arising from imperfections of information and competition. One of the principal reasons for government action in health care, however, has nothing to do with efficiency; even if markets were perfectly efficient, there would be a concern that those who are very poor do not receive adequate health care.

---

[24] National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, *Summary Health Statistics for the U.S. Population: National Health Interview Survey, 2011,* December 2012, Table 3.2.

[25] D. Squires, "The U.S. Health System in Perspective: A Comparison of Twelve Industrialized Nations," The Commonwealth Fund Pub. 1532, *Issues in International Health Policy* 16 (July 2011).

[26] S. Cliff, "Our Health Care Spending, In One Map," *Washington Post Wonkblog*, December 8, 2011.

[27] E. Fisher et al., "Slowing the Growth of Health Care Costs—Lessons from Regional Variation," *New England Journal of Medicine* 360 (February 2009): 849–852.

Many believe that no individual, regardless of income, should be denied access to adequate medical care. If choices must be made, they should be made on the basis not of wealth but of other attributes, such as age, the likelihood of success of a procedure, or, perhaps, random selection. This view holds that medical services are different from clothes, movies, automobiles, and most other commodities. Just as the right to vote should not be subject to the marketplace (individuals are not allowed to buy and sell their votes), and just as when there was a draft, individuals were not allowed to buy their way out of their military obligations (except during the Civil War), the right to live—access to medical services—should not be controlled by the market. The view that there are goods and services like health care, whose availability to different individuals should not depend solely on their income is known as **specific egalitarianism**.[28]

Not all economists agree about whether medical services should be treated differently from other commodities. Many hold that they should not: those who have more money and want to spend it on getting health care should be allowed to do so. Those who hold this view often point out that the relationship between medical care and life (death) is very weak; other factors, such as smoking, drinking, and eating patterns (consumption of vegetables and fruits, for example) have greater effects on longevity. If one wanted to improve the health status of the poor, one could do so in a more cost-effective way by waging campaigns against smoking and drinking, raising taxes on tobacco and alcohol, and encouraging the consumption of fruits and vegetables.

Still a third view—toward which many Western democracies seem to be gravitating—is that everyone should have the right to a certain minimal level of care. The provision of Medicaid can be thought of as reflecting that view.

REFORMING HEALTH CARE

Recent efforts in health care reform have targeted the big issues identified at this chapter's outset: high costs, limited health insurance coverage, and growing fiscal strains. This section describes some of the most important attempts to counter these problems, including the PPACA enacted in 2010, usually referred to as the Affordable Care Act, or just Obamacare (see case study, "Comprehensive Health Care Reform").

---

[28] J. Tobin, "On Limiting the Domain of Inequality," *Journal of Law and Economics* 13 (1970): 263–277.

COST CONTAINMENT

Economists look to improved incentives as the primary way in which costs can be contained.

ELIMINATING TAX DISTORTIONS Eliminating the tax incentives associated with employer-provided insurance has met with strong resistance from labor unions, largely because they have made more extensive health care a major thrust of labor negotiations over the past several decades, accepting better health benefits in lieu of higher wages. (Even though total compensation, including health benefits, has increased roughly in line with productivity, real wages have increased more slowly— in some cases, actually falling, especially for unskilled workers.) There is also worry that without such tax incentives, the number of uninsured would increase, exacerbating other problems in the health care system. A compromise proposal is to cap the tax deductibility, so that at the margin, individuals and firms face the correct incentives, and to make it a condition for eligibility for tax deduction that if employees are given a choice of plans, they must face the full marginal costs of the more expensive plans. Some go further and argue that there should be a tax on "Cadillac plans," because such plans raise excessively the usage of health care. However, most economists would argue that if individuals *choose* to buy such plans (or their unions choose to give up wage increases to get them) they should be allowed to do so. Of course, the government should not *subsidize* such plans through preferential tax treatment.

MORE EXTENSIVE USE OF MANAGED CARE As we noted, managed care has grown extensively, to the point at which the vast majority of those with employer-provided coverage are in HMOs. Ironically, the debate about managed care has shifted. Originally, there was concern about whether it would simply reduce costs by, say, 15 percent, leaving the rate of increase unchanged, or whether it could reduce the rate of increase. The experience in the 1990s, though it left this longer-range question unsettled, demonstrated clearly that HMOs could reduce costs markedly.[29] As they reduced costs, however, concerns were expressed about whether they had too strong incentives for cost containment, unbalanced by commitment to the quality of care. The issue became cystallized in a debate over "drive-through deliveries," in which HMOs allowed new mothers only a 24-hour hospital stay after the delivery of a baby. Other practices, such as a reluctance to pay for expensive new experimental

---

[29] Critics noted that some of the cost savings came from the fact that HMOs that owned their own hospitals managed to reduce the amount of uncompensated care they provided.

COMPREHENSIVE HEALTH CARE REFORM

Since the adoption in 1965 of Medicare and Medicaid as part of President Johnson's "Great Society," all attempts at comprehensive reform of the health care sector over the subsequent 45 years had been abject failures until enactment of the Patient Protection and Affordable Care Act (PPACA) in March 2010, often referred to as "Obamacare."

PPACA is an extraordinarily complex, 974-page law to be phased in over several years. The primary objectives of PPACA are to broaden the coverage and reduce the cost of health insurance and improve access to health care more generally.

Coverage and access is increased by complementing current employer-provided insurance with an expansion of Medicaid eligibility and extension of funding for the Children's Health Insurance Program (CHIP), subsidized private insurance premiums for qualifying low-income individuals, and creation of individual and small-business health insurance exchanges. In addition, health plans will not be allowed to deny coverage for most pre-existing conditions regardless of age, cancel coverage because of changes in health status, or place annual and lifetime limits on essential health benefits.

PPACA costs are to be offset by a combination of increased revenue (taxes, fees, and penalties) and reforms designed to improve the efficiency and effectiveness of the health care system, including measures to combat fraud and abuse in the Medicare and Medicaid programs; administrative simplification and creation of uniform electronic standards and operating rules to promote the spread of electronic health records; more coordinated care for patients with chronic conditions; creation of the Patient-Centered Outcomes Research Institute to provide physicians and patients with information on the effectiveness of medical technologies and interventions; creation of a complementary Innovation Center in the Centers for Medicare and Medicaid Services to develop, test, and evaluate new policies and programs to enhance the quality of care; and establishment of the Independent Payment Advisory Board, a panel of medical experts tasked with improving Medicare's payment system.

The Congressional Budget Office and the Joint Committee on Taxation estimate that PPACA will reduce federal deficits by $143 billion over a ten-year period (2010–2019), and by 2019, will result in 94 percent of the non-elderly, legally present U.S. population having some form of coverage. However, passage of the PPACA was extremely contentious, with all Republicans voting against it in both houses of Congress. The legislation then faced many legal challenges, culminating in the June 2012 Supreme Court decision that affirmed the authority of Congress, under its taxing power, to require, beginning in 2014, that most individuals have health insurance or pay a penalty (known as the individual mandate). However, the Supreme Court also ruled that in the complicated domain of federal authority versus states' rights in the health sector, the provision requiring states to comply with new eligibility criteria for Medicaid or risk losing their funding was constitutional, only as long as the states lose just the new funds for noncompliance, rather than losing all their funding. The result is that some states, including populous states with a significant percentage of low-income residents and limited Medicaid coverage, could continue to have large numbers of uninsured people.

SOURCES: Office of the Legislative Counsel, U.S. House of Representatives, *Compilation of Patient Protection and Affordable Care Act*, May 2010; H. Chaikind, C. W. Copeland, C. S. Redhead, and J. Staman, *PPACA: A Brief Overview of the Law, Implementation, and Legal Challenges*, Congressional Research Service, March 2011; and Henry J. Kaiser Family Foundation, *Timeline: History of Health Reform in the U.S.*, www.kff.org.

procedures, also drew widespread media attention. Drive-through deliveries were banned, and many states proposed a stronger role of government in regulating HMOs.

HEALTH SAVINGS ACCOUNTS Legislation passed in 2003 established **health savings accounts (HSAs)**.[30] These are intended to encourage individuals to buy insurance policies with larger deductibles and co-payments, and thus contain fewer incentives for moral hazard—they offer tax-preferred treatment of money saved for medical expense for individuals covered by high-deductible health plans. Funds allocated to these accounts are tax deductible, but annual contributions are limited and excess contributions are not tax deductible. HSAs eliminate one of the distortions of the existing system under which health care paid for through insurance companies is, in effect, tax deductible, but direct payments are not. Both HSAs and their predecessor, medical savings accounts (MSAs), have been very popular: in 2011, deductions for HSAs and MSAs cost an estimated $1.9 billion in tax expenditures.[31]

One criticism of HSAs relates to the relative strength of the incentive effects for reducing moral hazard, compared to the adverse selection effects. Clearly, those who think it unlikely that they will need much medical care, and rich individuals better able to bear the risk of paying a large deductible or co-payment and for whom the tax break is particularly significant, will be much more attracted to the HSAs. If healthier individuals are attracted to the HSAs, then those not having HSAs will face higher insurance premiums.

Thus, for some individuals, HSAs might prove very attractive; their overall medical costs—insurance plus out-of-pocket expenses—might well go down. The unanswered question is, to what extent are those lower costs a reflection of greater efficiency (a reduction in the distortions arising from moral hazard) and to what extent are they a reflection of cream skimming?

EXTENDING INSURANCE COVERAGE

As we noted at the beginning of this chapter, many people in the United States have no health insurance. A number of reform efforts have attempted to address the problem of the uninsured.

---

[30] HSAs were preceded by medical savings accounts (MSAs), established on a pilot basis for self-employed individuals and employees of qualifying small business employers as part of the Health Insurance Portability and Accountability Act of 1996. The Treasury Department extended the MSA pilot program periodically through 2007, after which no new MSAs could be opened because they had been superseded by HSAs, included as part of the Medicare Prescripton Drug, Improvement, and Modernization Act of 2003. See Internal Revenue Service, Department of the Treasury, *Health Savings Accounts and Other Tax-Favored Health Plans*, Publication 969.

[31] Office of Management and Budget, *Analytical Perspectives, Fiscal Year 2013 Budget*, Table 17-1.

MANDATED HEALTH INSURANCE   Extending health insurance is central to the PPACA. Today, most employers provide health insurance to their employees; beginning in 2014, the PPACA establishes an **individual mandate**, under which most individuals will be required to have health insurance or pay a penalty.

Many advocates of universal health insurance favor a **single payer system** like Canada's, which works much like Medicare, with the government paying medical bills. However, financing such a system would require new taxes—a political impossibility at present. Instead, the individual mandate constitutes a hidden tax, as ultimately the mandate imposes insurance costs on firms and consumers.

PRE-EXISTING CONDITIONS   As noted earlier, many insurance firms have refused to cover health conditions existing prior to the beginning of the insurance coverage. Thus, a worker with a kidney problem would find it impossible to change jobs because the new insurance policy would not cover the costs of continued treatment. Legislation passed in August 1996—called the Kennedy–Kassebaum bill after the senators who wrote it—was devised to prevent insurers from excluding pre-existing conditions from coverage. A number of questions were raised with the passage of the legislation: How much would insurance firms have to raise premiums to extend this coverage? Would employers, concerned that hiring a worker with a pre-existing condition would lead to soaring premiums, discriminate against job candidates they suspected of having such conditions? So far, it appears that the bill's impact on either the insurance or labor market has been limited. Thus, the PPACA strengthens these provisions by prohibiting exclusions for pre-existing conditions.

FILLING IN THE GAPS   When the PPACA was enacted in 2010, approximately 50 million Americans did not have health insurance. Roughly 30 million people are expected to gain coverage under the new law, and half of these will gain coverage through an expansion of Medicaid. (Beginning in 2014, state Medicaid programs must provide coverage to all eligible non-pregnant, non-elderly legal residents with incomes up to 133 percent of the federal poverty level.) Other significant gaps to be filled under PPACA are health insurance for employees of small businesses, via establishment of federally subsidized health insurance exchanges; health insurance for retirees not yet eligible for Medicare benefits (ages 55 to 64), through the Early Retiree Reinsurance Program; and health insurance for young adults, by extending coverage under their parents' plan until the age of 26. Other gaps had been filled by previous legislation, such as health insurance for low-income children who are ineligible for Medicaid

but cannot afford private insurance [32] and for employees who were fired, changed jobs, or got divorced. [33]

MEDICARE REFORM: EASING LONG-TERM FISCAL STRAINS

The long-run fiscal strains arising from the health care system are largely the result of growing Medicare expenditures. Efforts to reform Medicare focus on three issues: improved incentives, better management of health care provision, and more competition.

IMPROVED INCENTIVES The Medicare program itself has large co-payments and deductibles, but most elderly buy, at quite moderate costs, **Medigap** insurance—that is, insurance that covers most of these out-of-pocket costs. For individuals with Medigap insurance, there are few incentives to economize on the use of medical care. Those who buy Medigap insurance do not pay the full incremental cost associated with the policy, and because of this reduced incentives to economize on health care, those with Medigap insurance make more demands on the Medicare system itself. Medigap premiums cover the direct costs incurred by the insurance companies, but not the additional costs imposed on the government because of increased utilization. One set of reform proposals thus would either restrict the coverage that can be provided by such policies or force the policies to pay the full incremental costs.

One of the fastest-growing parts of the Medicare system is home health care. Many individuals do not need to stay in hospitals or long-term care facilities, as long as they have some limited assistance at home. Home health care provides that assistance. However, home health care has been subject to extensive abuse. Unlike doctors, nurses, and other traditional health care providers, who must pass stringent standards, relatively few skills are required of home health care providers, and entry into the business appears easy. Furthermore, although the demand for heart surgery may be relatively insensitive to price (and would be limited even at a zero price), almost anyone offered free assistance would find it useful. Ironically, however, although the Medicare system has deductibles and co-payments associated with most medical treatments, it has not imposed any co-payments on home

---

[32] The Children's Health Insurance Program (CHIP) was created by the Balanced Budget Act of 1997 as Title XXI of the Social Security Act and extended under the CHIP Reauthorization Act of 2009.
[33] Health benefit provisions in the Consolidated Omnibus Budget Reconciliation Act (COBRA) of 1986 require most employer group health plans to provide temporary continuation of group health coverage that otherwise might be terminated, and the Health Insurance Portability and Accountability Act (HIPAA) of 1996 allows the right to enroll in another group health coverage program without waiting until the next open season for enrollment.

health services since 1972. In 2011, the Medicare Payment Advisory Commission (MEDPAC)—an independent congressional agency established by the Balanced Budget Act of 1997 to advise Congress on Medicare financing and related issues—recommended co-payments for home health services, but there has been strong opposition because of claims that it is inefficient and regressive. Critics say co-payments would discourage home health care in favor of more expensive nursing facilities, and would fall most heavily on the poorest and oldest Medicare beneficiaries.

IMPROVED MANAGEMENT   As costs soared soon after the Medicare program was begun, attention focused on providing better incentives to the providers. Initially, Medicare had simply reimbursed hospitals for whatever costs they incurred. Such a system clearly provides perverse incentives. In 1983, Medicare switched to a system in which hospitals were reimbursed a fixed amount for the treatment of a particular illness.

Overall, the system called **diagnosis-related groups (DRGs)** has worked well. Initially, there was concern that hospitals might refuse to provide treatment in areas where the reimbursement rate was too low, or that they would be tempted to cheat on the diagnoses or to cut corners on the services provided, adversely affecting the quality of the care. (Of course, the intent of the system was to cut *unnecessary* costs, such as hospital stays that were longer than needed. The problem is that it is often difficult to distinguish between cutting corners and cutting unnecessary expenditures.) Neither of these fears has come true.

One ongoing concern, however, is the ballooning of out-of-hospital costs. Hospitals have tried to push patients out of the hospital as quickly as possible. Thus, although hospital costs have been contained, overall health costs have not performed as well. This has prompted some to suggest an extension of the system, called **bundling**, to include all the treatment associated with an illness, both in hospital and out of hospital.

Conservatives have long argued that private insurance companies could manage Medicare better than the government. In fact, Medicare beneficiaries have had the option to receive their benefits through private health plans since the 1970s, mainly through HMOs, as an alternative to the traditional federally administered Medicare program. However, this option was greatly expanded under the Bush administration through creation of Medicare Advantage: by 2012, 27 percent of Medicare beneficiaries were enrolled in a Medicare Advantage plan. It turned out, though, that to provide services comparable to those provided by the traditional Medicare program actually increased costs, by as much as 20 percent. In response, the PPACA mandates the gradual reduction of federal payments

to Medicare Advantage plans to bring them closer to the average costs of care under the traditional Medicare program. It also provides for bonus payments to Medicare Advantage plans based on quality ratings, and restricts the share of premiums that these plans can use for administrative expenses and generation of profits.[34]

STRENGTHENED COMPETITION   Although there are many doctors and hospitals in the country as a whole, competition within a given locale may be very limited. Providers have sought immunity from (or at least a loosening of) antitrust provisions, which would reduce competition even further. So far, these pressures have been resisted. Such immunity might, for instance, allow all the X-ray labs in an area to get together to set prices, or a single health provider—HMO or preferred provider organization (PPO)—to sign up all the bone specialists in a community.

Some have argued that the government should strengthen competitive forces by having competitive bidding for the supply of Medicare services. The winning bidder would, in effect, become a **preferred provider**, with the government paying 100 percent of its bid. Other providers might be reimbursed a fixed amount up to a fixed percentage of the winning bid. Critics of such plans worry about ensuring quality and about the development of a two-tier system, in which poor individuals would go to the preferred provider, while wealthier individuals would go to higher-priced providers.

The measures just discussed are not the only cost saving measures that have been proposed. For instance, there was a provision in the law establishing the prescription drug benefit under Medicare (Medicare Part B) restricting the ability of the government to negotiate with the drug companies, a provision that was estimated to increased government spending over ten years by as much as half a trillion dollars. Repeal of this provision and the adoption of standards to ensure cost-effective selection of drugs would accordingly save a considerable amount of money.

FINANCIAL REFORMS   The reforms just described are directed at improving the efficiency with which Medicare services are provided. There is another set of reforms focused more narrowly on lowering the financial drain on the government, either by reducing benefits—such as by increasing the age of eligibility from 65 to, say, 67, just as the normal age of retirement has been increasing under Social Security—or by imposing more of the costs on recipients. Medicare subsidies are, for instance, extended to all participants, regardless of their

---

[34] The Henry J. Kaiser Family Foundation, *Medicare Advantage Fact Sheet,* November 30, 2012.

income, which means that a millionaire retiree is being subsidized by a $20,000-a-year worker. Pressure to increase payments, at least for higher-income beneficiaries, has been mounting in recent years, particularly as some recent studies suggest that the Medicare program overall is regressive. With taxes proportional to income, and all individuals having the same entitlement, it had been thought that Medicare was highly progressive. But when account is taken of the fact that richer individuals tend to live longer and are better capable of extracting services out of the Medicare system, it turns out that the rich get more, so much more that it more than offsets their higher Medicare tax contributions.

REFORMING MEDICAID

Medicaid provides medical assistance to the poor and the disabled and long-term nursing care to the aged. Each of these comprises about a third of the program. The long-term nursing program has been a particular source of concern. Although Medicaid began as a program for the destitute, today about one-third of all nursing home and home health care expenditures are paid by Medicaid. Some elderly individuals transfer wealth to their children so they can qualify to receive these benefits. Although the government has attempted to limit these transfers (typically, funds transferred within three years of entering a nursing home are treated as if they were available to pay for nursing home care), it has had only limited success. Nursing home expenditures are expected to soar as the number of very old Americans—those over 85—increases. Furthermore, whereas just 5 percent of the general population over 65 is in nursing homes, a quarter of those over 85 are, and half of the nursing home population is 85 or older.

The private insurance market for long-term nursing care was slow to develop. In 2010, private insurance paid only 9 percent of total expenditures for nursing homes and continuing care retirement communities (barely higher than the 7 percent figure for 1994).

There have been three major groups of proposals to reduce government costs, all based on the premise that most individuals should make provision for themselves, but if they do not, it is difficult for the government not to provide assistance. One proposal is to mandate insurance: every individual would have to purchase, either privately or through the government, long-term nursing care insurance. Providing it through the government would entail effectively expanding the Social Security program. To finance long-term nursing care, the government would require an

increase in the Social Security tax of around 1.0 to 1.5 percentage points. The second proposal is to increase incentives—both carrots and sticks—to purchase insurance by providing tax preferences and asset protection (under "asset protection," individuals who purchased a minimal amount of nursing home insurance could keep a certain amount of assets to pass on to their heirs or for other expenses without losing eligibility for Medicaid when their insurance ran out), and by reducing the magnitude of public support for nursing care. The third proposal is to assist the private sector in providing meaningful long-term nursing care insurance by providing reinsurance against large losses (associated with long stays) or against hard-to-anticipate long-term trends, either in longevity or overall nursing home care expenses.

REVIEW AND PRACTICE

SUMMARY

1. Although decisions about health are difficult, resource allocations—choices among alternative uses of funds—must be made. Economic analysis may be useful in making those decisions in a systematic and consistent way.

2. U.S. health expenditures are a third higher than the OECD average as a share of GDP, and are the highest among OECD countries as a share of total government expenditures—the government now pays almost half of all personal health care expenditures, whereas out-of-pocket expenditures as a percentage of total health expenditures in the United States is among the lowest of the OECD countries. In spite of the high level of health expenditures, health outcomes are poorer than in other OECD countries. The four major public programs in the United States are Medicare, Medicaid, health care for veterans, and public support for research and development. In addition, there are two major categories of tax expenditures: employer-financed health insurance and tax deductibility of medical expenses exceeding a certain level.

3. Three fundamental problems facing the health care system in the past couple of decades before the passage of PPACA have been:
   a. A large and growing population of uninsured
   b. High and rising health care costs
   c. Growing federal health care expenditures

4. Among the reasons why markets may not provide an efficient allocation of resources in the health care sector are:
   a. Consumers and insurers have limited information.
   b. Competition is limited.
   c. Only a small fraction of health care costs are paid directly by the consumer.

5. Health insurance is associated with the problems of adverse selection and moral hazard. Insurance reduces individuals' incentives to take care of themselves and to economize on health care services, and insurance firms may increase their profits more by ensuring that they cover only the most healthy than by increasing their efficiency.

6. The tax system encourages excessive purchase of insurance and excessive consumption of health care services. Tax expenditures for health amount to billions of dollars. Most economists believe that these tax benefits should be capped, so that individuals will more nearly pay the full marginal costs of services.

7. Concern that the fee-for-service system leads to excessive consumption of health care has led to rapid growth of health maintenance organizations. More recently, there has been a concern that HMOs have been too zealous in controlling costs, sacrificing quality and service.

8. Prior to the enactment of comprehensive health care reform in 2010, including an individual mandate for medical insurance, attention was focused on incremental reforms, such as extending insurance coverage to the unemployed and to children, eliminating restricted coverage of pre-existing conditions, and providing medical savings accounts to encourage individuals to purchase policies with larger deductibles and co-payments.

9. Proposals to reform Medicare focus on improving incentives, introducing better management of health care providers, ensuring more effective competition, and placing more of the financial burden on beneficiaries.

10. The major concern today about Medicaid is the growing burden of nursing home care.

KEY CONCEPTS

Adverse selection

Bundling

Capitation fee

Co-payment

Cost shifting

Deductible

Diagnosis-related groups (DRGs)

Fee-for-service plans

Health maintenance organization (HMO)

Health savings accounts (HSAs)

Individual mandate

Malpractice suits

Managed care

Medicaid

Medicare

Medigap

Moral hazard

Preferred provider

Single payer system

Specific egalitarianism

Tax expenditures

Third-party payments

Uncompensated care

QUESTIONS AND PROBLEMS

1. List the various distortions in incentives that arise in the health care sector. (Be sure to include those that affect the purchase of insurance as well as the purchase of health care services directly.)

2. In what ways is the purchase of medical services similar to the purchase of a car? In what ways is it different?

3. We have noted that there is extensive disagreement on what should be done about the way medical care is provided in the United States. To what extent is this disagreement due to differences in judgments concerning how the market for medical services functions? Be specific. To what extent is the disagreement due to differences in values?

4. Consider the "market failures" that arise in medical markets and current proposals for altering the way medical care is provided in the United States. Discuss the extent to which each of the proposals is aimed at remedying particular market failures.

5. During the past fifty years there has been a decline in community-run hospitals and an increase in private (for-profit) hospitals. Are there reasons why hospitals should be particularly suited or unsuited to being run publicly? What do you think accounts for these trends?

6. Critics of malpractice suits claim that they have contributed significantly to the rise of medical costs and want legislation that would limit the size of awards and lawyers' fees, or otherwise discourage such suits. Many lawyers are concerned that any such legislation would impair the rights of victims of malpractice to be justly compensated for the damages they have suffered. Discuss the equity–efficiency trade-offs. What do you think should be done?

7. Assume that medical expenditures are fully deductible from taxes. Show diagrammatically the effect on the demand for medical services. If the elasticity of demand with respect to price is 0.7, what is the effect of deductibility on an individual in the 15 percent marginal tax bracket; in the 28 percent marginal tax bracket; in the 40 percent marginal tax bracket?

8. Some states have proposed requiring community rating—that is, insurance companies would not be allowed to charge individuals different premiums, regardless of their health,

age, or sex; everyone in the community would pay the same premium. Discuss the possible consequences of community rating. Evaluate such proposals from the perspective of equity and efficiency.

9. Many employers offer their employees a choice of plans, paying a fixed share of the cost of each. What inefficiencies does this introduce? Some employers, such as Stanford University, have instead offered a fixed payment, regardless of the plan chosen, and have insisted that all programs offer identical coverage. Within three years, the cost of providing this standard coverage fell by 20 percent in real terms. Explain why this may have happened.

10. Economists have criticized the tax treatment of health insurance. Why have unions resisted changing this tax treatment? What are the equity and efficiency consequences of capping the tax deductibility at some number, say, $5000, representing the average costs of health benefit plans in unionized firms?

14

EDUCATION

In the United States, education has long been recognized as a responsibility of government. The Land Ordinance of 1785, enacted even before the Constitution, set aside land in the newly established Western territories to fund public schools. Although the locus of responsibility remains at the state and local levels—education is the single largest expenditure at those levels—today the federal government finances about 15 percent of all public current and capital education expenditures, using these funds for specific purposes, such as providing education for the disadvantaged, encouraging the establishment of educational standards, promoting science education, and providing financial assistance to enable more people to attend college.

In U.S. national politics, bipartisan consensus on the importance of education is the norm. However, major controversies exist over how best to attain a high-quality education system that provides equality of opportunity, even to the poor.

Two major economic issues—a slowdown in productivity and increasing inequality—have informed much of the recent concern about education in the United States. During the decades between World War II

and the early 1970s, output per worker grew at a rate of 2.8 percent per year, dropped to 1.1 percent per year from 1973 to 1979, and then slowly rebounded to reach 2.5 percent annually from 2000 to 2007. Although explanations of this long interlude of productivity slowdown and subsequent substantial increase over the past two decades remain uncertain, many economists and policy makers agree that improvements in human capital—the skills and experience of workers—may hold the key to improving productivity growth, and strong education is seen as critical to that goal.[1]

The early 1970s also marked an important change in the pattern of income distribution in the United States. The fruits of economic growth were shared more equitably during the two decades prior to 1973 than during recent decades.[2] Much of this increase in inequality can be related to education. During the 1980s, the difference between the income of a college graduate and that of a high school graduate increased enormously[3] as the market placed a greater premium on skilled (educated) workers. Thus, providing better education and ensuring that a larger fraction of the population went to college were seen as ways of enhancing opportunity and reducing inequality. Not only would those receiving a college education enjoy higher incomes, but wages would also rise in the market for unskilled labor as the number of unskilled laborers dropped.

For these reasons, and others, economists and policy makers have placed a high priority on strong education—but how strong is the education system in the United States? One of the most impressive achievements of the U.S. system has been the growth in college attendance throughout the latter half of the twentieth century. The GI Bill of Rights, which provided an opportunity for World War II veterans to attend college, transformed higher education from a privilege of the elite to an expectation of the broad middle class. That expectation carried through to another generation as college enrollment rates soared in the 1970s and 1980s,[4] while high school dropout rates fell.

---

[1] U.S. Bureau of Labor Statistics, "Productivity Change in the Nonfarm Business Sector, 1947–2011," *Labor and Productivity Costs*, http://data.bls.gov.

[2] One way of seeing this is to divide the country into quintiles by income, distinguishing the poorest fifth of the population from the next poorest fifth, and so on. Before 1980, all quintiles saw their incomes growing, and the poorest quintiles saw their incomes growing fastest, so their share of aggregate income increased. From 1980 to 2000, the growth rate of the poorest quintile lagged behind that of the richest quintile, and their income has actually declined in constant 2009 dollars over the past decade, while the richest quintile has continued to enjoy income growth. Thus, since 1980, the lowest quintile's share of aggregate income has declined from 4.2 to 3.4 percent, whereas the highest quintile's share has increased from 44 to 50 percent. Rising inequality is even more striking when looking at the relative income growth of America's richest households over the past three decades: since 1980, the top 5 percent's share of aggregate income has increased from 17 to 22 percent. (Source: U.S. Census Bureau, *The 2012 Statistical Abstract*, Table 694).

[3] *Economic Report of the President, 1997* (Washington, DC: Government Printing Office, 1997), Chart 5-4, p. 169.

[4] *Economic Report of the President, 1996*, Chart 7-1, p. 193.

Even so, in recent decades questions have been raised about the caliber of the U.S. education system. International comparisons inform some of these questions. The United States, despite 89 percent of its population aged 25 to 64 having graduated from high school and 32 percent having graduated from college by 2010—among the highest rates of OECD countries—has not fared well on standardized math and science tests administered around the world.[5] For example, in 1995, eighth graders in the United States ranked twenty-eighth in math—behind Singapore, South Korea, Bulgaria, Russia, and most of the developed countries—and seventeenth in science. Even though America's performance had improved considerably by 2011, when it was ranked ninth in math and tenth in science, it was still behind much of East Asia and several European countries. Inequality in the United States may help explain such results. Students outside America's poor urban schools achieve results comparable to their peers in other advanced countries, and although the median math and science scores of American students sometimes rank below those of other high-income countries, the best American students (in the ninety-ninth percentile) rank above their peers from many of those countries.[6]

This brings us to a major issue in recent policy debates: education of the economically disadvantaged, particularly in America's inner cities and impoverished rural areas. Indigent students receive inadequate primary and secondary education, leaving them unprepared for colleges and universities. Without adequate skills, these people earn low wages, and are thus forced to continue to live in poor areas, continuing the cycle of poverty.

The problem has been exacerbated by some other trends in the economy. During the 1980s, while real wages of the poorest declined, real tuition at public two- and four-year institutions of higher learning increased markedly, so the indigent students who were qualified for higher education faced increasingly insurmountable financial hurdles. Thus, even though the increased returns to education noted earlier did lead to an increase in college enrollments, children of more affluent families were better able to respond to these market signals, and the gap between enrollment rates of the disadvantaged and other Americans actually increased.

All these issues have made education a critical concern for the U.S. public sector. This chapter describes the structure of the U.S. education

---

[5] Organization for Economic Cooperation and Development, *Education at a Glance, 2012*.
[6] International Association for the Evaluation of Educational Achievement (IEA), *Trends in International Mathematics and Science Study (TIMSS), 1995* and *2011*, available from the website of the National Center for Education Statistics (NCES), U.S. Department of Education, http://nces.ed.gov. Research conducted by the OECD yields even worse comparative results for the United States; see OECD, *Program for International Student Assessment (PISA), 2009*.

system, explores the rationale for public finance and provision of education in the United States, and analyzes some of these current issues in U.S. education policy.

THE STRUCTURE OF EDUCATION IN THE UNITED STATES

Traditionally, elementary and secondary school education has been the responsibility of local communities. They financed it (usually with property taxes), hired the teachers, and determined the curriculum. Figure 14.1 shows the trends in local, state, and federal spending on elementary and secondary education. From 1945 to 1972, the state percentage of education funds remained steady at 39 percent. During the mid-1970s, state expenditures began to increase at a faster rate than local expenditures,
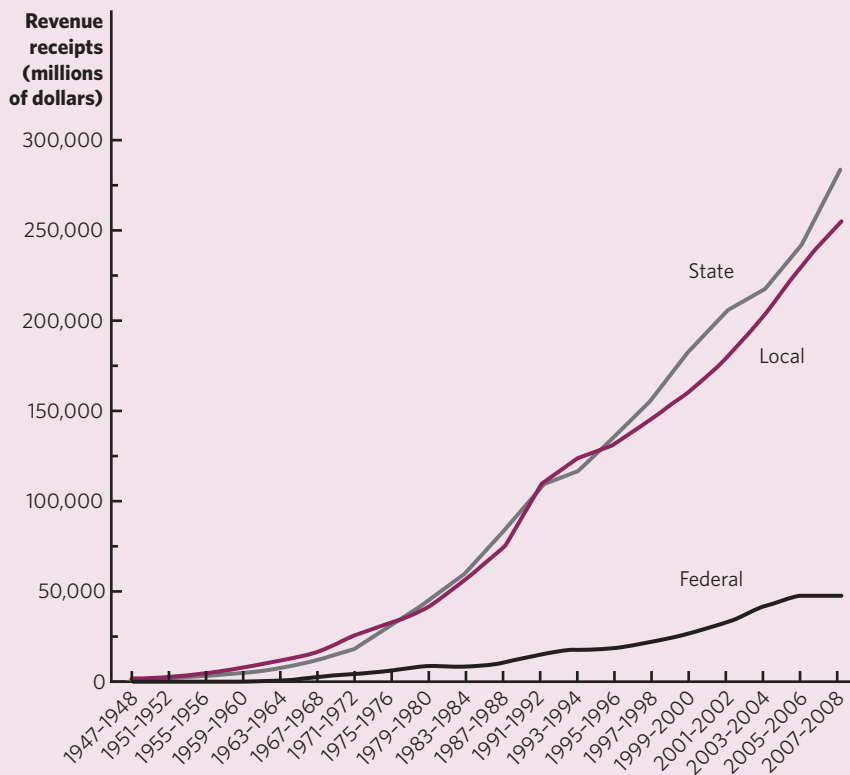
FIGURE 14.1

REVENUE SOURCES OF PUBLIC ELEMENTARY AND SECONDARY SCHOOLS
All sources of support have increased markedly in recent years.

SOURCE: U.S. Department of Education, National Center for Education Statistics, *Digest of Education Statistics 2010*, Table 180.

so that by the end of the decade, states actually paid for a higher percentage of total costs. Today, states finance almost half the cost of public elementary and secondary schools. Federal support for education, mostly for special programs such as aid to schools with a large number of disadvantaged children, increased in the 1960s and 1970s, and by 1980 represented 10 percent of total financing for public elementary and secondary schools. Federal support declined from that level to under 7 percent from the mid-1980s to the mid-1990s, but it has risen to between 8 and 9 percent of total expenditures for the past decade.

There is, however, a great deal of variability, both in the federal government's role and that of the states. For example, whereas in recent years the federal government has provided approximately one-sixth of the funds for primary and secondary schools in Louisiana and Mississippi, it has provided less than 5 percent of the funds in Connecticut and New Jersey. On average, states provided slightly more than half of the nonfederal funds, but this varied, from over 90 percent in Hawaii (where there are no local school districts) and Vermont (where public education is funded primarily by a state-determined share of the residential property tax), to only one-third in Illinois and Nevada. Revenue from private sources, while still comprising a small share of total funding for public elementary and secondary schools, is rapidly increasing in importance, providing 6 percent of nonfederal funds in Tennessee and 5 percent in Oklahoma.[7]

At the same time, the states have taken an increasingly active role in setting certain minimum standards. They set the number of days students must be in school per year; the lowest age at which children may drop out of school; and the minimum education requirements for teachers employed in public schools, often specifying, for instance, not only that the teachers have a college degree, but also that they have taken a specified number of courses in education. States also play a role in determining curricula. For example, many states require courses in American history in the eleventh grade and adopt certain approved lists of textbooks, from which the local community must select the books it will use. Several states have made provisions for state takeovers of local school districts when there is "academic bankruptcy"—that is, when the local school districts fail to meet certain minimal standards.

One consequence of local control is that many rich communities spend far more on education than most poor communities. The range in educational expenditures per student has been enormous, even within individual states, with rich districts spending two, three, or more times the amount spent in poor districts.

---

[7] U.S. Department of Education, National Center for Education Statistics, *Digest of Education Statistics 2010*, Table 181.

About 90 percent of all elementary and secondary school students attend public schools. Of those who go to private schools, about 40 percent go to Catholic schools and another 40 percent to other parochial schools, with the remaining one-fifth attending nonsectarian schools. Private institutions play a much more important role in higher education. About 40 percent of bachelor's, master's, and doctoral degrees are earned at private institutions, almost all of which are not-for-profit, and approximately 60 percent of first professional degrees are earned in private schools.[8]

Because the principle that states and localities should be responsible for making educational decisions is held so dear, the federal government has relied on a variety of indirect ways to try to influence states and localities. In 1957, after the Soviet Union launched Sputnik, the first space satellite, and thus appeared to be surpassing the United States in science, the federal government supported efforts to improve the science curriculum in elementary and secondary schools. In the early 1990s, when it became apparent that many of those graduating from high school who did not go on to college were having a hard time entering the labor force, the federal government sponsored a school-to-work initiative. This encouraged the states to design apprenticeship-like programs (so successful in other countries, such as Germany) that would facilitate the movement from school to work. When, again in the early 1990s, the federal government wanted to encourage higher standards in elementary and secondary schools, it did not impose national standards, but rather provided funding that helped the states establish standards for themselves, and several commissions were set up to help in the formation of national standards in the various disciplines.

FEDERAL TAX SUBSIDIES TO PRIVATE AND PUBLIC SCHOOLS

The personal income tax may have important effects on the demand for public and private education. Expenditures by state and local communities for education (as well as for other purposes) are implicitly subsidized by the federal government because state and local taxes are deductible on federal income tax returns, and interest on state and local bonds is exempt from federal taxation. This means that if my community taxes me $1000 to support public schools, the cost to me is far less than $1000. If I am in

---

[8] First professional degrees include those for medicine, dentistry, pharmacy, veterinary medicine, law, and theological professions. See U.S. Department of Education, National Center for Education Statistics, *Digest of Education Statistics 2010*, Tables 35, 62, and 287.

TABLE 14.1 ESTIMATED FEDERAL TAX EXPENDITURES FOR EDUCATION, FISCAL YEAR 2011 (MILLIONS OF DOLLARS)

	REVENUE LOSS	
	CORPORATE	INDIVIDUAL
American Opportunity Tax Credit		$13,060
Lifetime Learning Tax Credit		2,800
Exclusion of Scholarship and Fellowship Income		3,060
Exclusion of Interest on Student-Loan Bonds and on Bonds for Private Nonprofit Educational Facilties	750	1,830
Parental Personal Exemption for Students Age 19 and Over		5,600
Qualified Tuition Programs		1,610
Deductibility of Charitable Contributions for Educational Purposes	650	2,870
Deductibility of Student-Loan Interest		1,400
Deduction of Non-Business State and Local Taxes (Total = $41,060)		11,907
Other	260	1,850
Total	$1,660	$45,987

SOURCE: Calculated from Executive Office of the President, Office of Management and Budget, *Budget of the United States Government, Fiscal Year 2013, Analytical Perspectives*, Table 17-2. The expenditure resulting from the deductibility of state and local taxes assumes state and local jurisdictions spend 29 percent of revenues on education (calculated from "State and Local Governments—Revenue and Expenditures by Function: 2007 and 2008," *Statistical Abstract of the United States, 2012*, Table 436).

the 31 percent marginal tax bracket (so I pay 31 cents of each additional dollar of taxable income to the federal government), then by deducting $1000, my federal taxes are reduced by $310. The net cost to me of $1000 on public education is thus only $690. In contrast, if I spend $1000 on private education, it costs me $1000 that I could have spent elsewhere.[9]

The total value of tax expenditures specifically for education in 2011 was estimated to be approximately $48 billion, as Table 14.1 shows. Most of these expenditures came from the reduction of taxable income because of a variety of deductions, exclusions, and exemptions. Because the value of these income adjustments is greatest for higher-income individuals, this form of support for education is, in effect, *regressive*; that is, it benefits higher-income individuals and higher-income communities more than lower-income ones.

At the same time, the tax system serves to discourage private expenditures on education.[10] If I spend $1000 in tuition to send my child to a

---

[9] This example assumes that I itemize my deductions rather than claiming the standard deduction. Only individuals with large deductible expenses such as interest on a home mortgage itemize. See Chapter 22 for a more detailed discussion.

[10] However, the tax deductibility of gifts to private schools, which are treated as charities, strongly encourages private education.

private school, not only do I pass up the already-paid-for public education, but the expenditure is not even tax deductible. Later, the return to the education, which occurs in the form of higher wages, is taxed. The tax system thus discourages private spending on education.

WHY IS EDUCATION PUBLICLY PROVIDED AND PUBLICLY FINANCED?

In the United States, the public role in the provision of education has been so pervasive that it is generally taken for granted. In some other countries, however, although the government may provide funds to educational institutions, much of the education itself is provided by private, particularly religious, schools.

IS THERE A MARKET FAILURE?

Education is not a pure public good. The marginal cost of educating an additional child is far from zero; indeed, the marginal and average costs are (at least for large school districts) approximately the same. Furthermore, there is no technical difficulty in charging individuals for use of this service.

Those who seek to justify public education in terms of a market failure focus on the importance of externalities. It is often claimed, for instance, that there are important externalities associated with having an educated citizenry. A society in which everyone can read functions far more smoothly than a society in which few can read. There is also a large private return to being able to read, and even in the absence of government support, many individuals would learn this and other basic skills. Indeed, most individuals would go far beyond that. The question is: Given the level of education that individuals would choose to undertake privately if there were no government subsidy, would further increases in education generate any significant externalities? There is no agreement concerning the answer, but the case for government support based on these kinds of externalities—at least for advanced industrial economies like the United States—seems, at best, unproved.[11]

---

[11] See, for instance, D. M. Windham, "Economic Analysis and the Public Support of Higher Education: The Divergence of Theory and Policy," in *Economic Dimensions of Education*, A Report of a Committee of the National Academy of Education, May 1979.

There may be other important externalities associated with education. Public education may have played an important role in integrating new immigrant groups into American culture. Public education may have been essential in making the melting pot work. The benefits of this accrued not only to individuals, but also to the nation as a whole.

There is another indirect form of externality to investments in science and technology education: people with these scarce skills are the key to technological progress and, typically, innovators capture only a fraction of their overall contribution to the increase in productivity.

However, there may be a reason why individuals do not invest as much in education as they would like, even to the point at which the private return equals the cost of capital: they may lack access to funds to finance their education. Private lenders are not, for the most part, willing to lend to finance education; understandably, banks are concerned about getting repaid. Hence, those lacking funds of their own (or funds from their families) would be denied access to education without some assistance from the government.

There is another market failure: investments in education are very risky. Individuals do not really know what the payoff is going to be—but there is no way that individuals can divest themselves of this risk. Many individuals will be reluctant to invest in their education, even if the *average* returns exceed the cost of capital, simply because it would impose too much risk. Matters have become worse in recent years. It used to be that individuals whose income turned out low could discharge their debts through bankruptcy—and they can for debts other than student debts. Since 2005, however, it has become extremely hard for students to discharge their debts. Therefore, even if the school they attend does not prepare them for a better paying job, they are saddled with paying the loan.

There is still another reason why there may be underinvestment in education. At the elementary and secondary levels, parents make decisions on behalf of their children. Although most parents may view expenditures on their children altruistically, investing in their children as long as the returns are sufficiently high, some parents may not do so. Under a system of privately financed education, the children of such parents might receive an insufficient education. Thus, a rationale for public support of primary and secondary education is provided on distributional grounds: there is a widespread belief that children's access to education should not depend on their parents' financial ability or their sense of altruism. Indeed, such distributional concerns provide the strongest motivation for a public role in education more generally—for the financing of elementary and secondary education and for the extensive financial support of higher education.

Although these arguments provide a rationale for government financing of education, they do not provide an argument for government

production. Why is it that at the elementary and secondary levels, government production (public schools) dominates, whereas at the college level, private not-for-profit education plays a far more important role? In many other countries, private elementary and secondary schools are more dominant than they are in the United States. We examine the controversy over private schools later in this chapter.

THE FEDERAL ROLE

The arguments discussed earlier explain why *some* governmental entity should provide financing for education, but they do not explain the level—local, state, or federal—at which the support should be provided. Increasingly, many arguments for a public role in education are arguments for a federal role. Largely this is because to the extent that there are externalities, given the huge migrations that occur throughout the country, those externalities need to be addressed at the federal level. Inadequate education is associated with a variety of social problems and antisocial behaviors. Those who grow up in one state and receive an inadequate education there have a good chance of practicing their antisocial behaviors in another state.

Moreover, distribution concerns can be addressed fully only at the national level, given the huge disparities in per capita incomes (and corresponding disparities in education expenditures per pupil) across states.[12] States that spend less on education per pupil tend to generate lower performance, as measured by earnings.[13]

ISSUES AND CONTROVERSIES IN EDUCATIONAL POLICY

Education has long been a source of political and intellectual debate. There are different perspectives on the basic impact of education on students, how education should be produced and financed, and whether the

---

[12] For instance, the state with the highest per capita income (Connecticut) had a per capita income almost twice that of the lowest (Mississippi) in 2008; total education expenditures per elementary and secondary pupil in Connecticut were also almost twice that of Mississippi.
[13] D. Card and A. B. Krueger, *Wages, School Quality, and Employment Demand* (New York: Oxford University Press, 2011).

current system results in too much inequality. The following sections take up some of these issues and controversies.

EDUCATION OUTCOMES

There are different perspectives on the outcomes of the education process. One view, implicit in the foregoing discussion, holds that education increases the skills of individuals, and thereby wages. Called the **human capital** view, this perspective sees investment in people as akin to capital investment. The greater the investment, the greater the productivity.[14] However, there is controversy over why, and to what extent the higher wages earned by skilled workers reflect an increase in their productivity resulting from education.

One variant of the human capital view, emphasized by Samuel Bowles and Herb Gintis of the University of Massachusetts, focuses on the socializing role of education. Education teaches people how to perform well in the workplace by teaching how to obey orders, follow directions, and work in teams. When successful, this socialization teaches punctuality and reliability. In this perspective, those who go to school longer learn more of these social skills, or, in any case, have demonstrated a greater ability or willingness to cope with the demands of the school system. These social abilities (or drives) make these individuals more valuable in the workplace.[15]

Another view of education, called the **screening** view, argues that one of the important functions of education is to identify the abilities of different individuals. Those who go to school longer receive higher wages and are observed to be more productive. This is not because the schools have increased their productivity, but rather because the schools have identified those individuals who are the most productive, or who have the necessary drive and ambition. The school system is viewed as a screening device, separating the very able and highly motivated from the less able and less motivated.[16]

---

[14] See G. Becker, *Human Capital: A Theoretical and Empirical Analysis with Special References to Education,* 2d ed. (New York: National Bureau of Economic Research, Columbia University Press, 1975).

[15] A. Weiss, "Human Capital versus Signaling Explanations of Wages," *Journal of Economic Perspectives* 9, no. 4 (Fall 1995): 133–154. Weiss closely studied low-skilled workers in a manufacturing plant. Long-run success depended not on any particular skill but on social characteristics such as reliability (low levels of absenteeism) and punctuality.

[16] This view has been put forward by J. E. Stiglitz, "The Theory of Screening Education and the Distribution of Income," *American Economic Review* 65 (1975): 283–300; A. Michael Spence, "Job Market Signaling," *Quarterly Journal of Economics* 87 (1973): 355–374; and K. J. Arrow, "Higher Education as a Filter," *Journal of Public Economics* 2 (1973): 193–216.

In the screening perspective, the *social* returns to education are far less than the *private* returns. The private returns can be significant: those who go to college receive a substantially higher income than those who do not. If all that was going on were the identification of who was more able and who less, though, the total production of society would be unaffected; the social returns would be zero. In fact, education also identifies differences in skills, enabling a better matching of individuals and jobs, and this does increase overall productivity: there can be significant social returns to screening. There is general agreement today that some of the returns to education are the result of increases in skills and some are a result of screening; the controversy arises over their relative importance.[17]

DO EXPENDITURES MATTER?

The significance of the ways in which education differs from other commodities is posed most starkly by the controversy over whether increased educational expenditures lead to increased education performance. In the production of a standard commodity, an increase in input would necessarily lead to an increase in output. Earlier, we noted the low performance of American students compared to those abroad, as indicated by certain test scores. Figure 14.2 shows clearly that this is not because the United States is spending less—in fact, on a per-pupil basis, United States is spending more than almost any other country. Similarly, expenditures per pupil have increased markedly over the past fifteen years, with only a modicum of effect on test scores. The weak link between expenditures and performance had been demonstrated earlier, in a classic study by James Coleman of the University of Chicago and his co-authors.[18] Coleman's study spawned a huge body of literature, with his results still in dispute.[19]

---

[17] For instance, studies showing that wages do not depend closely on the subjects studied suggest that content (skill formation) matters little. Moreover, if skill formation were predominant, one would expect that the increase in returns would be a relatively steady process, so the increment in wages per year would be the same; in fact, though, the increase in wages from the completion of the fourth year of college is much greater than the increase in wages from the completion of the preceding three years.

[18] J. S. Coleman et al., *Equality of Educational Opportunity* (Washington, DC: Department of Health, Education, and Welfare, 1966).

[19] See the survey by Eric A. Hanushek, "School Resources and Student Performance," in *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success,* ed. G. Burtless (Washington, DC: Brookings Institution, 1996), pp. 43–73; and Hanushek's earlier article, "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature* 24, no. 3 (September 1986): 1141–1177. See also H. Wenglinsky, *When Money Matters: How Educational Expenditures Improve Student Performance and How They Don't* (Princeton, NJ: Educational Testing Service, Policy Information Center, 1997); A. B. Krueger, "Experimental Estimates of Educational Production Functions," *Quarterly Journal of Economics* 113 (May 1999): 497–532; and E. N. Wolff, W. Baumol, and A. N. Saini, "A Comparative Analysis of Education Costs and Outcomes: The United States vs. Other OECD Countries," Working Paper, June 2013.
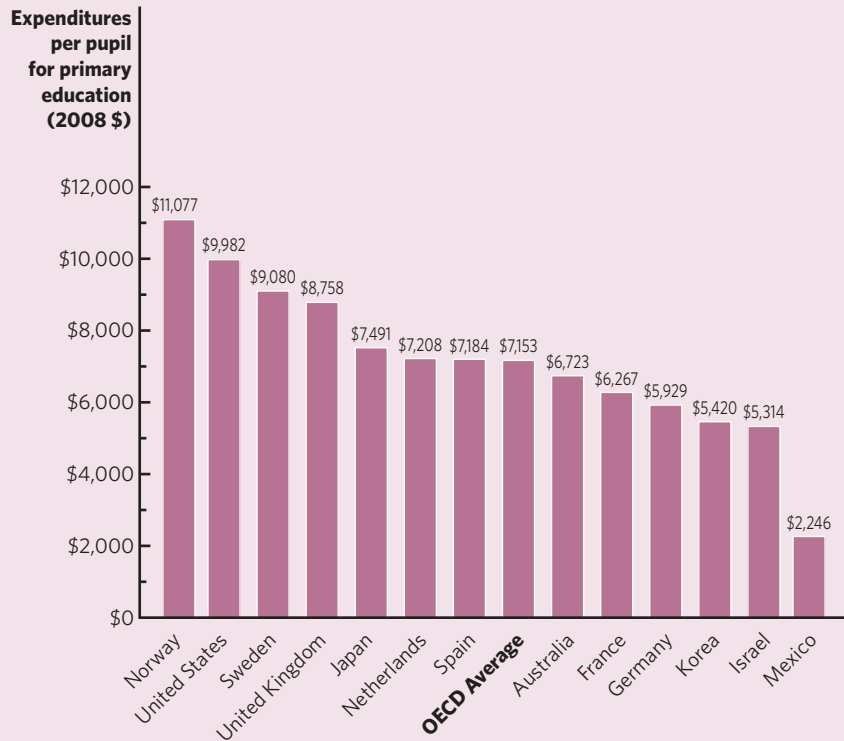
Many economists believe that increased expenditures do make a difference—a greater difference than Coleman's earlier study suggested. A set of studies that looked at the performance of identical twins (so natural differences were fully accounted for) showed that levels of education made a systematic difference in earnings. Another study compared states with different rules for when students could drop out of school; those who were forced to stay in school longer (controlling for other variables) seemed to do better.

The problem with interpreting the consequences of differences in expenditure levels is that much—and an increasing proportion—of education expenditures in the United States goes to purposes not directly related to teaching (such as administrative expenses) and to addressing the requirements of those with special needs. There are only limited data looking at, for instance, the consequences of smaller class sizes. The most famous experiment occurred in Tennessee, where students were randomly assigned to different class sizes; the results of that experiment,

though not conclusive, support the view that smaller class size (entailing greater expenditures) does lead to improved student performance.[20]

Another set of critiques of Coleman's findings argues that the real output of educational expenditures should not be higher test scores but higher productivity, leading to higher wages. Several studies have established a clear link between expenditures and earnings.[21]

Those who believe that schools have relatively little impact on earnings believe that home background is critical. Even in that view, it is still possible that by increasing expenditures on the disadvantaged, public schools can help offset deficiencies in home background. This raises fundamental questions about the allocation of resources *within* schools: How much should go to helping the academically gifted, the average students, and those at the bottom? The country's technological leadership depends on having the best scientists in the world, and this argues for putting resources at the disposal of the scientifically gifted. On the other hand, without adequate skills, those at the bottom will see their wages fall behind, as they have been doing in the last two decades (see Figure 14.3). Increasing inequality is likely to give rise to increasing social problems in the decades ahead. Education advocates argue that we should spend more on both, but given the current overall limitations on expenditures, the issue is: Should we direct more to the top or the bottom, or do we now have just about the right balance? The appendix to this chapter presents a framework for thinking about this issue.

SCHOOL VOUCHERS: CHOICE AND COMPETITION

Perhaps the most heated recent debate in the political sphere has concerned the question of school choice. Should parents be given more choice about where their children go to school? The simplest proposals for providing school choice entail **school vouchers**: each child would be given a

---

[20] See E. Word et al., "Student/Teacher Achievement Ratio (STAR)—Tennessee's K-3 Class Size Study," Tennessee Department of Education, 1994. This study corroborates other findings. In an analysis of many studies, Gene Glass concluded that "small classes were very much better than large classes; large classes were hardly any better than very large classes." See G. V. Glass, *School Class Size: Research and Policy* (Beverly Hills, CA: Sage Publications, 1982), p. 47. In a review of eleven studies, Robert Slavin found that the largest (but still modest) effects of class size on test performance were observed in the study with the largest reduction in class size (from twenty-three to fifteen). See R. E. Slavin, "School and Classroom Organization in Beginning Reading," in *Preventing Early School Failure: Research, Policy, and Practice,* ed. R. E. Slavin et al. (Boston: Allyn and Bacon, 1994), pp. 122–130. See also K. Akerhielm, "Does Class Size Matter?" *Economics of Education Review* 13, no. 3 (September 1995): 229–241; D. Card and A. Krueger, "School Resources and Student Outcomes," *Annals of the American Academy of Political and Social Science* 559 (September 1998): 39–53; and G. J. Whitehurst and M. M. Chingos, *Class Size: What Research Says and What It Means for State Policy* (Washington, DC: Brookings Institution, May 2011).

[21] See, for example, C. R. Belfield and H. M. Levin, eds., *The Price We Pay: Economic and Social Consequences of Inadequate Education* (Washington, DC: Brookings Institution, 2007).

FIGURE 14.3

DIFFERENCES IN INCOME ACROSS LEVELS OF EDUCATION

These charts plot median incomes of men and women with higher levels of education relative to high school dropouts. Since 1990, the earnings of minimally educated men and women have fallen behind their better-educated peers.
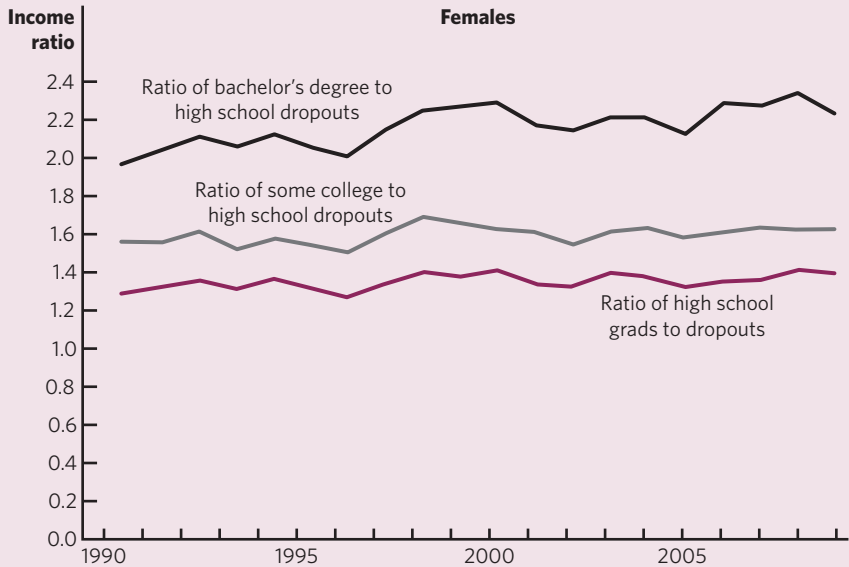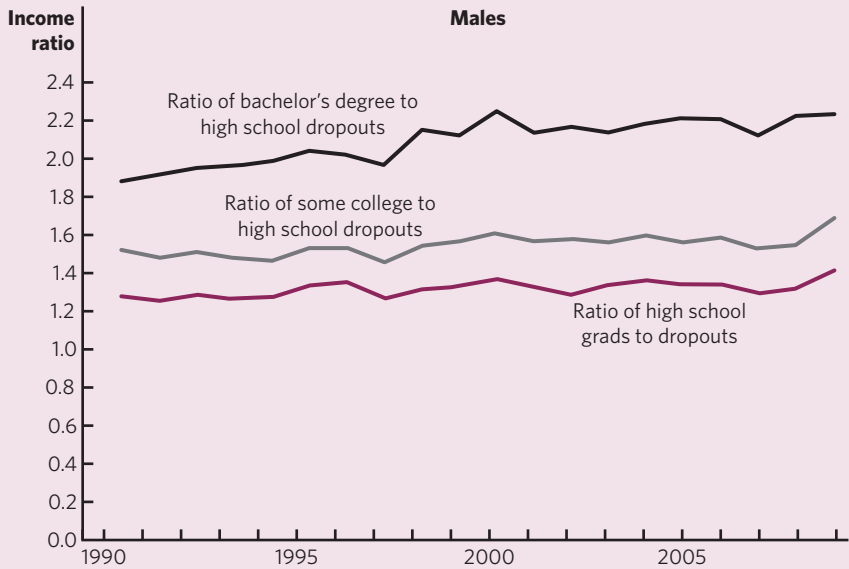
coupon—worth, say, $5000—to be used at the school of the parent's choice. Public schools would, under this proposal, have to compete directly with private schools; they would have to raise their revenue by persuading students to attend, just as private schools do. If parents valued the kinds of programs provided by the public schools, then the public schools would do well.

Implicit in much of the discussion favoring vouchers is a critique of the role of government in education: today it both finances education and produces it. Many believe, as we saw in Chapter 8, that, for a variety of reasons, the government is not an efficient producer, and that in the competition between schools, private schools would win out. Moreover, they say, when parents choose their children's school, they become more committed to that school and more involved in their children's education, and this contributes to school performance.

Interestingly, and contrary to the claims of critics, advocates of choice argue that private schools not only produce a higher-quality education at lower costs, but actually promote equality. In another famous study, James Coleman and his co-authors argued that America's public schools were actually more segregated, both racially and socioeconomically, than were its private schools. The segregation resulted because the public school system, especially at the primary level, is based on neighborhood schools, and neighborhoods are, in effect, segregated. They also argue that private schools are more effective in educating disadvantaged students (for instance, as measured by test scores).[22]

Finally, advocates of vouchers argue that America already has a system of choice, but constrained choice. Those with enough resources can choose to move to the suburbs or put their children in private school. It is the poor whom the current system deprives of choice. There is competition, but only for the children of the affluent, not for the poor. America believes in competition. It may be natural for those in any industry to argue that competition in their sector would be less desirable, and the postulates of perfect competition may not be satisfied perfectly. Still, as a country, we believe that the advantages of competition outweigh the disadvantages. Why should education be any different?

Voucher programs have been attacked on several grounds. The conditions that make competition work in conventional markets do not exist in education. Parents, particularly less well educated parents, often are not well informed; they are ill prepared to judge the effectiveness of schools in providing key skills. Indeed, this information problem is a key rationale for the establishment of national standards.

---

[22] J. Coleman, T. Hoffer, and S. Kilgore, *Public and Private Schools: An Analysis of High School and Beyond* (Washington, DC: U.S. Department of Education, National Center for Educational Statistics, November 1981).

THE VOUCHER DEBATE

PRO	CON
Competition promotes efficiency.	Conditions that make competition work are not satisfied.
Choice promotes commitment.	Will result in more social and economic stratification.
Risk of economic stratification.	Evidence on greater efficiency is questionable: What will happen to problem children?

This problem is even true for trade schools and colleges. Many with poorer backgrounds have been taken advantage of by schools that do not provide a good education and that engage in deceptive advertising—schools where completion rates are low and job placements are poor. Government attempts to regulate these effectively for-profit schools, or even to limit government loan programs for such schools, have been beaten back. The schools are highly profitable, and it pays for these schools to lobby heavily and make substantial campaign contributions. These for-profit schools, and their ability to manipulate the political system for their own advantage—even at the cost of those that they are supposed to serve—have become a major national scandal.

Moreover, in many areas, the number of schools is limited; many parents are inclined to choose schools on the basis of convenience rather than educational excellence. Competition is further limited by the fact that once children are enrolled in a school, transferring is likely to be difficult; and the less standardized the curriculum, the more difficult adjustment will be.

Critics of choice argue, furthermore, that the alleged superior performance of private schools is a mirage. Part of the seemingly superior performance arises from a selection effect: those who choose to send their children to private schools are more committed to education, and it is this commitment, rather than what the schools do, that accounts for any measured differences in performance. Second, there is a discipline effect: private schools reject or expel those who are disruptive or who fail to perform adequately in other ways. Public schools cannot do this. Although it may be true that today performance in private schools exceeds that of public schools for individuals of seemingly similar backgrounds, this would not be the case if there were a major expansion of private schools, especially if they were required to take all students, including those with discipline problems.

Moreover, say many critics of vouchers, the scheme would lead to a more socially and economically stratified society, with children of wealthy and well-educated parents going to one set of schools and children of poor and less well-educated parents going to another. Although regulations to prevent racial discrimination might be easily enforced, regulations to ensure the absence of socioeconomic stratification would be difficult to implement.

Voucher schemes remain very much on the political agenda. In part, this reflects the widespread dissatisfaction with American elementary and secondary education—although curiously, whereas most parents are critical of the state of education overall, they believe their own children are receiving a good education. The concerns over quality are especially strong in America's inner cities, which have high dropout rates.

Those seeking to introduce innovation and more competition *within* public schools have looked to alternatives to vouchers. One is **contracting out**: the school district contracts outside managers to run their schools. Advocates claim that the competition between management teams will lead to higher performance. School districts can make more informed decisions concerning management teams than many parents can concerning schools. Several school districts have experimented with such arrangements, and despite sufficiently ambiguous evidence of significant cost savings or improved student achievement, the contracting-out business has continued to grow.[23]

Still another initiative involves **charter schools**, which are self-managed or managed directly by the children's parents, rather than by a larger school district. A variation of this in England is "academies," public schools removed from local council control and given more autonomy in their staffing and teaching methods. There are now a large number of charter schools throughout the United States, and parents of children attending them generally are highly supportive. Supporters of the charter schools hope that their innovations will be imitated elsewhere, and that they will thereby spur an overall improvement in educational quality. Critics are more skeptical, however. Currently, charter schools, like private schools, cater to those most committed to education and most desirous of an alternative to the current public school system. Without wider scope for school choice, however, how will their impact be felt on public education more broadly? Studies examining charter schools suggest that, on average, they have not improved performance, although the best schools have done very well.[24]

Other reforms emphasize the broader role of incentives, such as merit pay for teachers. How effective such incentives can be depends to a large extent on how well teacher performance can be measured. The problem is

---

[23] For an extensive discussion of contracting out principles and practices, see P. Hill, L. Pierce, and J. Guthrie, *Reinventing Public Education: How Contracting Can Transform America's Schools* (Chicago: University of Chicago Press, 1997). For an analysis of contracting out results, see P. Burch, "After the Fall: Educational Contracting in the USA and the Global Financial Crisis," *Journal of Education Policy* 25 (November 2010): 757–766.

[24] For assessments of charter school performance, see M. Berends and H. Walberg, *Charter School Outcomes* (Hillsdale, NJ: Lawrence Erlbaum Associates, 2008); R. Zimmer et al., *Charter Schools in Eight States: Effects on Achievement, Attainment, Integration, and Competition* (Santa Monica, CA: Rand Corporation, 2009); and S. Imberman, "Achievement and Behavior in Charter Schools: Drawing a More Complete Picture," *The Review of Economics and Statistics* 93 (May 2011): 416–435.

VOUCHERS: THE SAN JOSE AND MILWAUKEE EXPERIMENTS

Education vouchers would seem to be an ideal subject for experimentation: communities could make vouchers available and see what happens. In fact, a few communities have done just that. One of the earliest experiments, involving choice only among public schools, occurred in San Jose, California. One study concluded, after examining the results of that experiment, that

> education vouchers stand little chance of succeeding in American elementary and secondary schools. [Parents eventually] learn about their alternatives and the rules governing choice. But there is still some question about the social consequences of parents' or students' program choices. Specifically, there is some concern that parents pick programs which reinforce their class-related social values, so that poor children have little opportunity to acquire the beliefs, attitudes, and social competencies necessary for social mobility to the middle class.*

One of the few experiments to provide vouchers that could be used in public or private schools was begun in Milwaukee in 1990. Curiously, as the experiment proceeded, both supporters and critics of vouchers claimed that the Milwaukee experiment confirmed their views concerning the effectiveness of vouchers.

Milwaukee provided a limited number of vouchers that parents could use to send their children to the school of their choice. In 1990, the first year of the voucher program, the number of vouchers was limited to 1 percent of the public school enrollment (approximately 950 students); the program was expanded to 1.5 percent in 1994.

Critics pointed out that relatively few parents availed themselves of the opportunity. Between 1990 and 1993, fewer than 600 parents applied each year, with an average of under 400 selected and 350 actually enrolled. To set this number in perspective, Milwaukee public school enrollment was 95,000, of which 42 percent came from families below the poverty level; any family with an income of up to 1.75 times the poverty level was eligible to apply. Even more telling, of those who elected to go to private schools, a large fraction transferred back into public schools within a year or so. By 1994, only 92 of the 354 students enrolled in a choice school in 1990 remained in the program. There seemed to be little effective competition with—and therefore spur to—public schools, and little spur to the expansion or creation of new private schools.

Advocates pointed to the higher performance of those who did switch to private schools and remained there, compared to that of peers with similar backgrounds who had remained in public schools. Statisticians clashed over the extent to which these differences could be attributed to the special nature of the children who had elected to go to and remain with the private schools.†

*G. Bridge, "Citizen Choice in Public Services; Voucher System," in *Alternatives for Delivering Public Services,* ed. E. S. Savas (Boulder, CO: Westview Press, 1977).

†C. E. Rouse, "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program," *Quarterly Journal of Economics* 113, no. 2 (May 1998): 553–602; and J. Greene et al., *Expanding Choice in Elementary and Secondary Education: A Report on Rethinking the Federal Role in Education* (Washington, DC: Brown Center on Education Policy at Brookings, February 2010).

that it is difficult to assess the teacher's *value added*: that is, if a teacher has students that enter the class poorly prepared, educational attainment is likely to remain weak at the end of the year—even if the teacher did a good job. A second problem is that teachers are typically poorly paid. Many, perhaps most, enter the profession not because of the money but because they love teaching. Small pittances paid as part of incentives reminds them of how poorly paid they are. This may actually lead to poorer performance, and those who have opportunities elsewhere move into other better paid jobs. In many contexts, it has been shown that incentive pay systems backfire.[25]

Many of the critics of America's education system like to blame unions—but the role of unions is more ambiguous than the rhetoric often suggests. Some of the highly unionized states have far better outcomes than the less unionized states. Unions have fought hard to maintain public support for public education, and to ensure that there are sufficient funds—including, not surprisingly, funds to pay teachers adequately. Without their actions, salaries and working conditions might have been much worse, and thus school systems would not have been able to recruit good teachers, and education outcomes would have suffered. Still, in some places and in some areas, teachers unions have resisted changes and reforms that might have had positive benefits. What is clear is that without the buy-in of those who are responsible for education—the teachers—it is hard to see how one can get educational reforms that would result in substantial improvements in performance. That is why some suggest that the best approach is a "deal": more flexibility on the part of unions in return for better support for public education.

SCHOOL DECENTRALIZATION

A widely discussed school reform, school decentralization, aims to address problems associated with the control of schools by large educational bureaucracies.[26] Decentralization has been an issue in large urban school districts, where such bureaucracies have been seen as unresponsive to parents. The effect of decentralization is to shift authority to

---

[25] There is a large and growing literature discussing these reforms, as well as many field experiments testing different incentives. See, for instance, E. Hanushek and D. Jorgenson, eds., *Improving America's Schools, The Role of Incentives* (Washington, DC: National Research Council, 1996); D. Neal, "The Design of Performance Pay in Education," Chapter 6 in *Handbook of the Economics of Education*, vol. 4, ed. E. Hanushek, S. Machin, and L. Woessmann (Waltham, MA: North-Holland, 2011); M. Springer et al., *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT)* (Evanston, IL: Society for Research on Educational Effectiveness, 2011); and E. Hanushek and L. Woessmann, "Symposium on Performance Pay for Teachers," *Economics of Education Review* 30 (2011).

[26] For an extensive discussion of decentralization, see J. Hannaway and M. Carnoy, eds., *Decentralization and School Improvement* (San Francisco: Jossey-Bass, 1993).

individual schools and their principals, giving greater influence to parents. Advocates argue that school decentralization creates stronger incentives for parents to monitor teachers and schools, to the benefit of school performance. Teachers, too, the argument goes, have stronger incentives to perform better because they have greater influence in the educational process at their schools.

Criticisms of decentralization have been far more muted than those of the voucher programs; a number of important decentralization initiatives are already in place, including those in two of the country's largest cities: New York and Chicago. However, there is concern that unless union rules and attitudes are changed, little headway will be made in fundamental reform. Another worry is that if there is too much discretion at the level of schools, too much inequality will be generated: some schools will be far better than others, and this seems unfair. Supporters of decentralization point out that this, of course, is already true. Thus, any attempt to improve quality will entail some temporary inequality, until the poorer schools are lifted to the level of the better schools, but it makes little sense to keep the best schools down simply to avoid a disparity.

PERFORMANCE STANDARDS: NO CHILD LEFT BEHIND AND RACE TO THE TOP

Many believe that before decentralization, choice, or any other reform can be effective, clear performance measures for schools must be established. Without such measures, parents, teachers, and educational reformers cannot begin to determine whether schools are performing well or poorly, and therefore have no basis for educational reforms.

Establishing clear performance goals was the main objective of the No Child Left Behind (NCLB) Act of 2001, a bill to overhaul the Elementary and Secondary Education Act of 1965 as well as to replace the Goals 2000: Educate America Act of 1994. Under this legislation, states are required to conduct annual student assessments linked to state standards and participate in National Assessment of Educational Progress reading and mathematics tests in fourth and eighth grades, develop adequate yearly progress (AYP) standards and take specific steps when AYP is not met, and ensure that their teachers are "highly qualified." NCLB also provides grants for the development and implementation of standards-based assessments, curriculum development in priority subject areas, targeting of disadvantaged students, and the promotion of educational innovations.

Even though there is widespread support for the establishment of performance goals, the legislation has encountered considerable opposition,

for several reasons. Some voice concern that certain skills are easier to test than others. For example, there are good ways of testing basic skills, like phonics, but testing creativity and what are called "higher order" cognitive (thinking) skills is far more difficult. One of the strong points of American education has been how it has encouraged creativity. Will the tests divert teachers' attention toward teaching basic skills, and away from encouraging creativity and thinking skills? Some critics also claim that teachers are likely to gear their teaching in response to the tests ("teaching to the test"), and that high-stakes testing has led some schools to intentionally manipulate student test scores. (Several scandals have suggested that this has actually happened.)

On the other hand, advocates claim that our schools are failing to teach basic skills to many youngsters, and getting them to do that would be a major advance.

Still others are worried about potential inequalities. Schools with disadvantaged children worry that they will be criticized for their poor performance, and they are not persuaded by reassurances that in being judged, the background of their students will be taken into account. There is also a concern that children from disadvantaged backgrounds will likely perform relatively poorly on these standardized tests, and that employers will then use these tests as a simple and acceptable screen for determining whom to hire, putting already disadvantaged groups in an even worse position.

Most importantly, critics of NCLB argue that parents already know that schools are not performing as well as they would like. They believe that without more fundamental changes in the organization of schools, information itself will do little but increase parents' sense of helplessness.

Critics also say that the resources required to achieve the goals of "leaving no child behind" have not been provided—and are unlikely to be provided in the current era of budget stringency.

Congress is now debating whether to "end or mend" NCLB. This legislation has indeed compelled states to design and implement school accountability systems and take concrete action to improve performance based on annual student assessments. However, the results on student achievement thus far are mixed. NCLB has generated substantial and almost universal gains in mathematics, but has had virtually no positive effect on reading skills, and has generated only modest impacts among disadvantaged cohorts in math, thus making little progress in

ENHANCING THE EFFICIENCY OF SCHOOLS

1. Choice and competition: school vouchers, contracting out, charter schools
2. Incentives: merit pay
3. Reorganization: decentralization
4. Performance standards: No Child Left Behind and Race to the Top

closing achievement gaps. There is also concern that although NCLB has focused attention on measuring and improving educational performance for all students, its narrowly defined and inflexible parameters for success have created perverse incentives to do well, such as narrowing the curriculum, setting artificially low standards, and excluding good teachers who do not qualify under such restrictive assessment metrics.[27]

To address the shortcomings of NCLB, the Obama administration launched the Race to the Top program under the American Recovery and Reinvestment Act of 2009. Race to the Top is a $4.35 billion competitive grant program to assist states in implementing innovative education reform in four areas: upgrading of standards and assessments, improvement of data collection and utilization, enhancement of teacher effectiveness and equity in teacher distribution, and turnaround of lowest-performing schools.

INEQUALITY

One of the underlying purposes of public finance and provision of education is to promote equality of opportunity. However, there is a considerable concern that today the public school system does not come close to living up to that ideal. This is reflected in recent statistics showing that the United States has become the country with the least (or at least one of the worst) equality of opportunity among the advanced industrial countries—that is, a child's lifetime prospects are more dependent on the income and education of his or her parents than in other similar countries.[28]

The major reason for the shortfall is not hard to find. Education has remained primarily a local responsibility. The resources available in some communities, such as wealthy suburban communities or towns in which there is a large factory, are a multiple of those available in others. Demands for public services in general are greater in some communities than in others: urban communities typically require far higher expenditures for maintaining public safety than do rural communities. Finally, different communities have expressed different "preferences" for education. Communities with the same resources may spend different amounts.

---

[27] For a rigorous analysis of the academic impact of NCLB, see T. Dee and B. Jacob, "The Impact of No Child Left Behind on Student Achievement," National Bureau of Economic Research Working Paper 15531, November 2009. For a summary of the NCLB reauthorization policy debate, see R. R. Skinner, "The No Child Left Behind Act: An Overview of Reauthorization Issues for the 111th Congress," Congressional Research Service Report RL33749, May 2009.

[28] See J. E. Stiglitz, *The Price of Inequality: How Our Divided Society Endangers Our Future* (New York: W. W. Norton & Company, 2012).

Whatever the reasons, large differences in support for public education can have significant consequences. Those lucky enough (or rich enough) to live in communities that provide good schools have better prospects than those who live in communities that do not, and because America is becoming more segregated on the basis of income—with rich families increasingly living in communities in which other rich people live—this means that educational outcomes, and lifetime prospects, are increasingly differentiated.

The system of local funding has been tested in the nation's courts. A number of state supreme courts have found that relying on local property taxes to finance education violates provisions of state constitutions ensuring equal access to public education. At the federal level, the U.S. Supreme Court ruled, in *San Antonio Independent School District v. Rodriguez*, 411 U.S. 1 (1973), that local funding in Texas did not violate the "equal protection clause" of the U.S. Constitution (the Fourteenth Amendment) even though it resulted in large variability in expenditures.

Such decisions raise several important questions. Should spending be required to be the same in every community within a state? Or should the provision of a minimal level of education in every community be the only requirement? Mandating equality would seem to preclude communities' spending more on their children. Several states have, in fact, put caps on the levels of expenditures, thereby attaining equality not only by raising the minimum level but also by lowering the maximum. If only a minimum standard is required, how is it to be determined? Obviously, if the standard is set low enough, it will have no effect at all. If it is set too high, how will poorer communities find the funds? Higher standards necessitate a greater role for the state in financing education. If equality is insisted upon, what adjustments should be made for differences in the costs of education in different communities, and in the nature of the student bodies? Is equality of spending enough? Some communities might use more of their funds to develop better athletic facilities, others for basic-skills development, still others for special educational programs. The result is different treatment of similar individuals who happen to reside in different communities. Ensuring complete equality would require eliminating community control and establishing a centralized educational system within each state.

At issue are not only the basic trade-offs of equity and efficiency; most people, to some degree, believe that all parents should have the right to make decisions concerning their children's education. With local control of education, parents at least feel that they can have some influence over the outcomes. Thus, local autonomy of schools has become for many a principle in its own right; some individuals might still favor local control even if

APPROACHES TO ENHANCING EQUALITY

- Provide minimum level of expenditures for all districts.
- Cap expenditures of high-spending districts.
- Federal and state financial assistance to low income communities and communities with large numbers of disadvantaged children.
- Key issue: how to maintain local responsibility and autonomy while providing equality of opportunity.

it could be shown that central control was both more equitable and more efficient. To others, the trade-off at issue is between the rights of parents (to decide about their children's education) and the rights of children (to equality of opportunity, regardless of who their parents are).

Both state and federal governments have responded to these concerns by attempting to increase equality while retaining at least some degree of local autonomy, especially by providing funds to communities with low incomes and with many disadvantaged children. New Jersey, for instance, has broadened its role in financing education, with poorer districts getting much more aid than better-off districts; set minimal educational standards that all districts must attain; and, in some cases, placed ceilings on the amount that higher-spending school districts can spend.

LIMITATIONS ON EQUALITY IMPOSED BY PARENTAL CHOICE

Some attempts to attain greater equality, such as the ceilings on expenditures by richer school districts, have been criticized as misguided and self-defeating, as the underlying problem is not just the degree of inequality within public schools but also the total extent of inequality in our society. As long as the government is not willing to prohibit individuals from going to private schools, any attempt to introduce too much equality into the public educational system will result in individuals' transferring to private schools.

In England, there have been periodic proposals for the government to actively discourage private schools on the grounds that private education leads to social stratification (only the upper and upper-middle classes send their children to private schools). In the United States, however, restrictions on private schools might well be unconstitutional. In any case, although there has been controversy about whether private schools should receive public support, no one has suggested that private schools be actively discouraged.

AID TO HIGHER EDUCATION

Interestingly, many of the controversies that plague elementary and secondary education have been resolved at the higher-education level. There has even been a voucher plan: the GI Bill of Rights, which spurred the

growth of higher education in the United States after World War II, in effect gave veterans a voucher they could use at any university, public or private. Private schools, as we have already noted, play a large role at the college and university level, and there is a high level of decentralization.

The federal role dates from the setting aside of land for a university in the Land Ordinance of 1785, to the founding of the land grant universities in 1862, to the extensive support of research universities through grants and contracts in the post–World War II era.

A fundamental difference between higher education and elementary and secondary education is that the students have reached an age at which they can make decisions for themselves. It is the students who should judge whether the returns to further education warrant further investment. At least that would be the case if individuals had good information, and had the skills required to assess whatever information is presented. However, many private for-profit schools do not provide the kind of information that individuals would need to make an informed decision—and it is precisely the skills in assessing such information that a college education is intended to hone. In many other areas, we have learned that there is a need for consumer protection, and this seems one for which the need, given the evidence of market exploitation, is especially strong.

Beyond this, government has a key role to ensure access, so students have the financial resources to go on to college. Currently, this is accomplished in five ways. First, states greatly subsidize higher education, typically charging tuitions in their public colleges and universities that are a fraction of the total costs. Second, the federal government provides grants, such as Pell and TEACH grants, for students who meet eligibility criteria based on financial need. Third, the federal government sponsors loan programs for higher education, such as Perkins and Ford loans. Fourth, it supports a limited work-study program, called the Federal Work-Study Program. Fifth, it provides tuition tax deductions and credits.[29]

The state subsidies have been widely criticized as being untargeted (or mistargeted). Indeed, because enrollment rates typically are higher among the children of the affluent, on average, it is the more affluent who benefit from such subsidies. Moreover, to the extent that the subsidies can be thought of as directed at the children themselves, rather than at the parents, lifetime incomes, even of children from poorer families, will be higher than the average income of the population. Therefore, critics say, government should focus its attention on loans or grants to children from poor families.

Federal tuition tax credits were introduced in 1997 (effective in 1998). There are now two federal tax credits to help pay higher education costs:

---

[29] For a summary of federal student aid programs, see U.S. Department of Education, *Funding Your Education: The Guide to Federal Student Aid, 2012–13.*

the American Opportunity Credit, up to $2500 per eligible student annually for the first four years of an undergraduate degree program, and the Lifetime Learning Credit, up to $2000 per eligible student annually for all years of post-secondary education, including nondegree and skills improvement courses. These, too, have been criticized as untargeted: the money will not go to the very poor (although 40 percent of the American Opportunity Credit is refundable, so even those who do not pay taxes can receive up to $1000). Critics worry that the tuition tax credits may simply induce colleges to raise tuitions, especially in states that charge very low tuition for their community colleges.[30] They worry that enrollment rates of the very poor may even go down, as poor people who cannot take advantage of the tuition tax credit find college less affordable, unless states increase scholarship funds to offset any tuition increases. (Proponents argue that even if tuitions increase, the increased availability of funds for education will be helpful.)

Critics of the tax deduction and credit program argued that it would have cost far less and been far more effective in increasing enrollment rates, particularly among the disadvantaged, if the government had announced a broader policy of guaranteed financial access, a combination of loans, grants, and work to ensure all Americans access to higher education.

Federally guaranteed loans were first authorized as part of the Higher Education Act of 1965. Although today there is little controversy about the value of student loans, there has been recent debate over how best to provide them. The federal government now has two major loan programs: Federal Perkins Loans, in which private institutions lend to students while the government bears the risk of default; and William D. Ford Federal Direct Loans, in which the federal government lends directly to students. The latter began in 1994 under the Clinton administration, which claimed that the federal government could administer the program more effectively and at lower cost than private sector institutions—a claim that the subsequent years bore out. Part of the rationale for the program is that the division of risk bearing and administration under the existing program gave rise to insufficient incentives for screening and debt enforcement (collection) on the part of private institutions. Equally troublesome, the private financial sector was charging students *as if* they were bearing the risks, when they were not. It became a lucrative source of finance—so lucrative that there were scandals involving bribes to colleges to use particular lenders. Moreover, the government would be able to introduce more flexible loan instruments, which indeed it has done,

---

[30] The credit is 100 percent of the first $1000 in tuition and 50 percent of the second $1000. Raising tuition from $500 to $1000 will thus cost the student nothing—what the student pays will be unchanged, although what the federal government pays will increase substantially.

including repayment plans that are graduated (stepwise increase in payment size over the loan term), extended (up to 25 years rather than the standard 10 years), income-based (maximum monthly payment is 15 percent of discretionary income), pay as you earn (maximum monthly payment is 10 percent of discretionary income), income contingent (payments based on a capacity-to-pay formula), and income sensitive (payments based on annual income). Borrowers who work in the public sector for a decade might also be eligible for the Public Service Loan Forgiveness Program after making 120 full, on-time loan payments.[31] Since 1994, the federal government has become the major lender, issuing 93 percent of all student loans in academic year 2011–2012.

As the cost of higher education continues to increase while both state aid for education and median household income continue to decline, student loan debt has grown dramatically. This trend has persisted throughout the Great Recession, despite a substantial reduction in other types of consumer debt; at the end of 2012 it was approximately $1 trillion, far exceeding total household credit card debt. Not only is the magnitude of student loan debt worrisome, but the rise in arrears is also troublesome—at over 10 percent, student loan balances 90 or more days delinquent for the first time exceed the "serious delinquency" rate for credit card debt. These problems are exacerbated by U.S. bankruptcy laws. Unlike most other types of consumer debt, student loans are extremely hard to discharge.

With incomes of most Americans stagnating and tuition costs soaring, America faces a bind: if current trends continue, young Americans (and their families) will face increasing debts to finance their education. This could have severe macroeconomic consequences, just as other aspects of household debt did, in the run up to the global financial crisis. However, it will also mean that many Americans will decide that it is just too risky and will not get the education they need to be competitive in the twenty-first century. This will obviously harm their own economic prospects, but if (as seems likely) enough people make that decision, it could affect America's overall economic prospects.

One way out of the dilemma is further government support to higher education, which would enable a lowering of tuition, or at least grants to more individuals so they could take on less debt. With budget cutbacks at both the state and national levels, however, the prospects of this are bleak.

Since the late 1980s, several countries have adopted **income-contingent loans (ICLs)** as a possible solution; the United States introduced ICLs in 1994. The key distinguishing characteristic of ICLs is that collection of the debt depends on the borrower's capacity to pay.

---

[31] For a detailed explanation of these federal student loan products, see http://studentaid.ed.gov/repay-loans/understand/plans.

For example, in 1989 Australia introduced a risk-sharing ICL, the Higher Education Contribution Scheme (HECS). Under this program, student debt is repaid through the tax system, with the amount due dependent on personal income: there is no repayment obligation unless income exceeds a certain threshold, and then the repayment amount increases as income rises.[32] The program not only reduces the risks associated with investing in education, but it also significantly lowers transactions costs. It has been an important ingredient in Australia's successful attempt to extend access to a college education: the percentage of Australia's population 25 to 64 years old who have graduated from college rose from 18 percent in 1999 to 27 percent in 2010. This brought Australia closer to the U.S. graduation rate. Although the United States started from a better position of 28 percent in 1999, the U.S. graduation rate did not grow as quickly, reaching 32 percent in 2010.[33]

---

[32] For further information about Race to the Top, see the U.S. Department of Education website www .ed.gov. For more data on student loan debt, see the Federal Reserve Bank of New York's quarterly reports on household debt and credit. For a detailed explanation of ICLs, see B. Chapman, "Income Contingent Loans for Higher Education: International Reforms," Chapter 25 in *Handbook of the Economics of Education*, vol. 2, ed. E. Hanushek and F. Welch (Amsterdam: Elsevier, 2006), pp. 1435–1503. HELP is the Australian government loan program to assist students finance their tuition fees (FEE-HELP or VET FEE-HELP), student contributions (HECS-HELP), and overseas study expenses (OS-HELP). In 2012–2013, the HELP repayment threshold was A$49,096, above which the repayment rate rose in 0.5 percent increments from 4 to 8 percent of HRI. More information about HELP can be found at http://studyassist.gov.au and www.ato.gov.au.
[33] Organization for Economic Cooperation and Development, *Education at a Glance, 2012*.

REVIEW AND PRACTICE

SUMMARY

1. The past fifty years have seen changes in the structure of education in the United States, including an increase in the fraction of funds provided by states and the federal government. Although direct federal expenditures for education are low relative to state and local spending, federal tax expenditures provide large implicit subsidies. In addition, federal involvement in guiding education policy is growing.

2. Education is not a pure public good. Important externalities are associated with education, but these externalities do not provide a fully persuasive justification for the role of the government. The major justification for public support of elementary and secondary education is the belief that the quality of education obtained should not depend solely on the resources of the child's parents. There are important market failures, including imperfections of capital and risk markets.

3. There are concerns about both the efficiency and equity of the U.S. educational system. Proposals to improve the efficiency of the system are (a) a voucher program to enable more choice and promote competition; (b) decentralization reforms, to enable teachers and parents to exert more control over schools; (c) charter schools to provide more

scope for innovation; (d) contracting out, to provide more competition for school management, without the disadvantage of social stratification that may arise from voucher plans; and (e) national standards with clearer goals and performance measures, both for schools and teachers. Most of these "experiments" have had, at best, mixed results.

4. The heavy reliance on local finance for schools results in a great deal of inequality in expenditures. Reforms include providing more state aid.

5. Although education is not the only determinant of an individual's future wages, there is a systematic correlation between the level of education and wages. There is controversy, however, concerning the explanation of this correlation. Some claim that it is primarily due to the increased skills that children obtain at school (the human capital view), whereas others claim that it is due to the schools' identifying the very able and differentiating them from the less able (the screening view).

6. The poor performance of American students on standardized international exams in spite of heavy expenditures is related in part to the high level of inequality in the United States. In addition, large amounts of expenditures go to administration and special programs. The overall link between expenditures and performance has been hotly debated.

7. The government has long played an active role in higher education, although its dominance is not as great as at the elementary and secondary school levels. Some believe that the effects of government aid to higher education are regressive, as those who benefit from college are likely to have higher incomes. They believe that direct subsidies should be replaced by loan programs. With tuitions soaring and incomes of most Americans stagnating, however, student debt is already soaring. Australia has pioneered an alternative with low transactions costs called income-contingent loans.

8. The federal government has long played an important role in supporting higher education. Student loans and federal grant programs have increased accessibility to higher education, but stagnating incomes in the middle and rising tuitions now threaten that accessibility. There is considerable controversy over the extent to which tuition tax credits will increase enrollments. Most agree, however, that such tax credits are not well targeted.

KEY CONCEPTS

Charter schools

Compensatory education

Contracting out

Human capital

Income-contingent loans (ICLs)

School vouchers

Screening

QUESTIONS AND PROBLEMS

1. **a.** Discuss the equity and efficiency arguments for raising tuition at state universities. To what extent do your answers depend on whether there is a good college loan program available?

   b. Discuss the equity and efficiency arguments for providing college loans at subsidized interest rates.

2. List some characteristics of our educational institutions or outcomes of our educational system that seem to be more consistent with the screening view of education than with the human capital view.

3. The property tax base per student is often as high in industrial centers as it is in the suburbs. Why might you still expect expenditure per pupil to be lower in industrial centers than in the suburbs?

4. Discuss the trade-offs involved in deciding on the appropriate level and form of decentralization/centralization within education. Bear in mind that different aspects—finance, control of curriculum, control of hiring—can be decentralized to different extents, and that the issues of centralization and decentralization relate not only to the division of responsibility among the federal government, the state, and the local community, but

also to the division of responsibility within the school district, among central office administrators, school principals, and teachers.

5. Discuss the trade-offs between parental choice and equality of opportunity. To what extent should the principle of consumer sovereignty extend to parental rights to choose the amount and form of education for their children?

6. Provide an economic analysis of the issue of tracking (putting students of similar abilities in the same classes). What evidence would you like to have to decide whether tracking is desirable? What are the trade-offs? (How does your answer depend on whether teachers find it easier to teach classes in which students have relatively similar abilities, and the extent to which less able students learn from having more able students inside the classroom?)

7. A community is considering how to allocate expenditures between education and other goods. Draw the budget constraint, putting "education" on the horizontal axis, and "other goods" on the vertical.

   a. Contrast the budget constraint of a poor community with that of a richer community. Using indifference curves, explain why a poorer community is likely to spend less on education than a richer community.

   b. Show how the deductibility of property taxes, used to finance public education, affects the budget constraints of the two communities. How does it affect choices? If the price elasticity of the demand for education is unity, what should be the effect on the demand for education in a community in which the median voter is in the 33 percent marginal tax bracket? Or in a 15 percentage tax bracket?

   c. Assume now that the poor community is subsidized (and the rich community taxed) so that the extra educational expenditures it gets from a 1 percent tax is the same as that of a rich community. Draw the budget constraints. Will the poorer community continue to spend less on education than the richer community? On what does your answer depend?

   d. Whose indifference curve is relevant for the analysis? (Recall the discussion of median voters from Chapter 9.)

8. List the various concerns about inadequacies of the U.S. educational system. Evaluate various reform initiatives in terms of the extent to which they address these concerns.

   a. If you thought that the primary problem was education of the disadvantaged, what reforms would you emphasize?

   b. If you thought that the primary problem was the lack of an adequate supply of highly trained scientists and engineers, what reforms would you emphasize?

   c. If you thought that the primary problem was inadequate performance of elementary and secondary schools, which reforms would you emphasize?

   d. If you thought that the primary problem was the lack of enrollment in universities, what reforms would you emphasize?

9. Assume that spending more on the education of a student increases his or her productivity. Draw the relationship between productivity/wages (assuming that wages increase in tandem with productivity and that there are diminishing returns). Assume that more able students have, for each level of education, a higher level of productivity.

   a. Assume that one wanted to maximize total output for a given level of expenditure. How would you allocate educational expenditures? Would you necessarily spend more on the more able students?

   b. Assume that you wanted to reduce inequalities in income. How would this affect patterns of educational expenditure?

APPENDIX: HOW SHOULD PUBLIC EDUCATIONAL FUNDS BE ALLOCATED?

Every school district faces the problem of allocating its educational budget. It can allocate more funds for special education, for remedial classes for the disadvantaged, or for accelerated classes for the gifted. As more funds are allocated to any one individual, there is some increase in that individual's productivity. This is the return to education.

If we wished to maximize national output, and if efficiency alone were our goal, we would allocate funds so the increase in productivity from spending an extra dollar on one individual would be the same as the increase in productivity from spending an extra dollar on another. If very able individuals not only reach a higher level of productivity than others at each level of education but also benefit more from education, so the *marginal* return to education is higher, such a policy entails spending a greater amount on the education of the able than on that of the less able. Some would say this is unfair; the government should ensure equality of expenditures in public education. However, when educational expenditures are equalized, those who are more able—or who have home backgrounds that give them an advantage—will still be better off. Accordingly, some believe that government should engage in **compensatory education**: it should attempt to equalize not *input* (expenditure), but *output* (achievement). It should attempt to compensate for the background disadvantages facing some groups in our society. One of the major federal programs is directed specifically at encouraging local communities to provide such compensatory education.

As more and more funds (of a fixed budget) are allocated to the less able, and less and less to the more able, total output falls, as the marginal return to education (under our assumption) for the less able is smaller than for the more able. Under these assumptions, there is a trade-off between efficiency and equality, as depicted in Figure 14.4A. What point one chooses on this locus depends on one's values, on how one is willing to trade off efficiency versus equality.

Some maintain, however, the trade-off curve does not look like Figure 14.4A but like Figure 14.4B—that is, some movement toward compensatory education may actually increase national output. In this view, those who are advantaged have a higher output than the disadvantaged at each educational level, but the *marginal* return to further education for the more able is actually lower than for the less advantaged. This implies that

EQUITY–EFFICIENCY TRADE-OFFS IN EDUCATION EXPENDITURES

(A) Poorer individuals with poorer home backgrounds have, at any level of educational expenditures, a lower marginal return to education. Accordingly, maximizing output requires spending more on richer individuals. By spending an equal amount, output is reduced but equality is increased. Compensatory education—equalizing productivity—requires spending even more on poorer individuals, reducing output but increasing equality. (B) Here, poorer individuals may have a lower level of income at any given level of educational expenditures, but a higher marginal return. Then output maximization (equalizing marginal returns) requires spending more on poorer individuals. Spending an equal amount on each will then lower equality and efficiency.

A

Equality (measured by some inequality index)

Compensatory education—equalized productivity

Equal expenditure

Output maximum

Efficiency (e.g., national output)

B

Equality

Compensatory education—equalized productivity

Output maximum

Equal expenditure
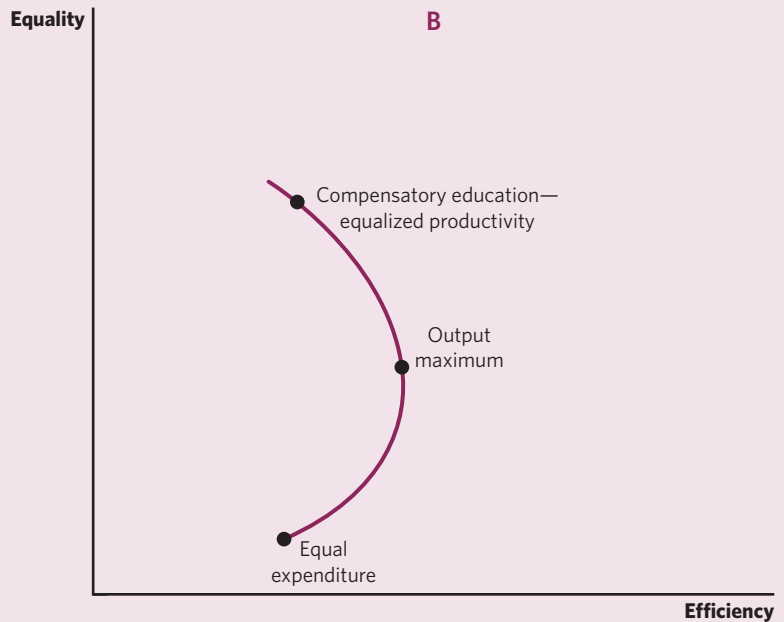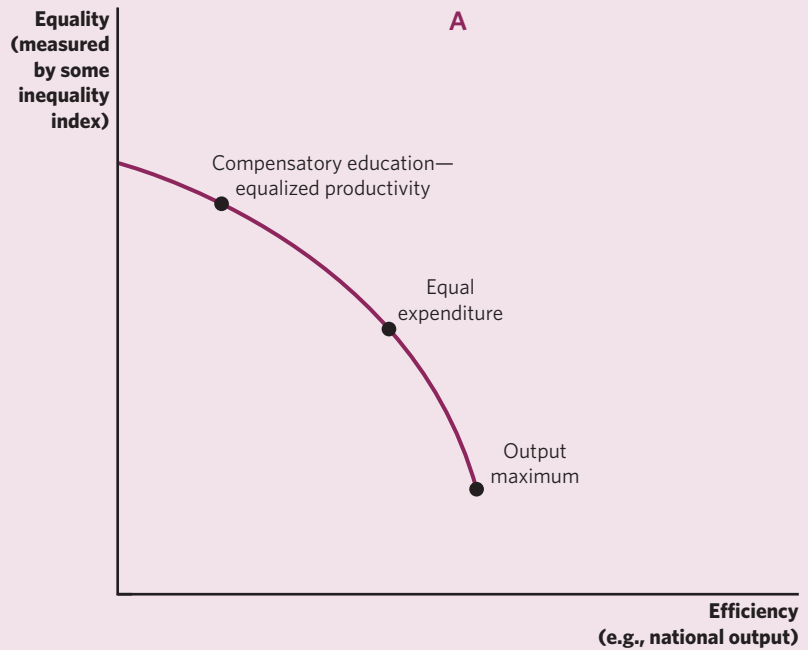
Efficiency

we can get both more efficiency (higher output) and more equality with at least some degree of compensatory education.

The differences in the education–productivity relationship between one individual and another may result either from differences in innate ability or from differences in environment (home background). There is a long-standing controversy about the relative contribution of these two factors in explaining performance. In the case of two individuals with the same innate ability but different home backgrounds, the nature of the education–productivity relationships may depend on whether education in the home (home background) is a *substitute* for or a *complement* to schooling. If home background is a complement to schooling, it means that it increases the return to education. If it is a substitute, the more education that occurs in the home, the smaller the return to formal education.

15

WELFARE PROGRAMS AND THE REDISTRIBUTION OF INCOME

All societies have made some provision for the poor and destitute. We call programs that transfer cash and consumption goods to the poor **public assistance** or **welfare** programs. The manner in which the very poor are taken care of has changed dramatically over time. In medieval Europe the church took responsibility, often establishing almshouses. In modern societies, governments play a major role. In this century, there have been two dramatic changes in the manner in which government fulfills that role. The first was in 1935, when the federal government first took on a major responsibility for welfare under the **Aid to Families with Dependent Children (AFDC)** program. The second was on August 22, 1996, when President Clinton signed a historic bill, the Personal Responsibility and Work Opportunity Reconciliation Act of 1996, which even changed the name of the assistance program from AFDC to **Temporary Assistance for Needy Families (TANF)**. This act culminated a movement for reforming the welfare system that had been the subject of increasing national attention.

In his presidential campaign four years earlier, Clinton had accompanied his promise to "end welfare as we know it" with the slogans "a hand up, not a handout" and "making work pay." Welfare, it was felt, had

created a kind of dependency; there was widespread agreement that the welfare system had to be restructured to help those on welfare get off welfare and become productive members of the labor force. This required economic incentives—it had to pay to work rather than to be on welfare. But it was felt that the carrot of increased income would not suffice: there had to be a stick—the termination of welfare payments after a fixed period of time (two to five years). However, that carrot and stick might not be enough; there needed to be complementary systems and training as well if people were to move from welfare to work.

The antipathy to welfare was partly based on its seeming failure, but it was also largely driven by misperceptions. The federal deficit had become a major source of concern in the 1990s. There was a widespread impression that welfare was largely responsible, and that many on welfare were taking the system for a free ride. In fact, however, total welfare expenditures have never amounted to a large fraction of total government expenditures. In 1996 total welfare expenditures were less than 10 percent of total federal government outlays, and welfare expenditures excluding Medicaid were only 4 percent of total outlays. Even complete elimination of welfare expenditures—a far more Draconian measure than anyone was proposing—would not have eliminated the deficit. Indeed, the magnitude of the spending cuts enacted under the new legislation did not significantly alter the status of the federal deficit.

Ironically, shortly after the passage of the bill, not only did the fiscal deficit cease to be an immediate problem—the federal government turned a large surplus for the first time in three decades—but welfare rolls also declined precipitously, partly because of the booming economy that had brought overall unemployment rates down to 4.3 percent, a level that also had not been seen in three decades.

Critics of these reforms worried, however, what would happen when the country went into a downturn and jobs became scarce. Would the new welfare system work, especially to protect children? The Great Recession that began in 2008 provides telling evidence of deficiencies, with child poverty in particular increasing—children under 18 years living in poor families rose from 18 to 22 percent from 2006 to 2010. Alternative measures of poverty that include some assessment of the value of in-kind benefits show a somewhat less dramatic increase in overall poverty, but still very high levels of childhood poverty.

This chapter provides a brief review of the history of U.S. welfare policies and programs, a look at the United States in comparative perspective, a summary of the major analytic issues, and a discussion of the 1996 welfare bill, explaining why it was so controversial and describing ongoing efforts at further reforms.

FOCUS QUESTIONS

1. What are the major welfare programs? How have they grown over time? To what extent did the growth of welfare programs account for the growth in the deficit in the 1980s? How much of the budget do they consume today?

2. What are the dimensions of the poverty problem in the United States? How has it changed over time?

3. What is the effect of welfare programs on labor supply? What other distortions are associated with welfare programs?

4. In what forms are welfare benefits received? What is the distinction between categorical and broad-based assistance? What are the efficiency and equity issues associated with each of these forms of assistance?

5. What has been the impetus for welfare reform? What have been the major reform proposals? What were the major components of the 1996 welfare reform bill, and what have been the consequences?

A BRIEF DESCRIPTION OF MAJOR U.S. WELFARE PROGRAMS

In the United States, although states and localities had long provided some form of general assistance to the needy, supplementing church and other voluntary programs, the federal government took on major responsibility with the New Deal in the 1930s. The Social Security Act of 1935 established Aid to Families with Dependent Children to provide assistance to families without a major breadwinner and **Supplemental Security Income (SSI)** to provide funds to aged and disabled individuals with low incomes (supplementing Social Security payments).

The next major expansion of federal welfare programs occurred when President Johnson declared a "War on Poverty" in the 1960s. As part of that war, a number of programs were introduced, including Medicaid, to provide medical assistance; Medicaid has subsequently become the largest assistance program in dollar terms. The following sections describe briefly the major programs.

AFDC AND TANF

Since 1935, AFDC had been the primary cash program in the U.S. welfare system. The program was a combination of federal and state programs. The states not only administered AFDC, but also set benefit levels and had some discretion over rules. The federal government provided a portion of the funds, which varied from approximately one-half to three-fourths, depending on the state's per capita income. Programs in which federal outlays depend on state expenditures are called **matching programs**. The federal matching subsidy presumably resulted in the states' providing higher levels of benefits than they would have if they had had to pay the full (marginal) costs themselves.

Starting in 1997, TANF replaced AFDC. TANF represented a marked departure from the earlier system in two ways. First, it replaced the old system of matching grants with **block grants**—a fixed amount of money—with states given considerable discretion as to how that money could be spent, including discretion in determining the eligibility of needy families and the benefits and services those families receive. Second, TANF focused on moving individuals from welfare to work. The states were given broad flexibility in the design and operation of their welfare-to-work programs, but the use of TANF funds had to be consistent with federal priorities of strong work requirements, time limits to receiving assistance, a reduction in welfare

dependency, and the encouragement of two-parent families. Welfare expenditures under TANF have been rising only slightly, and TANF has indeed succeeded in substantially reducing the number of people on welfare. However, TANF has not had a significant impact on reducing poverty levels, and during the 2008 recession, TANF did not perform its countercyclical role of safety net during hard times. Food stamps/SNAP (see below) have only partially filled the gap, with approximately one out of seven Americans receiving these benefits, but with an increasing focus on the federal deficit, many in Congress would like to dramatically downsize this program.

Under both AFDC and TANF, there has been considerable variation in the amount of benefits provided by the states. For example, in 2009 the highest benefits per resident, in New York, were more than ten times the lowest level of benefits, in Idaho. Expenditures have varied greatly over time as well. Total expenditures (in constant 2009 dollars) increased from $35 billion in 1987 to $44 billion in 1995, fell to $30 billion in 2006 in the aftermath of the Clinton welfare reform, and then rose again to $34 billion in 2009 in response to the economic crisis. The total number of beneficiaries has also fluctuated widely, with recipients of cash assistance more than doubling, from 6.7 million in 1969 to 14.2 million in 1993, then dropping 72 percent to 4.0 million in 2008, before climbing to 5.0 million in 2010.[1]

Over time, the details of the federal program have varied. Any **means-tested** program of assistance—that is, any program whose benefits are targeted to those with low incomes—must reduce benefits as income rises. Before 1979, welfare recipients' benefits were cut $67 for every increase of $100 in earned income beyond a certain minimal amount, an effective marginal tax rate of 67 percent. After 1979, the effective tax rate was raised to 100 percent, although some allowance was made for child care expenses and the adverse effects were somewhat mitigated by the Earned Income Tax Credit (see next section).

EARNED INCOME TAX CREDIT

The **Earned Income Tax Credit (EITC)** program supplements the income of low-income families with children[2] by an amount that depends

---

[1] U.S. Census Bureau, *Statistical Abstract of the United States, 2010*, Table 12; and U.S. House of Representatives, Committee on Ways and Means, *Background Material and Data on Programs within the Jurisdiction of the Committee on Ways and Means: 2011 Green Book*. The latter reference is a report published every few years by the House Ways and Means Committee, which provides the most complete compilation on the welfare programs. It is generally referred to as simply the *Green Book* (reflecting the color of its cover). Much of the data for this chapter are drawn from the 2011 report, which will be referred to as simply the *2011 Green Book* in the following references.

[2] The 1993 law added a small credit for childless workers. The maximum amount was $464 in 2011, but all EITC figures are indexed for inflation, so currently, the maximum credit is somewhat greater.

on their income and number of children. Although this amount was small in the 1980s, and was intended to offset Social Security contributions, in 1993, the EITC was greatly expanded to encourage people to work by ensuring that their after-tax earned income would be greater than their untaxed income from welfare payments. By 2011, a worker with three dependents could have his or her income supplemented by as much as 45 percent. Thus, a worker receiving $5 an hour has his or her wage increased, in effect, to $7.25 an hour. The EITC gradually increases as earned income rises, plateaus at the maximum benefit level ($5751 in 2011), and then is phased out as income continues to rise.

Under the old AFDC program, EITC benefits were not counted as income for the purposes of determining benefit levels, so individuals were still better off as a result of working. For example, if a worker receiving the maximum benefit at the time earned an extra $100 of income, although his or her AFDC benefits would go down by approximately $100, the EITC payment would go up by $40, so he or she faced an effective tax rate of "only" 60 percent. TANF has no provision regarding treatment of earned and unearned income. States set their own income limits and make their own rules governing the treatment of earnings and other income, so the interaction between TANF and EITC may vary from state to state.

The number of families receiving EITC benefits grew from 6.2 million at its inception in 1975 to 26.5 million in 2009; the average credit per family grew from $200 to $2200, and the total expenditures grew from $1.25 billion to $58.8 billion.

FOOD STAMPS/SNAP

TANF is the principal **cash welfare program**. Most assistance to the poor is directed at **in-kind benefits** like medical care, or to finance particular categories of expenditure like food and energy.

Food stamps, first introduced in 1964 on a nationwide basis as part of President Johnson's War on Poverty, are designed to assist poor individuals in buying food. The federal government bears essentially all the costs, and sets uniform benefit levels. The benefits depend on a measure of income, which allows a variety of adjustments, the most important of which is that housing expenditures are deductible.

The Food Stamps Program (FSP) has been modified considerably over the past fifty years, mostly regarding eligibility requirements, benefit levels, and administrative procedures. The latest round of reforms was enacted under the 2008 farm bill (formally, the Food,

Conservation, and Energy Act of 2008). These included changing the name to the **Supplemental Nutrition Assistance Program (SNAP)** in an effort to counter the stigma associated with **food stamps**. After two decades of pilot projects and phased implementation, the 2008 farm bill also made electronic benefit transfer (EBT) the program's standard issuance vehicle to improve efficiency and reduce fraud. Paper coupons were replaced by benefit cards—SNAP benefits are deposited electronically into the recipient's account every month; the recipient has access to this account when paying for food, using the benefit card like a bank debit card.

Participation in FSP/SNAP peaked at more than 27 million per month in 1994, before gradually dropping to about 17 million in 2000 after the 1996 welfare reform bill limited the benefits of working-age adults without children. Such individuals cannot receive SNAP benefits for more than three months in a thirty-six-month period if they have not worked twenty hours a week, completed a job training program, or participated in a **workfare** program. (Workfare programs provide assistance in exchange for work.) However, participation rates and program costs have risen steadily since then as a result of a combination of outreach efforts and poor economic conditions, reaching almost 45 million recipients per month and $72 billion in benefits in 2011.[3]

SNAP is the most important "food" program, but two others should be noted. Thirty-two million children receive free or subsidized lunches, and twelve million children receive free or subsidized breakfasts. These programs were started when, in the nationwide draft for World War II, the extent of malnutrition in poor children was first widely recognized; moreover, educators have long argued that inadequate nutrition adversely affected learning.

An additional nine million pregnant women, infants, and children under age 5 who are at risk of inadequate nutrition receive food support under the **Special Supplemental Nutrition Program for Women, Infants, and Children (WIC)**. Inadequate nutrition at these points can have lifelong effects, particularly on a child's ability to learn. While the other programs described so far are **entitlement programs**—that is, anyone meeting the criteria is entitled to receive the benefits; thus, total federal expenditures depend simply on the number of individuals who are eligible and who decide to apply—the federal government provides a fixed amount of money to support the WIC program.

---

[3] The SNAP gross monthly income eligibility limit is set at 130 percent of the poverty threshold (poverty line), a measure of the minimum subsistence level required for, say, a family of four. In 2011, the amount of the threshold for a family of four was $2422 per month. See the website of the Food and Nutrition Service, United States Department of Agriculture, for a detailed history, full eligibility requirements, and program data for SNAP. http://www.fns.usda.gov/snap.

MEDICAID

Established in 1966, this program provides medical assistance to the poor, especially poor children; medical care to the disabled; and nursing home care to a large proportion of the aged. Medicaid is a federal-state government matching program, with the federal government paying between 50 and 83 percent of the costs, depending on the state's per capita income. The states were given considerable discretion in determining **eligibility requirements** and coverage. Medicaid, together with the **Children's Health Insurance Program (CHIP)**, provides coverage for 54 million low-income individuals, including more than 43 million children, comprising half of all low-income children in the United States.[4]

Historically, families receiving benefits under AFDC were eligible for Medicaid. Although TANF families are not automatically eligible, states are still required to provide Medicaid to children and family members who would have been eligible for AFDC using the program's terms as of July 16, 1996 (making adjustments for inflation). Eligibility for Medicaid is thus based on a **threshold test**: those with incomes above the threshold (essentially, the cutoff level of AFDC) are not eligible. Because many employers do not provide medical benefits to low-income workers, many of those on welfare find themselves in a bind: even if they would like to work, they would lose eligibility for Medicaid benefits if they accept a job. This is particularly important for those with children requiring medical attention. As a result, they are reluctant to move off welfare. This situation is referred to as **welfare lock**. States can remove adults from Medicaid rolls if they refuse to comply with the work requirements of TANF.

An important part of President Obama's attempt to increase access to health care was to expand Medicaid eligibility to adults with income up to 138 percent of the federal poverty level (FPL), extending benefits to 21.3 million additional individuals by 2022. Because Medicaid is a joint federal/state program, the new law had strong fiscal incentives for states to participate. In 2012, when the Supreme Court ruled that the individual mandate under the Affordable Care Act was constitutional, it also ruled that the act's provision requiring states to comply with new eligibility criteria for Medicaid or risk losing their funding was constitutional only as long as states would lose just new funds for noncompliance, not all their funding (see Chapter 13). As this book goes to press, approximately thirty

---

[4] CHIP was established as part of the Balanced Budget Act of 1997. It was the largest single expansion of health insurance coverage for children in more than thirty years. CHIP, like Medicaid, is jointly funded by the federal government and the states, although the federal matching rate for state CHIP programs is usually about 15 percentage points higher than the Medicaid matching rate. The purpose of CHIP is to provide health coverage for children in families with incomes too high to qualify for Medicaid but too low to afford private coverage.

states have decided to expand Medicaid under the Affordable Care Act, with the remaining states declining to increase Medicaid coverage at this time but retaining the option to do so in the future.

HOUSING

In 2008, 5.1 million housing units (providing homes for 9.6 million people) received a total of $35.8 billion in subsidies from numerous Department of Housing and Urban Development (HUD) programs. Forty-four percent of those assisted received help under the Housing Choice Voucher program, allowing them to find and lease housing in the private market using the vouchers. Housing vouchers are like food stamps, they give the recipient a fixed amount to be spent on housing chosen by the participant. Multifamily assistance, subsidies provided to private landlords for their low-income rental housing, made up one-third of available units. Public housing assistance, whereby local housing authorities receive HUD funding to build, operate, or make improvements to housing owned by local agencies, comprised the remaining one-fourth of available units. The Internal Revenue Service-administered Low-Income Housing Tax Credit (LIHTC), which subsidizes the acquisition, construction, or rehabilitation of property rented to low-income households, also supported 1.7 million housing units, at an estimated cost of $5.4 billion in tax expenditures.[5]

This emphasis on tenant-based and multifamily assistance, which relies on subsidies for housing provided by the private sector, contrasts sharply with the government's original direct provision of housing for the poor. Public housing was greatly criticized on several grounds. In many cases, costs were high for the quality of housing that was provided—the government seemed an inefficient producer and manager of housing. Worse, government housing projects isolated the poor; they became "warehouses of the poor." The government was unable to maintain them adequately, and they became infested with crime, drugs, and rats.

By the mid-1990s, the problems had become so manifest that the federal government set about tearing down the worst of these projects, ending "public housing as we know it." In one famous incident, the city of St. Louis leveled its Pruit-Igoe project with dynamite rather than trying to maintain the facility. Thus, much public housing assistance has been

---

[5]Some LIHTC properties also receive HUD subsidies. (Sources: Office of Policy Development and Research, Department of Housing and Urban Development, *A Picture of Subsidized Households 2008*; and Joint Committee on Taxation, *Estimates of Federal Tax Expenditures for Fiscal Years 2008–2012*, October 2008, Table 3.)

replaced by these other programs, designed to provide improved incentives and greater social integration. It is worth noting that government assistance for housing the rich and middle class—tax deductibility of mortgage interest—far exceeds government assistance to the poor.

OTHER PROGRAMS

There are many other means-tested programs: the **Low-Income Home Energy Assistance Program (LIHEAP)**; and a variety of education programs such as Head Start, which provides preschool education for children of low-income families, and Pell grants, which help pay for college education for children in low-income families. Still other programs provide job training to unskilled and economically disadvantaged individuals.

Figure 15.1 shows overall expenditures on means-tested programs. Much of the increase is accounted for by Medicaid, which almost doubled, from 27 percent of welfare expenditures in 1975 to 52 percent in 2009. Figure 15.2 compares U.S. spending on means-tested programs with spending by other OECD countries. In contrast to countries like Australia, the United Kingdom, and Canada that have well-developed

FIGURE 15.1

UNITED STATES EXPENDITURES ON MEANS-TESTED PROGRAMS

Overall expenditures on means-tested programs have quintupled as a share of GDP since 1965.

SOURCE: Office of Management and Budget, *Budget of the U.S. Government, Fiscal Year 2013, Historical Tables*, Table 8.4.

Percent of public social expenditures

Australia: 34.9%
Canada: 21.6%
Chile: 3.5%
France: 14.5%
Germany: 13.1%
Italy: 5.0%
Japan: 3.0%
Korea: 10.2%
Mexico: 11.7%
Sweden: 4.0%
United Kingdom: 24.1%
United States: 7.5%
OECD Average: 10.6%
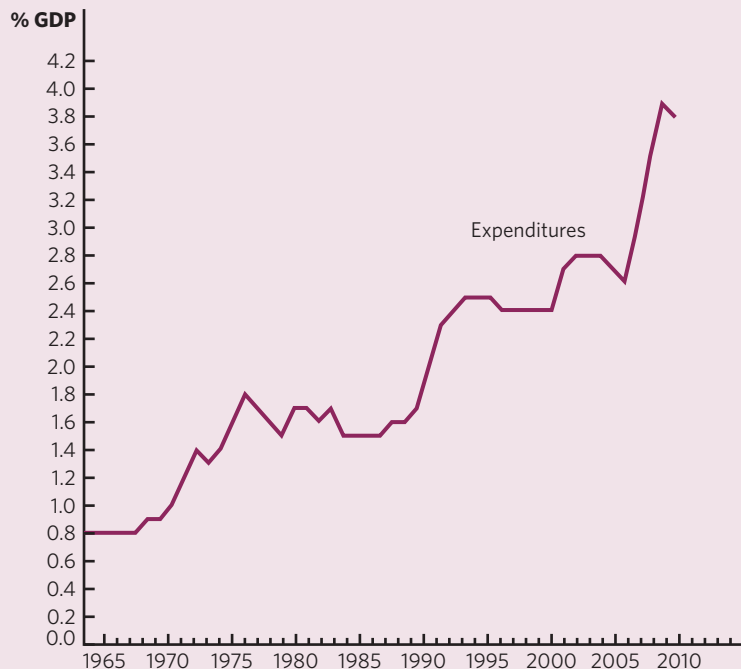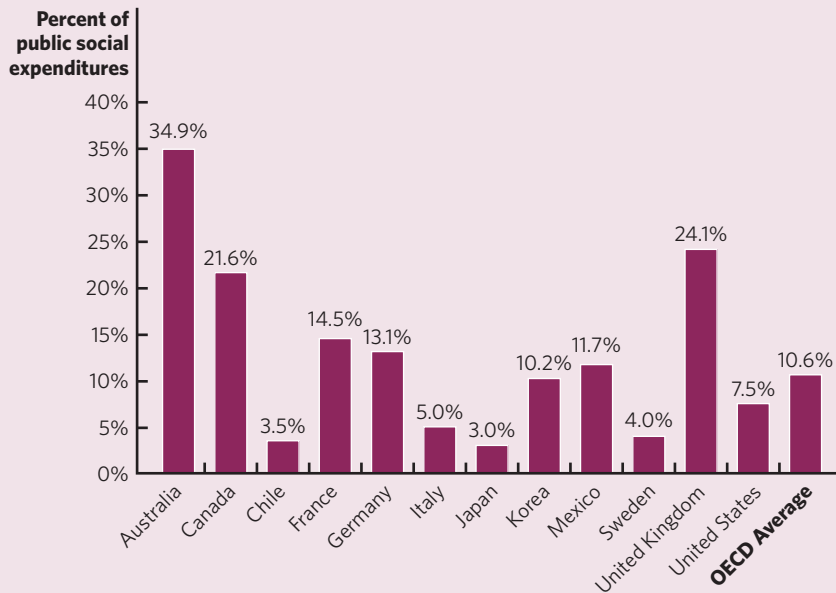
FIGURE 15.2

OECD EXPENDITURES ON MEANS-TESTED PROGRAMS IN 2007
The United States spends less than the OECD average on means-tested programs, in contrast to other Anglophone countries such as Australia, the United Kingdom, and Canada.

SOURCE: OECD Social Expenditure Database (SOCX), www.oecd.org/els/social/expenditure.

social insurance systems (see Chapter 16), means-tested programs in the United States account for a relatively small share of the total public social expenditures.[6]

RATIONALE FOR GOVERNMENT WELFARE PROGRAMS

Chapters 3 through 7 set forth the basic theoretical rationale for government programs: markets may produce efficient outcomes (ignoring the market failures), but they do not necessarily produce a distribution of income that is socially acceptable. Welfare programs focus on one aspect of the distribution of income: those at the very bottom.

---

[6] The OECD comparison includes general means-tested expenditures such as cash and in-kind payments for income maintenance and social assistance, means-tested support payments to the elderly and disabled, and means-tested spending on the unemployed; it does not include specific housing subsidies, spending to improve job or earnings prospects ("active labor market programs"), or means-tested medical expenditures. For a detailed explanation of the calculations underlying this comparison, see W. Adema, P. Fron, and M. Ladaique, *Is the European Welfare State Really More Expensive? Indicators on Social Spending, 1980–2012*; and "Manual to the OECD Social Expenditure Database (SOCX)," OECD Social, Employment and Migration Working Papers No. 124, Paris, 2011.

Government programs are often thought of as **safety net** programs. There is a view, for instance, that a civilized society cannot allow individuals to starve, or to die as a result of inadequate health care. These programs are also often thought of as a form of **social insurance**. Knowing that adversity can strike anyone, individuals may think, "There but for the grace of God go I." The knowledge that there is a safety net adds to their sense of economic security, and thus to their overall well-being. The welfare programs are not formally part of the social insurance system, in which individuals pay explicit premiums to obtain, say, unemployment insurance against the risk of losing their jobs, but they still perform this insurance function—they are intended to provide funds to allow individuals to get back on their feet. Safety net and social insurance programs are increasingly grouped together under the general category of **social protection**.

Similarly, programs aimed at children have been justified both as a reflection of basic values and as an investment in the future. As a basic value, it is argued that all children should, to the extent possible, have the opportunity to live up to their full potential. Children with inadequate nutrition and health care cannot do this. Further, children growing up in poverty are less likely to finish high school and more likely to end up in a life of crime: in fiscal year 2009, more than half of all convicted federal offenders had not graduated from high school, in contrast to the 5 percent who were college graduates.[7] Thus, the overall returns to society from reducing childhood poverty appear to be large.

DIMENSIONS OF THE PROBLEM

The **poverty rate** is the fraction of the population whose income lies below a threshold that is intended to measure the minimal level required to maintain a subsistence living standard. The threshold clearly needs to be adjusted each year to reflect increasing prices (inflation). Although there is some controversy about the measure (what does it really mean to be in poverty?), movements in the poverty measure do track what is happening at the bottom of the income distribution. The overall poverty rate, which was more than 22 percent in 1960, was reduced to a much lower level of about 12 percent in the 1970s. Since then, however, the poverty rate has averaged 13.4 percent, fluctuating from a low of 11 percent in 2000 to peaks of more than 15 percent in 1983, 1993, and 2010. The primary reason for the increase in poverty since the 1970s has

---

[7] M. Motivans, Bureau of Justice Statistics, U.S. Department of Justice, *Federal Justice Statistics 2009—Statistical Tables*, Table 4.4.

been reduction in the real wages of those at the bottom of the income distribution since that time—even those in the middle have not fared well. However, the poverty rate in the United States is an absolute number based on an assesment of a family's basic needs, whereas many European countries base their measure of poverty on *relative deprivation*, say, half of median income. When poverty is viewed as a relative concept, as inequality increases, typically poverty also increases. Given the large increase in inequality in the United States since 1980, this would mean that, using such measures, the United States would have seen an almost steady increase in poverty rather than the steady average with modest fluctuations described above.

There is considerable controversy about the factors contributing to the decline in the well-being of those at the bottom. Weaker unions have played a part. So too has globalization—competition, in effect, from unskilled, low-paid workers abroad. Those at the bottom are disproportionately from minority groups, and there has been remarkably little progress in reducing discrimination against these groups. Technology, too, has played a role: new technologies seem to be biased in favor of higher skills, and the quality and quantity of education provided at the bottom has not kept pace. This is why most experts believe that combating poverty requires a comprehensive agenda, of which welfare programs such as TANF and SNAP are an important part, but which also need to include EITC, increase of the minimum wage, programs fighting discrimination, and provision of better access to high quality education and health care for low-income families.

Particularly disturbing is the rapid rise in the number of children in poverty. The poverty rate for children rose from 15 percent in 1970 to 22 percent in 2010. Much of this arose from an increase in the number of children being raised in households with single earners: more than half of all children in poverty resided in households headed by single women in 2010.

The problems of poverty in the United States are seen most strongly among the black population. In 2010, 27.4 percent of blacks lived below the poverty line, almost triple the 9.9 percent poverty rate for non-Hispanic whites.[8] High poverty, combined with inadequate access to health, contribute to the high infant mortality and low life expectancy rates for blacks compared to those for whites. In 2007, whites experienced an infant mortality rate of 5.6 per 1000 live births, whereas blacks had a rate of 13.2; life expectancy in 2007 was 78.4 years for whites, but only 73.6 for blacks.[9]

---

[8] U.S. Census Bureau, *Historical Poverty Tables*, Tables 2, 3, and 10.
[9] U.S. Census Bureau, *Statistical Abstract of the United States: 2012*, Tables 104 and 116.

ANALYTIC ISSUES

Over the years, policy debates surrounding the welfare programs have focused on a set of incentive and equity issues: Do welfare programs discourage work? How can people most effectively be moved from welfare to work? Do noncash programs have other adverse incentive effects? On what grounds can they be justified? Do welfare programs contribute to the long-run welfare problem by causing dependency and encouraging out-of-wedlock births?

LABOR SUPPLY

Welfare programs use income as a basic criterion for determining eligibility. As income rises, benefits are reduced; if income rises above a given threshold, a family may become ineligible for Medicaid benefits. So individuals care about their total income—what they earn plus what they receive from the government. The total net income of poor individuals thus rises far more slowly than their before-subsidy income. It is as if poor individuals face very high marginal tax rates. (Recall that the *marginal* tax is the extra tax an individual pays as a result of earning an extra dollar of income.) Thus, prior to the 1996 reforms, a very–low-income individual on welfare who earned an extra $100 would have AFDC benefits cut by $100, but have EITC increase by $40. The increased EITC payments, however, would reduce food stamps by about $10. Hence, the net income would increase only by $30—in effect, there was a 70 percent tax rate. Such high tax rates discourage work, or at least discourage reporting the income earned from work.

Because the marginal return—the extra net income received from working an extra hour—is reduced, individuals on welfare have less incentive to work. A survey of studies shows a midpoint estimate of work reduction for AFDC and food stamp recipients of 30 percent attributable to AFDC and of 10 percent attributable to food stamps.[10],[11] The main effect seems to be a

---

[10] R. Moffit, "Incentive Effects of the U.S. Welfare System: A Review," *Journal of Economic Literature* (March 1992): 1–61. There is a wide range of estimates. See, for instance, J. Hausman, "Labor Supply," in *How Taxes Affect Economic Behavior,* ed. H. Aaron and J. Pechman (Washington, DC: Brookings Institution, 1981), pp. 27–72; Hausman obtains estimates that are thirty-seven times as large as those obtained by R. Moffit in "An Economic Model of Welfare Stigma," Rutgers University mimeo, 1980. For more recent studies, see B. L. Wolfe, "Incentives, Challenges, and Dilemmas of TANF," Institute for Research on Poverty Discussion Paper No. 1209-00, University of Wisconsin—Madison, May 2000; J. Liebman, "The Optimal Design of the Earned Income Tax Credit," in *Making Work Pay: The Earned Income Tax Credit and Its Impact on American Families*, ed. B. D. Meyer and D. Holtz-Eakin (New York: Russell Sage Foundation, 2002), pp. 196–234; and N. Eissa and H. W. Hoynes, "Behavioral Responses to Taxes: Lessons from the EITC and Labor Supply," in *Tax Policy and the Economy*, vol. 20, ed. J. M. Poterba (Cambridge, MA: MIT Press, 2006), pp. 73–110.

[11] These numbers may exaggerate the effects, because there may be a larger reduction in *reported* income and work than in actual income and work.

reduction in labor force participation (whether the recipient works or not); for those who work, there appears to be little effect on hours.[12]

The EITC was designed to provide greater incentives for individuals to participate in the labor force—it increased the overall return to work. (At the same time, for many individuals, it decreased the marginal return to working an additional hour, as benefits were cut as income rose; but the earlier cited studies suggest that this effect is not significant.) One of the aims of TANF was to complement the "carrot" that the EITC provided for those with children to participate in the labor force. Eligibility for receiving welfare had strict time limits and work requirements.

Welfare programs, like Medicaid, that employ thresholds may have even more dramatic adverse effects on work. The extra return to earning a dollar that pushes a family over the threshold is very negative. Loss of eligibility of medical benefits is cited as one of the main impediments to moving people from welfare to work, especially because many employers of low-wage workers do not provide health care benefits to their employees.

Although many workers may not work because of the loss of benefits, many others work but do not report their income. For instance, a welfare recipient may be employed as a household worker, performing services such as housecleaning or child care, but not report the income. Given the inherent lack of data, economists remain uncertain about the prevalence of this practice, and therefore about the full effects on labor supply of welfare programs.

DIAGRAMMATIC EXPOSITION   We can use standard budget constraints and indifference curves to illustrate the adverse effects of welfare on work in the standard economist's model. In Figure 15.3A, *BB* gives the before-tax budget constraint, showing how the consumption (after-tax income) of a worker, Alfred, increases as work increases (leisure decreases). The slope of the budget constraint is the wage: if Alfred's wage is $6 an hour, an extra hour of work increases consumption by $6. The indifference curves have the shape illustrated because Alfred does not like to work (or at least does not like the work available to him)—he prefers leisure. Alfred requires extra consumption to compensate him for working more, and because the more he works the less leisure he has, the more valuable leisure is to him at the margin. Furthermore, because the more he works the higher his consumption, the less valuable is his marginal increase in consumption. Accordingly, the extra consumption he requires to compensate him for an extra hour of work—the marginal rate of substitution—increases the more he works. That is why the indifference curve not only is upward sloping, but becomes steeper as he works more. The original equilibrium

---

[12] R. Moffit and A. Rangarajan, "The Work Incentives of AFDC Tax Rates: Reconciling Different Estimates," *Journal of Human Resources* (Winter 1991): 165–179.

FIGURE 15.3

EFFECT OF WELFARE
PROGRAMS ON
LABOR SUPPLY

Welfare programs can adversely
affect labor supply. (A) Stylized
version of welfare program
before 1979, when payments
were reduced as the individual
earned more. Both the income
and substitution effects led
to reduced work. (B) Stylized
version of welfare program
after 1979, when benefits
were reduced dollar for dollar,
beyond a certain minimal
amount. No one worked beyond
this level. (C) A simplified ver-
sion of the current system, with
TANF, EITC, and food stamps.
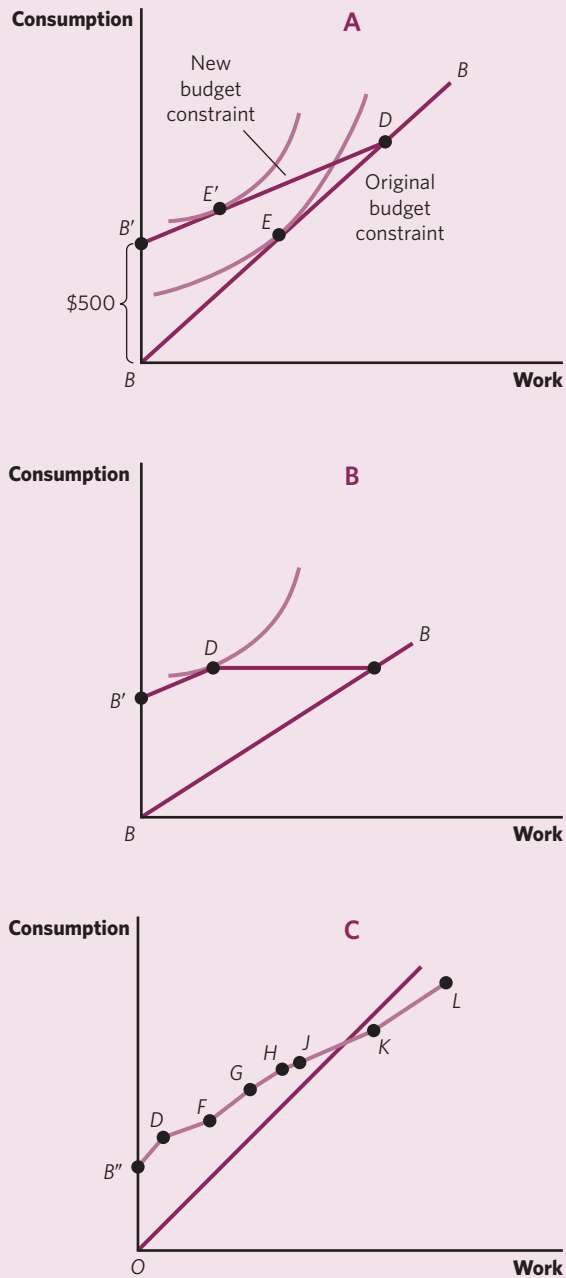
is at point *E,* where the slope of the budget constraint equals the slope of the indifference curve: the wage equals the marginal rate of substitution.

B'DB is the new budget constraint under a welfare system in which Alfred gets a fixed payment—say, $500 per month, which decreases the higher his income becomes. (At point *D*, the benefit is completely phased out.) We assume—as was the case prior to 1979—that for every $3 of income Alfred earns, he loses $1 of benefits. Thus, the new budget constraint is flatter; his after-tax wage is not $6 an hour, but $4 an hour. The new equilibrium is at *E'*, with a lower level of work than before. Work is reduced for two reasons. First, because Alfred is better off, he takes some extra leisure. This is the income effect; normally, as incomes rise, individuals consume more of any good. Second, because the return for working an extra hour is reduced, Alfred's incentives for working are reduced; this is the substitution effect. Both have the effect of reducing work.

Figure 15.3B shows the budget constraint under a welfare system in which Alfred loses a dollar for each dollar he earns beyond a certain minimum amount. Not surprisingly, Alfred chooses to work only enough to generate that income; he would be foolish to work more than that, as he earns no marginal return.

Figure 15.3C shows a simplified version of the current system, combining TANF, EITC, and food stamps. In this stylized version, there are seven segments to Alfred's budget constraint. There is a certain basic benefit, given by *OB''*. There is a certain minimal amount that he can earn without losing benefits. Thus, the slope of the budget constraint in the segment *B''D* is just the wage ($6 an hour), augmented by the EITC, but reduced by Social Security taxes. Beyond this point, benefits are reduced the more Alfred works; but the EITC supplements his wage by 40 percent, so his net tax rate is about 70 percent—the return per hour worked is $1.80. Thus, in the interval *DF*, the budget constraint is relatively flat. At *F*, TANF benefits are eliminated, but Alfred still has food stamps, pays Social Security taxes, and receives the EITC. Thus, the slope of the budget constraint in the interval *FG* is much steeper than in *DF*. Then, however, at an only slightly higher income, Alfred starts paying income taxes. This is reflected in a flatter budget constraint in the interval *GH*. At a slightly higher income, EITC benefits start being cut (segment *HJ*). This greatly reduces the marginal return to working. Soon thereafter, eligibility for food stamps is exhausted. In the interval *JK*, the slope of the budget constraint is thus slightly steeper. Eventually, EITC benefits also are exhausted, and the individual has to pay Social Security and income taxes (segment *KL*).

From the budget constraint, one can see that, depending on the part of the budget constraint at which an individual is located, there

INCENTIVE EFFECTS OF WELFARE PROGRAMS

1. The EITC provides positive incentives to participate in the labor force.
2. Welfare programs provide disincentives to work longer hours; effectively, there is a high marginal tax rate on working longer hours.
3. Benefits with thresholds—which suddenly disappear when incomes exceed a certain level—have particularly adverse incentive effects near the cutoff level.

can be a variety of effects on labor supply. In general, the marginal return to working is less than the wage—in some cases, far less. This is the substitution effect; the large substitution effect implies that there are large distortions, and that labor supply will be reduced. In addition, for most poor individuals, their after-tax and transfer-payment income is greater than it would be without taxes and subsidies; the income effect too induces less work. The income and substitution effects thus reinforce each other.

The standard economic model often suggests bigger labor supply effects than are observed. This is partly because it omits important considerations, both economic and noneconomic. Individuals know that if they are unemployed for an extended period of time, future job prospects will be jeopardized. Thus, after searching for a while, they may accept a job even if it is far below their qualifications. Moreover, there is a considerable body of research suggesting that the economist's model of labor supply is, at best, incomplete. In the economist's model, if one gave an individual the same money that he would have received on the job, but left him without employment, he would be better off because consumption would be unchanged and leisure increased. In practice, individuals value meaningful employment. It gives them a sense of worth and provides important source of social connectedness. The loss of welfare especially from extended periods of unemployment is far greater than can be accounted for by the loss of income.

The economic effects of limiting unemployment insurance or other sources of income support are also more ambiguous than this standard analysis would suggest. An important aspect of job search is finding the right job—a job that uses the individual's skills as fully as possible. If people are forced to terminate search prematurely, there will be a less perfect "match," and economic productivity will suffer as a result.

CASH VERSUS IN-KIND REDISTRIBUTION

More than 70 percent of welfare benefits are not unrestricted cash, but are directed to the purchase of food, housing, energy, or medical care. Today, Medicaid and other medical care alone account for more than half of all welfare expenditures. Two dollars out of every ten transferred

through welfare are provided in the form of subsidized food, housing, and energy. The present system is criticized on three grounds:

1. It introduces inefficiencies in resource allocations when there are substitution effects; and when there are no substitution effects, the consequences are not different from those of a direct transfer of income.

2. It is inappropriate for the government to attempt to distort individuals' consumption decisions; that is to say, in-kind benefits are paternalistic.

3. It is administratively costly: each program must be run separately; several different agencies must determine the eligibility of each individual for each program. (Eligibility standards to determine who is qualified to receive aid under each program are based primarily on income, but adjustments for family size and other circumstances are generally made, and may differ markedly from program to program.)

The sections that follow discuss the first two criticisms in greater detail. The third criticism applies broadly to most focused welfare programs, whether they are directed at particular groups of individuals **(categorical programs)** or particular forms of assistance (in-kind benefits). The eligibility standards for the various programs could clearly be coordinated and simplified, but there will always remain significant administrative costs associated with such programs. The broader question, addressed in the following sections, is whether there are benefits from such noncash and categorical programs that justify these additional costs.

INEFFICIENCIES FROM IN-KIND BENEFITS

In-kind benefit programs often distort individuals' choices, typically because they reduce the cost of obtaining the good—and thus induce individuals to consume more of the good than they would otherwise. In some cases, the programs reduce the marginal cost of the good to zero—up to the limit provided by the government. Clearly, if the government is giving away a good, individuals will consume it (as long as it has a positive benefit), whether or not the value to them is less than the cost of production. Of course, one of the reasons for in-kind benefits is precisely to promote the consumption of certain goods. *Whether* government should do that is another matter, to which we shall turn shortly.

SNAP The SNAP food assistance program provides an excellent illustration of how in-kind benefits can distort behavior, as we saw earlier in this chapter. Three aspects of behavior are affected:

1. Because (under the current system) an individual with a given income receives a fixed amount of SNAP assistance, in principle SNAP should have the same effect on behavior as would a comparable transfer of income, as long as the individual consumes at least as much in food as provided by the SNAP assistance—which is the case for 85 to 90 percent of participating households.[13] For these households, it appears that SNAP has a slightly larger effect on consumption than one might have expected, with $100 of assistance leading to an increase in food expenditures of between $20 and $45. As one of the intents of food stamps is to improve the nutrition of the poor, advocates of food stamps view this as a marked success.[14]

2. Because the value of SNAP assistance given decreases as an individual's income increases, SNAP has an adverse effect on labor supply.

3. Because the value of SNAP assistance given increases the smaller the individual's income *net of housing expenditures,* SNAP encourages housing consumption. It is ironic that whereas SNAP is intended to encourage food consumption, its major effects may be to encourage housing consumption and to discourage work.

Although these are theoretical possibilities, there is little empirical evidence suggesting that such effects are significant. By contrast, modern behavioral economics suggests that labeling assistance as "food" or "housing" will lead people to consume more food or housing than they otherwise would; the evidence is consistent with that hypothesis.

MEDICAID AND THRESHOLDS All means-tested programs can have adverse effects on labor supply, as we have noted, but the effects of certain in-kind benefits can be particularly dramatic. These are programs under which the government provides a certain free benefit to those who are

---

[13] J. Currie, "Welfare and the Well-Being of Children: The Relative Effectiveness of Cash and In-Kind Transfers," in *Tax Policy and the Economy*, vol. 8, ed. J. M. Poterba (Cambridge, MA: MIT Press for the National Bureau of Economic Research, 1994), pp. 1–43.

[14] B. Devaney and T. M. Fraker, "The Effect of Food Stamps on Food Expenditures," *American Journal of Agricultural Economics* (February 1989): 99–104. These results contrast with estimates of 5 to 13 cents additional food expenditures for an additional dollar of cash, reported by Thomas M. Fraker in *The Effects of Food Stamps on Food Consumption: A Review of the Literature* (Alexandria, VA: U.S. Department of Agriculture, Office of Analysis and Evaluation, Food and Nutrition Service, 1990), and are consistent with experimental findings that show that when individuals receive the equivalent in cash instead of food stamps, food consumption actually decreases by between 18 and 28 cents for each dollar of food stamps cashed out. See T. M. Fraker, A. P. Martini, and J. C. Ols, "The Effect of Food Stamp Cashout on Food Expenditures: An Assessment of the Findings from Four Demonstrations," *Journal of Human Resources* 30, no. 4 (Fall 1995): 633–649; and M. K. Fox, W. Hamilton, and B. Lin, eds., "Literature Review," *Effects of Food Assistance and Nutrition Programs on Nutrition and Health*, vol. 3 (Washington, DC: U.S.Department of Agriculture, Economic Research Service, Food and Nutrition Research Program, 2004).
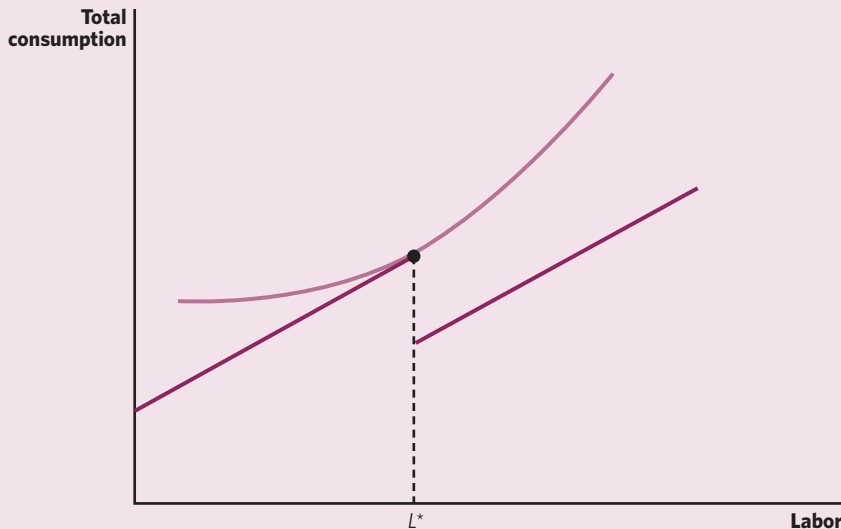
EFFECT OF THRESHOLDS ON LABOR SUPPLY
Programs like Medicaid, where there is a given level of income above which individuals lose eligibility, discourage work.

eligible—that is, for those whose income is below a certain threshold—and nothing for those above it. The most important program having this feature is Medicaid. There is an obvious administrative reason for such provisions: it makes it much easier to determine whether an individual qualifies for a program. An alternative would require the individual to face a fee schedule that depended on his or her income. Now, a doctor simply has to know whether the individual qualifies for the program. Under the alternative, either the doctor or the government would have to send poor individuals a bill for each service rendered, which would depend on the individual's income.

Let's say Figure 15.4 shows the budget constraint for David, who consumes a fixed amount of medical care services. Labor is shown on the horizontal axis, and consumption, including Medicaid benefits, on the vertical. As David works beyond a certain level, his income, L^*, exceeds the threshold for Medicaid, so his total consumption actually falls. Clearly, David will work an amount either just short of L^* or considerably greater than L^*.

HOUSING Public housing programs raise a number of complicated issues associated with in-kind programs. Direct provision increases the quantity of housing supplied at any price, and thus benefits not only those who receive public housing, but also others, as it drives down the equilibrium rent. On the other hand, government has proven itself an inefficient producer of housing—with the costs per square foot considerably higher than in comparable privately provided housing. That is one of the motivations for vouchers and tax credits.

Tax credits also operate directly on the supply side. Those who produce low-income housing receive a subsidy through the tax system. More housing is supplied at each price, lowering the equilibrium price both for those living in subsidized housing and for others. Programs that require housing receiving such special tax treatment to be mixed-income—that is, that the apartment buildings not just have low-income units—have reduced the isolation of the poor. However, these programs have also been criticized: much of the tax benefit goes to the brokers who package the deals. The "bang for the buck" may be relatively low.

Housing advocates, accordingly, have argued that money should be given directly to the poor, rather than to real estate developers. This is what housing vouchers do: they subsidize the rent that individuals pay. Figure 15.5 shows vouchers as shifting the demand curve for housing to the right. In the short run, when the supply is relatively inelastic (Figure 15.5A), the main effect may be on equilibrium rents—then those not receiving the vouchers are actually made worse off. In the long run, when the supply is relatively elastic (Figure 15.5B), supply responds—just as it would if the government had directly provided the housing itself. Still, unless the long-run supply curve is horizontal, those not receiving the subsidy are hurt.

One particularly troublesome feature of the design of U.S. housing programs arises from the fact that there is a limited budget for assisting the poor in acquiring housing. The government faced a dilemma: it could provide a little subsidy to all poor people, or a large subsidy to a few. It chose the latter strategy. Specifically, it required that public housing meet a minimum standard. (The government did not want to be accused of being a slumlord.) However, the minimum standard it chose was quite high. As a result, only about a quarter of those eligible for housing assistance receive it. There is concern that the result is worse than a random lottery, with most poor people unaffected, and those lucky enough to get the subsidy receiving a big windfall. The reason that it is worse than just a lottery is that it inhibits those lucky enough to get the subsidy from moving from welfare to work. Thus, public housing represents another form of welfare lock. America is a mobile country; every year about 16 percent of Americans move from one community to another, often in search of a job, or a better job. For poor individuals in public housing, though, moving becomes very unattractive. An individual who moves to another community goes to the bottom of the queue of applicants for housing assistance. In most cases, this individual will lose his or her benefits by moving. If the value of the housing subsidy is, say, $300 per month, the loss represents almost 40 percent of the income of a minimum-wage worker. When combined with other costs of moving—and other taxes, implicit and explicit—moving to another community for work becomes decidedly unattractive.

Rent

A

Demand curves

R_1

R_0

After
vouchers

Supply
curve

Before
vouchers

Quantity

Rent

B

Demand
curves

Supply
curve

R_1
R_0

After
vouchers

Before
vouchers

Quantity
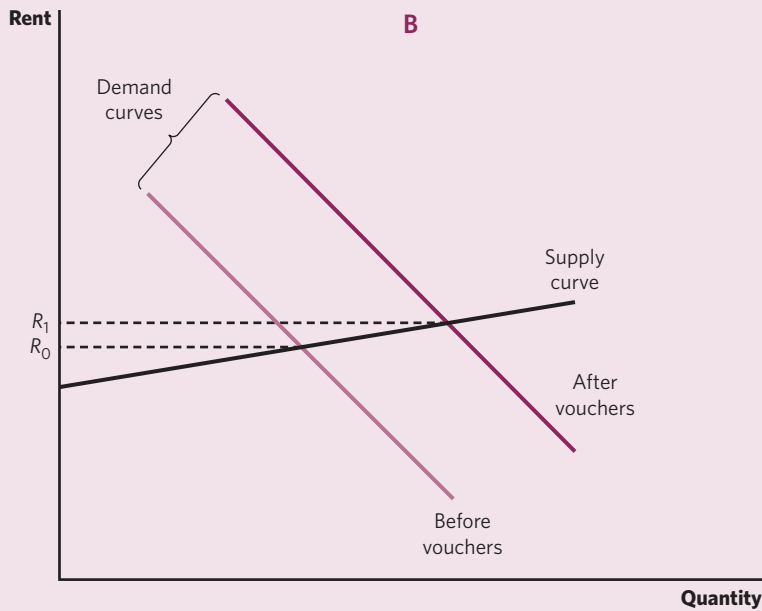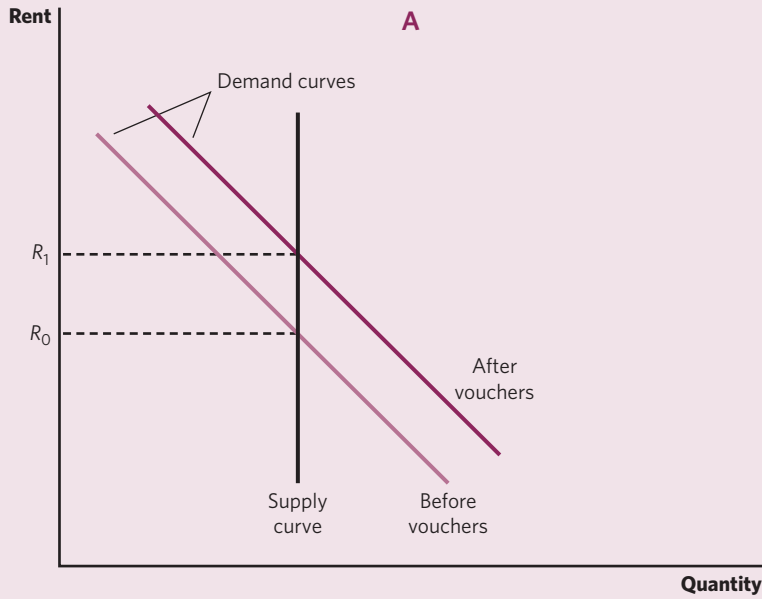
FIGURE 15.5

**HOUSING MARKET
EQUILIBRIUM
WITH VOUCHERS**

Vouchers shift the demand
curve to the right. (A) In the
short run, supply is inelastic. The
main effect is on price. Those
not receiving the vouchers are
worse off, as rents increase.
(B) In the long run, supply is more
elastic. However, unless the
long-run supply curve is per-
fectly horizontal, there will be
some increase in rents.

ARE IN-KIND BENEFITS PATERNALISTIC?

We have noted that in-kind benefits are typically distortionary—that is, there is a substitution effect associated with them. Individuals would be better off if they were just given the same amount in cash; alternatively, the government could make them just as well off, and reduce its total expenditures. Ironically, the only circumstances in which in-kind benefits are not distortionary are when they have no substitution effect, only an income effect; then they simply increase administrative costs, but the effect on consumption of food (or whatever the benefit being provided) is the same as a cash benefit.

Most advocates of in-kind benefits are bothered more by those cases in which it turns out that the benefits are ineffective, than by those in which they are distortionary. In their view, the reason for government's providing in-kind benefits is that it wants to ensure that the money transferred is spent on "good" uses—on housing, food, and medicine. In this perspective, what society cares about is not just the *perceived* well-being of the recipients, but also the outward manifestations of poverty—the slums, malnutrition, and the like that result from it. Some economists find this paternalistic view of the government objectionable; they argue that it violates the principle of consumer sovereignty.

Research in behavioral economics and the economics of information has shown, however, that many individuals, especially those of limited income and education, may not be fully cognizant of the effects on themselves and their children of inadequate nutrition or preventive health care, and may be prone to exploitation by, for instance, underregulated banks and merchants. As we noted, these in-kind programs may lead to better nutrition and health than would be the case with a cash grant.

Moreover, in Chapter 13, we encountered the view, called *specific egalitarianism*, that society should concern itself not only with the distribution of purchasing power in general but also with access to particular goods, services, and rights. The right to a minimal level of medical care, food, and shelter, in this perspective, ought to be viewed as a basic right.

There is another rationale for in-kind benefits: the government is concerned not only with what gets consumed, but also with who consumes it. Some in-kind assistance is intended to target particular beneficiaries, especially

---

IN-KIND BENEFITS

Arguments in favor:
- Targets aid where it is most needed
- Specific egalitarianism
- Political expediency

Arguments against:
- High administrative costs
- Inefficient (distortionary): same gain in well-being of recipients could be achieved at lower costs
- Rules exacerbate distortions: SNAP encourages consumption of housing
- Ineffective (often do not significantly increase expenditures on desired items, like food)
- Paternalistic

children and those needing medical care. In a world of perfect information, the government might be able to target these groups, and in that case, cash payments would clearly be preferred. However, in a world of imperfect information, restricting transfers to in-kind payments increases the proportion of funds going to the intended groups.

A study of this in the United States found that in-kind transfers have stronger effects on the well-being of children than do cash transfers, and that more narrowly targeted programs, such as Medicaid and Head Start, have greater impacts than more broadly targeted programs like SNAP. In contrast, a study of the impact of cash transfers of equal value to the in-kind Food Support Program (*Programa Apoyo Alimentario*) in rural Mexico found that although both forms of transfer had a large positive impact on food and total consumption, and led to a significant reduction in poverty as well, the magnitude of impact was the same for both.[15]

Finally, there are *political* arguments for in-kind benefits. Different congressional committees oversee each of the programs. SNAP, for example, is under the jurisdiction of the congressional agricultural committees. One of the reasons why the committees supported expansion of the SNAP food assistance program was that they viewed the program as benefiting their special interest—the agricultural sector—just as the housing industry is one of the strongest advocates of housing programs. In this perspective, overall support for welfare benefits would be reduced if these benefits were converted into cash payments. There is also a political dimension to the economic paternalism discussed previously: it is easier for politicians to convince their constituents to support in-kind benefits because of the widespread concern that cash assistance will be abused—for example, used to buy alcohol or cigarettes rather than nutritional food staples.

CATEGORICAL VERSUS BROAD-BASED AID

Should aid be given to all poor people or only to the poor who fall into certain categories? A number of programs raise this question. For instance, the Supplemental Security Income program transfers income only to the aged or disabled poor.

The most basic criticism of categorical aid programs (whether in-kind or cash) is their relatively high administrative costs, as noted in the discussion

---

[15] J. Currie, "Welfare and the Well-Being of Children: The Relative Effectiveness of Cash and In-Kind Transfers," in *Tax Policy and the Economy,* vol. 8, ed. J. M. Poterba (Cambridge, MA: MIT Press, 1994), pp. 1–44; and E. Skoufias, M. Unar, and T. Gonzalez-Cossio, "The Impacts of Cash and In-Kind Transfers on Consumption and Labor Supply: Experimental Evidence from Rural Mexico," Policy Research Working Paper No. 4778, Washington, DC, World Bank, 2008. See also C. Blackorby and David Donaldson, "Cash versus In-Kind, Self-Selection and Efficient Transfers," *American Economic Review* 78, no. 4 (September 1988): 691–700.

CATEGORICAL AID

Argument in favor:

- Targets aid where it is most needed, reducing overall distortionary effects and costs

Arguments against:

- Unfair to treat different poor people differently
- Distortionary effects in meeting eligibility standards (effect of TANF on family dissolutions)
- High administrative costs

of in-kind benefits. The greater expense of these programs arises primarily from the costs associated with ascertaining eligibility. For instance, administrative costs per beneficiary for the SSI program are now three times more per beneficiary and six times more per dollar of benefit payments than administrative costs for the Old-Age and Survivors Insurance Program.[16]

EFFICIENCY AND EQUITY ISSUES   Besides considerations of administrative costs, both efficiency and equity issues arise in comparing categorical versus broad-based aid. Categorical aid may have the effect of inducing individuals to fall into the benefited category; it may have a distortionary effect. This is not true of Social Security in the sense that individuals do not become older faster simply to take advantage of the program, although people can choose to retire at an earlier age to collect Social Security. However, there were allegations that AFDC had contributed to the breakup of families, since in many states, eligibility could be affected by the presence of a man in the household.

The major advantage of categorical aid over broad-based programs is that under certain circumstances it can provide more effective redistribution, with less loss in efficiency. It can enable the targeting of aid to the most needy, who, at the same time, will not have adverse incentive responses. We have repeatedly emphasized the trade-offs between equity and efficiency considerations in the design of redistribution programs. Providing a basic level of income through transfer payments that decline as the individual's income increases (from wages or other sources) may discourage work. This will normally imply that a lower level of redistribution is more desirable for individuals whose response to incentives is large, than for individuals whose response to wage incentives is small (e.g., those over age 70).

There is an equity argument against categorical programs. Some believe that the government should not discriminate in favor of or against any particular groups. Two individuals who are equally poor should receive the same amount from the government, whether they are young or old. There should not be "favored" categories, such as single-parent households or the aged. Admittedly, older individuals may have more medical expenses, and one might want to adjust the transfers to take this into account; however, this is already effectively done through Medicare.

---

[16] S. Szymendera, "Social Security Administration (SSA): Budget Issues," Congressional Research Service Report R41716, August 2011.

IS MEANS TESTING OBJECTIONABLE IN ITS OWN RIGHT?

A far more fundamental set of questions is raised by means testing. Should benefits be targeted at the poor alone? Any program like Medicare that provides benefits for all individuals, irrespective of their income, has a disproportionate effect on the poor. One way to help the poor is to provide equal benefits to everyone in society. Such programs have several distinct advantages. First, they reduce the distortions associated with the phaseout of means-tested benefits. Second, means-tested programs are often seen as carrying a stigma. That is one reason why Social Security and Medicare were not means tested. Not all those who are eligible for public assistance participate. The administrative complexity may provide part of the explanation, but another reason may be that some of those eligible view means-testing programs as demeaning. Finally, there is a political argument, which is sometimes put thus: "Means-tested programs are mean"—that is, because they lack political support, they tend to be stingy.

In recent years, as the government has faced perceptions of increasing financial stringency, these arguments have been overwhelmed by the desire to target the limited funds to those most needy. Thus, there has been increasing support for making wealthier individuals pay a higher fraction of their Medicare costs. (Note that whether there is "financial stringency" is largerly a matter of perception and politics. Countries far poorer than the United States have provided broad social benefits without means testing.)

OTHER DISTORTIONS

So far, we have discussed how welfare programs can adversely affect labor supply, and how in-kind benefits can distort consumption decisions. Critics of welfare programs argue that they have even more fundamental effects on behavior, such as that they have contributed to family breakdown, increased teenage pregnancy, and concentrated welfare dependency in the states that pay the highest benefits, as the poor migrate to those states.

It is clear that welfare has, or has had, features that might result in those behavioral effects. For instance, because welfare benefits typically depend on household income, the departure of a low-wage father may increase total "family" income by making the family eligible for assistance.[17] It is also true that a disproportionately large number of

---

[17] According to one study, the combined AFDC and food stamp benefit for an AFDC family was higher than the reported income of 62 percent of black men aged 20 to 24 and of 29 percent of black men aged 25 to 34. See F. Levy and R. Michel, "Work for Welfare: How Much Good Will It Do?" *American Economic Review* 76 (May 1986): 399–404.

CONDITIONAL CASH TRANSFER PROGRAMS

One of the most important innovations in welfare programs over the past decade is **conditional cash transfers (CCTs)**, sometimes referred to as *contingent welfare programs*. The objective of CCTs is to offer basic income support to low-income families, while at the same time promote investments in human capital to reduce long-term poverty and inequality. CCTs are incentive-based social welfare programs that provide direct cash transfers to poor households conditional upon behavioral change that will improve development prospects of the next generation, such as sending children to school and bringing children to health centers.

Brazil's *Bolsa Família* (Family Allowance) is the world's largest CCT. As of 2010, it had 12.7 million beneficiary households (more than 49 million individuals) and an annual budget of $5 billion. *Bolsa Família* was created in 2003 by combining and redesigning previously existing cash transfer programs for education (*Bolsa Escola*), food and nutrition (*Cartão Alimentção* and *Bolsa Alimentção*), and cooking gas (*Auxilio Gás*). Its key components are monthly transfers per child or pregnant/breastfeeding mother, conditional on a school attendance rate of at least 85 percent for children age 7 to 15, vaccinations for children 0 to 6 years, and prenatal care visits for pregnant women.

Bolsa Família has had remarkable education and health impacts to date. It has significantly increased school attendance, rates of grade progression, school retention of children after the age of 14, prenatal care visits, the probability that a child receives all seven vaccines required by the age of 6 months, and healthy body mass index (BMI) among children.

Bolsa Família's success is not unique. Several other CCT programs have also helped both to reduce current poverty and to increase future opportunities for low-income families, such as *Bolsa Família*'s predecessor, Mexico's widely studied *Progresa/Oportunidades* program. Approximately thirty countries now have CCTs. Most large-scale CCTs to date have been in Latin America, which amplifies the potential impact of CCTs, given that this region has the world's highest degree of income inequality.

Three distinguishing characteristics of CCTs underlie their success and worldwide growth. First, CCTs creatively mitigate the trade-off between humanitarian relief and economic efficiency by moving beyond a focus on just short-term poverty alleviation and income redistribution to include promotion of long-term economic growth through investment in human capital development. This also addresses the criticism of unconditional cash transfers, such as that they produce perverse incentives that reduce labor supply, crowd out private transfers, and encourage dependency. Second, CCTs use incentives to encourage utilization of public services, particularly health and education services, designed to offer a pathway out of poverty. Third, CCTs are designed and implemented in the context of "evidence-based public policy," in contrast to traditional safey net programs; especially in developing countries, CCTs have been closely studied, exhaustively documented, and rigorously evaluated.

However, a word of caution is in order. CCTs have been most successful in middle-income countries with both a strong civil service and a relatively widespread network of local schools and health centers. It is difficult for conditional transfers to work if the specified conditions cannot be met easily by low-income families because of weak public sector institutional capacity and insufficient public infrastructure.

SOURCES: L. Aber and B. Rawlings, "North–South Knowledge Sharing on Incentive-Based Conditional Cash Transfer Programs," Special Protection & Labor Discussion Paper No. 1101, Washington, DC, World Bank, January 2011; L. Brenzel, "Brazil—Bolsa Familia Project: Results-Based Financing (RBF) for Health—Country Snapshot," Washington, DC, World Bank, November 2009; A. de Brauw et al., "The Impact of Bolsa Família on Education and Health Outcomes in Brazil," paper by the International Food Policy Research Institute presented at the Second Generation of CCTs Evaluations Conference, World Bank, October 2011; and F. Veras Soares, R. Perez Ribas, and R. Guerreiro Osório, "Evaluating the Impact of Brazil's Bolsa Família: Cash Transfer Programs in Comparative Perspective," *Latin American Research Review* 45 (2010): 173–190.

welfare recipients live in one-parent or no-parent families.[18] However, the causal link has been questioned. Out-of-wedlock births also increased dramatically among those not on welfare during the 1970s and 1980s (although they declined somewhat among both groups in the 1990s), and the incidence of out-of-wedlock births does not seem to be higher in states offering larger benefits. Indeed, in many states in which there has been a marked rise in such births, benefits come nowhere near covering the additional costs of a child. Although there may be some effect, welfare benefits simply cannot account quantitatively for the magnitude of the change.[19] Indeed, studies have not found effects of welfare on illegitimacy large enough to explain the observed increases in female-headed families.[20] To be sure, poverty itself contributes to family breakdown and welfare dependency.[21]

Although there has been extensive migration, especially from the South to northern urban areas, since AFDC was first adopted in the midst of the Depression, there is debate about the extent to which that migration was induced by AFDC, in say, northern states. Other factors were probably more important. Jobs were more available, pay was higher, and there was less racial and class discrimination. Although, from an economist's perspective, if one is going to be unemployed, simply receiving checks, it makes sense to do so where the checks are largest, migration does not seem to be particularly sensitive to changes in welfare benefits. This may be because most welfare recipients plan to return to work; job opportunities and other factors, such as proximity to friends, dominate. These, plus the costs of moving, more than offset the differences in welfare benefits among the states.

---

[18] In 2010, 53 percent of TANF cash assistance went to one-parent families and 42 percent went to no-parent families; only 5 percent went to two-parent families. (Source: House Ways and Means Committee, *2011 Green Book*, Table 7-10.)

[19] One interesting theory attributes the increased incidence of out-of-wedlock birth to the increased availability of abortions. Previously, a significant fraction of marriages were shotgun marriages—as evidenced by the number of firstborn children arriving, say, seven months or less after marriage. This number has decreased dramatically; the availability of abortions has meant that men no longer feel the moral obligation to marry, as the woman has a choice. See G. A. Akerlof, J. L. Yellen, and M. L. Katz, "An Analysis of Out-of-Wedlock Childbearing in the United States," *Quarterly Journal of Economics* 111, no. 2 (May 1996): 277–317.

[20] See R. Moffit, "Incentive Effects of the U.S. Welfare System"; David Ellwood and Lawrence Summers of the Harvard Kennedy School, found that variations in benefit levels across states were not associated with variations in divorce rates, illegitimacy, or percentages of children in families headed by single parents. See D. Ellwood and L. Summers, "Poverty in America: Is Welfare the Answer or the Problem," in *Fighting Poverty: What Works and What Doesn't,* ed. S. Danziger and D. Weinberg (Cambridge, MA: Harvard University Press, 1986), pp. 78–105.

[21] See J. Currie, "Welfare and the Well-Being of Children"; according to Moffit, "The results show consistent evidence of the strong correlations between parental welfare receipt and later behavior of daughters. Daughters of welfare families are much more likely to participate in the welfare system themselves at a later date, and are more likely to have births in general and premarital births in particular." See "Incentive Effects of the U.S. Welfare System," p. 36.

WELFARE REFORM: INTEGRATION OF PROGRAMS

This chapter began by noting that some of the dissatisfaction with welfare programs was misplaced; although they are blamed for the increasing deficit, they really had little to do with it. However, the discussion in the preceding section has uncovered several other complaints, about actual or perceived distortions in behavior induced by the system. In recent years, there have been major initiatives for reform—one aimed at simplifying and integrating the welfare programs, and another aimed at trying to "make work pay."

As we have seen, the current welfare system consists of a potpourri of programs. Some programs, such as TANF, originated to deal with particular problems. Others grew in part in response to special interest groups—for example, the food stamp program gave the agriculture industry an opportunity to increase the demand for their products, and thus the income of farmers.

Regardless of the origins of programs within the welfare system, their lack of integration has two negative effects. First, it greatly increases administrative costs and burdens. Each program has, for instance, a different eligibility form. The bureaucratic complexity discourages some who could really benefit from the programs from taking advantage of them. It has been estimated that only about two-thirds of those eligible for AFDC received it, only 60 percent of those eligible for food stamps received food stamps, and only 83 percent of those eligible for EITC actually collected it.[22]

The second negative effect of the lack of integration is that the programs interact in ways that increase the magnitude of the distortions. Each means-tested program has a phaseout, an interval of incomes over which benefits decline. The decline in benefits is equivalent to a tax. Even though legislators may be aware of this effect, as they consider each program in isolation they seldom think of the combined effects of all of these taxes. As we have seen, the combined effect may result in a total marginal tax rate well in excess of 60 percent. If all the programs were consolidated into a single cash program, not only would overall administrative costs be reduced, but attention would also be paid to the combined effects.

At various times, there has been support for integrating all of the welfare programs together, and "cashing them out"—allowing those on

---

[22] J. K. Scholz, "Tax Policy and the Working Poor: The Earned Income Tax Credit," *Focus* (Winter 1993–1994): 1–12.

welfare to spend the money however they wish. Such proposals amount to an expanded EITC. They are sometimes referred to as **negative income tax** proposals.

In this system, all individuals would be required to file tax returns, but just as only individuals above a critical threshold level would have to pay income taxes, those below that critical threshold level would receive a check from the government. Consider a tax regime in which everyone receives a check from the government of, say, $1000 per year, but then pays to the government, say, one-third of each dollar earned. Those with an income of less than $3000 would receive something net from the government; those with an income greater than $3000 would pay to the government more than they receive. Those who favor this system argue that it is not only administratively simpler, but it is also less demeaning than the present system, which forces individuals to present evidence to several agencies concerning their low income.

Advocates of a negative income tax, however, have argued for more than just the consolidation of current welfare programs and a conversion to cash benefits. They have argued for a change in the structure of total benefits, for two key reasons. First, the high marginal tax rates on the poor discourage work effort among people whose attachment to the labor force is already weak; second, the total benefits under the current system amount to less than what is required to remain out of poverty.

However, this gives rise to one of the fundamental dilemmas in welfare policy. If we want to provide benefits that are large enough to ensure a minimal, say, poverty, level of income even for those who do not work, and if we want the phaseout to be slow so the implicit marginal tax rate is low, then the benefits must extend high up into the income distribution. For instance, if the minimal benefit for a family of three were $12,000, and the marginal tax rate, including the full Social Security contributions of approximately 15 percent, were kept to 30 percent (ignoring state taxes), then benefits would extend up to families with an $80,000 income; if the marginal tax rate were increased to 55 percent—more than one out of two dollars earned by the individual thus going to the government—benefits would still extend up to $40,000. The costs of such an arrangement would be high—too high to be politically acceptable. Thus, either initial benefits must be below the poverty level or effective marginal tax rates must be very high.

BASIC TRADE-OFF IN THE DESIGN OF NEGATIVE INCOME TAX

Higher benefits for the very poor entail either greater expenditures (raising tax rates, and reducing incentives for higher-income individuals) or faster phaseouts, implicitly entailing higher marginal tax rates for lower-income individuals.

THE WELFARE REFORM BILL OF 1996

As we have noted, the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 marked a major departure from the past in two important ways. First, it replaced the AFDC system, in which the federal government paid a fraction of the costs, depending on the per capita income in the state, with the block grants of TANF. Second, it imposed a number of stringent requirements designed to encourage movement from welfare to work. Two beliefs underlay the reforms: one that the programs were too costly (partly based on the misperception noted earlier concerning the size and growth of welfare programs and their role in the growing deficit), or another that the programs had failed in their primary purpose—rather than being a safety net, for too many people, welfare had become a way of life.

BLOCK GRANTING

Although the states administered the AFDC program, many of its basic features were dictated by the federal government. One of the most important features of the 1996 welfare reform was to end the federal welfare entitlement. State matching grants—in which the federal government contributed, say, a dollar for every dollar spent by the state—were converted to block grants—fixed amounts of money—giving the states much more discretion in how they administered the funds. Critics worried, however, that converting to block grants would lead to reductions in expenditures. The 50 percent matching rate had effectively lowered the "price" of welfare. A state could get two dollars of welfare benefits for its citizens by spending only one dollar. Welfare, it was argued, was like any other good: lowering the price raised the quantity demanded; raising the price would lower the quantity demanded.

Critics worried further that there would be a "race to the bottom." Lowering a state's welfare benefits relative to those of neighboring states would provide incentives for those on welfare to move; each state would have an incentive to encourage those free riding on the welfare system to leave. The upshot of these incentives, critics argued, would be competition among the states to drive out those dependent on welfare.

Advocates of block grants argued that there was not strong evidence of a large elasticity of demand—that is, of a large responsiveness to price—so that fears of welfare cuts were exaggerated. Moreover, there were incentives to drive out welfare recipients under the old regime: the states still

paid half of the cost of the benefits, although to be sure, these incentives would be enhanced if the states had to pay, at the margin, 100 percent of the benefits. Nonetheless, it was hard to see evidence of a race to the bottom; even though there were differences in benefit levels among states, they were related more to the states' income and demographics. Moreover, advocates of block grants argued, the increased efficiency resulting from state control and the elimination of federal bureaucratic requirements would increase the benefit received per dollar spent—at least partially offsetting the fact that the federal government was no longer picking up 50 percent of the tab.

To address concerns that there would be an excessive reduction in welfare support, proponents of block grants agreed to a "maintenance-of-effort provision," under which states receiving the block grants would agree to continue spending at least 75 percent of the amount that they had previously spent on welfare. Advocates argued that with the greater efficiency resulting from the devolution of responsibility to the states, the effective level of welfare benefits would not be cut at all. But critics raised two objections: Advocates of block grants exaggerated the negative impact of federal regulations and bureaucracy, so that any major cut in expenditures would be felt.

Moreover, with time, the adverse impact of block grants would worsen, particularly in states in which for one reason or another there was a marked increase in the welfare load (for instance, because of an economic downturn). If, under the old regime, benefits would have expanded by only 15 percent, then a 75 percent maintenance-of-effort requirement would be equivalent to more than a one-third cut in benefits, relative to what they otherwise would have been.

Furthermore, hidden in the 1996 bill were provisions that allowed funds spent on other social programs directed at the poor to be counted toward the maintenance-of-effort requirement, substantially lowering the effective maintenance-of-effort requirement.

ANALYTICS OF STATE RESPONSES TO BLOCK GRANTS

We can use some simple diagrams to illustrate why economists worried that converting from a matching system for welfare to a block grant would reduce overall support for welfare.

Figure 15.6 shows how the states might be anticipated to respond. *BB* gives a state's budget constraint between welfare and other expenditures before the federal program, and shows a hypothetical state indifference
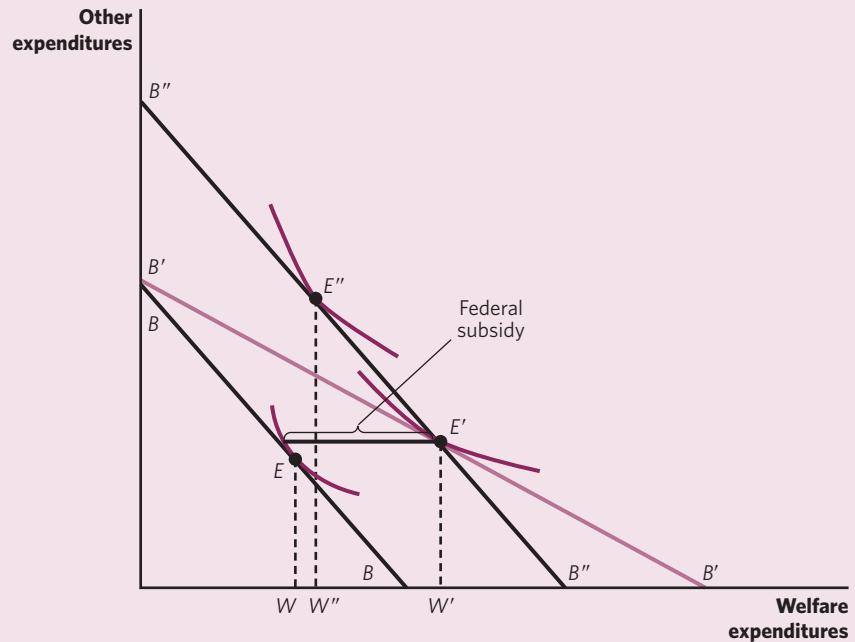
curve, depicting the trade-off between welfare and other expenditures. The state chooses point *E*, maximizing its (economic) welfare, subject to its budget constraint.

BB' gives a state's budget constraint under the former system of federal matching funds. The state is induced to spend more on welfare, both because of an income effect (the state is better off as a result of the federal subsidy) and a substitution effect (for each dollar of increased welfare expenditure, the state has to give up only 50 cents of nonwelfare expenditures). The state now chooses point *E'*. The increased funding for welfare and the lower price both lead to more expenditure on welfare than before. There are positive income and substitution effects.

B"B" now gives the budget constraint when the federal subsidy is converted into a lump sum, not dependent on the amount spent. The state is better off, but unless the income effect is large, expenditures on welfare are decreased; although, because of the income effect, they remain higher than with no subsidy. The figure shows what has been the concern of critics of block granting: a large decrease in welfare expenditures. (However, note that, as depicted, the indifference curve in the new equilibrium is at a higher level, reflecting the general proposition that nondistorting subsidies are more efficient.)

TIME LIMITS

The 1996 welfare reform introduced time limits into the welfare programs. Federal rules stipulate that federal TANF funds may not be received by a family that includes an adult who has received sixty months of TANF funds previously. However, a state may exempt up to 20 percent of its caseload from the five-year limit based on hardship. States may continue to assist all families beyond five years if they choose; those funds will be counted toward the states' maintenance-of-effort requirements. Many states adopted a shorter two-year period. Advocates of these time limits hoped not only that they would push people off the welfare rolls, but also that they would discourage people from joining them in the first place. The prospect of having to re-enter the labor market five years down the road would motivate prospective welfare recipients to remain in the labor market, where they could maintain and expand their job skills. Moreover, because employers prefer to hire workers with current job experience, remaining in the labor force would enhance the long-term employment prospects of employees.

Evidence supporting this theory came in the months immediately following the passage of the welfare reform bill, during which time welfare rolls decreased substantially—to 8.4 million recipients by June 1998—not only undoing the large increase that had occurred in the early 1990s, but also putting the number well below the 1990 level of 11.46 million.[23] Although improved economic conditions—the continuing decline in the unemployment rate to a three-decade low of just over 4 percent—explained a large part of the overall decline, the remaining decrease seems attributable, at least in part, to the 1996 welfare reform.[24]

MANDATORY WORK

In addition to time limits, the 1996 reform required states to impose work requirements. Adults had to engage in some form of work after a maximum of two years of TANF benefits, and to participate, unless the state opted out, in community service after two months. This work requirement does not apply to single parents of children under age 6 who cannot obtain child care.

---

[23] House Ways and Means Committee, *1998 Green Book*.

[24] The Council of Economic Advisers concluded that the 20 percent decline in welfare caseload during the 1993–1996 period can be attributed to the stronger economy (40 percent), welfare reform policies (30 percent), and the other factors such as the earned income tax credit (the remainder). Council of Economic Advisers, *Explaining the Decline in Welfare Receipt, 1993 to 1996* (Washington, DC: Government Printing Office, May 1998).

The work can be unsubsidized private employment, subsidized private employment, public employment, on-the-job training, community service, vocational education, or secondary education. To be considered working, the individual had to work for at least a weekly minimum average of twenty hours in 1997–1998, rising to thirty hours after the year 2000.

THE WELFARE REFORM DEBATE OF 1996

The debate over the 1996 reform took up a wide range of questions about values, political judgments, and economic assumptions.

WELFARE RECIPIENTS AND LABOR MARKETS   Did welfare recipients have the foresight—without further training and education—to realize that it was in their best interest to remain in the labor market? A lack of such foresight would undermine the intended incentive effects of the time limit.

Did welfare recipients have the skills required to be effective members of the labor force, or were many simply unsuitable for participation in the labor force? Welfare advocates argued that more training funds were required. Indeed, they proposed that meaningful reform would cost more in the short run than simply continuing the old system, but the bill that passed made minimal provisions for training. Advocates of the reform bill countered that "soft skills," such as the basic ability to get up and go to work every morning, were most important, and that these skills were best acquired in the labor market.

Some in the policy debate worried that many welfare recipients could not function effectively in the labor market. These advocates argued for large exceptions from the work requirement. More ardent welfare reformers objected that these large exceptions would undermine the whole reform effort. Those who really could not function in the labor market should be treated as disabled, and be subjected to the same disability standards as everyone else.

Was it unwillingness to work or a lack of jobs that led people to welfare? Could the labor market provide low-skilled jobs in sufficient numbers? Most economists believed the labor market had the capacity to create jobs—after all, between 1993 and 1996, 11 million jobs had been created—and that it would create jobs that corresponded to the skills that were available. They also believed that, at least in many locations, the lack of participation in the labor force was not due to a shortage of jobs: immigrants, with few skills, managed to get low-skilled jobs ranging from work in fast-food restaurants to driving for taxicab services. The problem was neither lack of jobs nor lack of training, but the lack of the soft skills noted earlier.

But in the first decade of the twenty-first century, the pace of private sector job creation greatly slowed to the point where one out of six Americans who would have a full-time job could not get one, and those with low skills typically went to the back of the job queue. An increasing fraction of jobs paid wages that left families in poverty.

TIME LIMITS   What would happen when the time limits expired? Should children suffer because their parents were unable, or unwilling, to work? One option, called *workfare*, would be for government to provide jobs in exchange for welfare. But this raises a host of issues: How much would workfare jobs pay? Would they undermine employment opportunities of other people participating in the work force? Would workfare jobs negate the incentive for welfare recipients to find regular employment?

SOCIAL WORKERS IN THE WELFARE SYSTEM   In the absence of incentives, would the workers who ran the old welfare system—who, critics said, were more in the business of sending out checks than of getting welfare recipients to work—be transformed into job placement officers? Or, if welfare recipients were to be moved to work, would greater reliance have to be placed on private placement agencies?

Those who focused on the present system's failures were not sanguine about the ability of the welfare system to transform itself. Without adequate incentives, they saw current social workers as simply perpetuating the current system, perhaps under some other guise. Not surprisingly, those who saw less of a necessity for reform adamantly opposed turning over social services to for-profit agencies. They were less concerned about the current system's inadequacies than about the potential of for-profit agencies to exploit the poor.

CHILD CARE   Who would take care of the children while a welfare mother was working? The welfare reform bill made some provision for child care services, but far less than was required to cover child care expenses. On the other hand, lower-income mothers in working families were struggling to find adequate child care arrangements. Did it make sense to provide child care arrangements for former welfare recipients, but not for low-income mothers who had never been on welfare? In a time of severe budget stringency, few believed that child care costs could be provided to all low-income mothers. Indeed, in many cases the government would be paying child care costs that exceeded the earnings of those receiving the benefits. It was precisely this dilemma that resulted in the compromise under which some, but limited, child care expenses were provided for welfare mothers.

In the end, the policy debate—about consequences and values—was subsumed under the political debate: the strong desire of many Americans to see significant reforms, and the desire of the state governors to assume more

control over their welfare system. The reductions in welfare rolls in the years immediately following passage of the 1996 reform bill provided some validation for some of the arguments underlying the reform. However, two questions remained: What would happen if and when the economy went into a prolonged economic downturn? What would happen to the "hard-core" welfare recipients, amounting to perhaps a quarter or more of the beneficiaries—those with few skills and little interest or ability to acquire them? Two contrasting pictures are painted for the future. One sees welfare mothers begging on the street, an image of America that harks back to the nineteenth century and across to less developed countries. Another sees welfare restored to its original intent of temporary help in times of adversity.

Although it is too soon to assess the long-term impact of the current "Great Recession" on the poor, and the role of the social safety net in mitigating this impact, preliminary data indicate mixed results in assisting

THE PERSON OR THE PLACE?

This chapter has looked at strategies to reduce poverty that focus on people—on transferring cash and in-kind benefits to low-income families. However, poverty in the United States is concentrated geographically, and there is a place-based strategy designed to improve living conditions and job prospects in those areas.

Poverty feeds on itself. High crime rates, substance abuse, weak family structures, poor schools, and limited job opportunities combine to perpetuate poverty. From 1993 to 2000, Congress funded the **Empowerment Zone (EZ)**, Enterprise Community (EC), and Renewal Community (RC) programs* to stimulate community development and business activity in impoverished census tracts. National competitions were established during three rounds of funding to designate approximately 200 urban and rural communities as EZs, ECs, or RCs. In addition to size (population and area), indicators of need determined which communities were qualified to compete. These included poverty level, unemployment rate, and other signs of "general distress," such as high incidence of crime or narcotics use, amount of abandoned housing, number of people on welfare, average years of school completed, and school dropout rate. Designated communities developed long-term strategic plans to promote job creation and economic growth, and received a combination of grants, loans, technical assistance, and preferential tax treatment to implement these plans. An integrated, cross-sectoral set of public investments in social infrastructure and services, coupled with wage credits and investment incentives, were used to leverage private sector investment and foster public–private–community partnerships.

Under the supervision of the Department of Housing and Urban Development for urban areas and the Department of Agriculture for rural areas, each department of the federal government was charged with devising programs to help meet the needs of these special communities. For example, the Department of Justice proposed enhanced law enforcement programs, and the Department of Transportation was charged with ensuring that public transportation systems enabled those who lived in the designated areas to travel to places

where there were jobs. Federal, state, and local communities were charged with finding ways to change regulations to facilitate the creation of new enterprises. A set of community development financial institutions (CDFIs) was to be created, which would help channel private money to these new enterprises. Tax credits would encourage the creation of jobs within EZ/EC/RC areas and the employment of people living in these areas in jobs both within and outside their neighborhoods. These job incentives were perhaps the most controversial part of the proposal. There was concern that employers outside the designated areas would not be induced to hire more workers from these communities; they would simply look through their list of existing workers, to see who was eligible. Some argued that the focus should be on job creation within the designated areas, whereas others, pointing out that most of America does not live where

it works, asked why the EZ/EC/RC areas should be any different.

The EZ/EC/RC programs were predicated on the belief that building up communities requires more than just economic development, but that economic development can occur most effectively if there is a broader array of support, both within the community and from all levels of government.

By the end of 2011, all EZ, EC, and RC programs had expired, and legislation was being considered for a similar Growth Zones program to be administered by the Department of Commerce. Although some EZ, EC, and RC areas have shown substantially improved indicators of poverty, unemployment, and economic growth, most studies have found modest impact at relatively high cost, and it has been difficult to ascertain whether the improvements have been caused by these programs or by external factors.

*ECs were located in smaller cities and smaller rural areas than EZs; RCs had indicators of greater poverty than EZs or ECs, including some areas that remained distressed after having previously received an EZ or EC designation.

SOURCES: Office of Community Renewal, Department of Housing and Urban Development, www.hud.gov; Joint Committee on Taxation, *Incentives for Distressed Communities: Empowerment Zones and Renewal Communities*, Report No. JCX-38-09, October 2009; Government Accountability Office, *Revitalization Programs: Empowerment Zones, Enterprise Communities, and Renewal Communities*, March 2010; D. Bondonio and R. T. Greenbaum, "Do Local Tax Incentives Affect Economic Growth? What Mean Impacts Miss in the Analysis of Enterprise Zone Policies," *Regional Science and Urban Economics* 37 (2007): 121–136; M. Busso and P. Kline, "Do Local Economic Development Programs Work? Evidence from the Federal Empowerment Zone Program," Cowles Foundation Discussion Paper No. 1638, February 2008; E. L. Glaeser and J. D. Gottlieb, "The Economics of Place-Making Policies," *Brookings Papers on Economic Activity* (2008): 155–239; and J. A. Elvery, "The Impact of Enterprise Zones on Resident Employment," *Economic Development Quarterly* 23 (February 2009): 44–59.

the most vulnerable. For example, from 2000 to 2010, the poverty rate rose from 11.3 to 15.1 percent for all races, 22.5 to 27.4 percent for blacks, 21.5 to 26.6 percent for Hispanics, and 16.2 to 22.0 percent for children. From 2000 to 2009, SNAP also grew, with participation increasing from 17.2 to 40.3 million people and benefits from $18.8 to $50.4 billion in constant 2009 dollars. However, welfare rolls actually declined during the same period: recipients of cash assistance dropped from 6.1 to 4.4 million people, while federal and state expenditures on cash assistance under AFDC and TANF declined in constant 2009 dollars from $14.0 to $9.3 billion. This has created a growing group of "working poor"—those who earn too much to qualify for TANF but not enough to support themselves.[25]

[25] U.S. Census Bureau, *Historical Poverty Tables*, Tables 2 and 3; Food and Nutrition Service, United States Department of Agriculture, *SNAP Annual Summary*; and *2011 Green Book*, Tables 7–3 and 7–9.

Evidently, TANF no longer serves the original role intended—providing a safety net for those with children. Fortunately, SNAP has partially filled the gap. However, the fact that such a large fraction of the American children are in poverty, through no fault of their own, is deeply disturbing to many, and sets the United States apart from most other advanced countries that have been far more successful in reducing childhood poverty.

CONCLUDING REMARKS

The public assistance programs described in this chapter have almost surely helped reduce poverty in the United States. Although in some instances they may have contributed to the vicious cycle of poverty, by removing incentives for recipients to be actively engaged in the labor force, more broadly they have been fighting a rising tide of inequality that has affected not only the United States, but many other countries as well. This long-term trend—in evidence now for more than two decades— is largely related to the increasing premium placed on skills in the labor market. Since President Johnson, there has been an increased emphasis on getting at poverty's root cause by improving educational and job opportunities (see Chapter 14), beginning with programs such as Head Start, which focuses on preschool children, improved grade school and high school education, job training programs, and programs to facilitate the movement from school to work and to increase access to higher education.

REVIEW AND PRACTICE

SUMMARY

1. Public assistance provides cash and in-kind benefits to the poor. Expenditures for in-kind transfers have grown rapidly in recent years, but cash assistance (as a percentage of government expenditures) has fallen. Increases in welfare programs have played a relatively small role in the increase in the overall federal deficit.

2. The earned income tax credit, greatly expanded in 1993, increases the incentives for working (participating in the labor force), although for those with incomes above a certain level, it decreases incentives to work longer hours.

3. Welfare benefits decrease as income increases; the decrease in benefits has the same effect as a marginal tax rate, discouraging work effort. Programs with more generous benefits for

the very poor must either cost more or have faster phaseouts. Faster phaseouts are associated with higher marginal tax rates, and thus greater work disincentives. Eligibility thresholds have particularly adverse incentive effects.

4. The in-kind redistributive programs have several disadvantages:

   a. They are administratively costly.

   b. In some cases, they have only an income effect (i.e., they have the same effect as a transfer of cash); in other cases, they have a substitution effect, and in those cases, the government could make the poor better off at less cost through a cash subsidy.

   c. The effect of many eligibility standards is to discourage work and, when compounded with payroll and state income taxes, can result in very high marginal tax rates.

   d. The structure of eligibility standards provides unintended results; for instance, the food stamp program subsidizes the consumption of housing.

   e. The programs are paternalistic. But, in spite of the concerns, such programs continue to play an important role, partly because they may succeed in increasing spending in ways that are viewed as socially desirable, such as adequate housing and nutrition. There are also political arguments in favor of such assistance. Moreover, behavioral economists have suggested that the effects of the programs differ from those predicted by the standard economist's model.

5. Categorical programs have similar disadvantages:

   a. They are administratively costly.

   b. They are viewed by some as inequitable, as individuals with the same income may be treated differently.

   c. They are sometimes distortionary, as individuals attempt to qualify for subsidies.

   However, when groups differ in their labor supply responses (or other responses) to government programs, the government may be able to obtain a higher degree of redistribution, for the same loss of inefficiency, by providing categorical aid.

6. The 1996 welfare reform ended the federal entitlement to welfare, converted the system of matching grants to the states to block grants, and put time limits on welfare. While there was extensive movement from welfare to work, TANF seemed to provide (or offer safety net) for poor families with children. SNAP took up an increasingly important role, but there are still a large fraction of children in poverty. Critics of the 1996 bill argue that it did not provide adequate assistance for training, while advocates argue that what is required is not expensive training but the acquisition of "soft skills"—good work habits—that are best acquired in the workplace.

7. The lack of integration of the various welfare programs results in high marginal tax rates, with strong adverse incentive effects. Although there are proposals to "cash out" benefits and consolidate programs, given fiscal constraints, either benefits for those with little income will fall far below the poverty threshold, or generous initial benefits will have to be phased out quickly, resulting in high marginal tax rates.

KEY CONCEPTS

Aid to Families with Dependent Children (AFDC)

Block grants

Cash welfare program

Categorical programs

Children's Health Insurance Program (CHIP)

Conditional cash transfers (CCTs)

Earned Income Tax Credit (EITC)

Eligibility requirements

Empowerment Zone (EZ)

Entitlement programs

Food stamps

In-kind benefits

Low-Income Home Energy Assistance Program (LIHEAP)

Matching programs

Means-tested

Negative income tax

Poverty rate

Public assistance

Safety net

Social insurance

Social protection

Special Supplemental Nutrition Program for Women, Infants, and Children (WIC)

Supplemental Nutrition Assistance Program (SNAP)

Supplemental Security Income (SSI)

Temporary Assistance for Needy Families (TANF)

Threshold test

Welfare

Welfare lock

Workfare

QUESTIONS AND PROBLEMS

1. It has sometimes been suggested that the government should restrict the use of SNAP assistance to "healthful" foods. Discuss the merits of this proposal. Assuming that it would be easy to distinguish between "healthful" and "unhealthful" foods, describe the effect of such a restriction on an individual's consumption of the two kinds of foods.

2. Consider a welfare program (such as housing) with an eligibility standard that requires that an individual's income be below some threshold level. Draw the individual's budget constraint with and without the subsidy (put labor on one axis, consumption on the other).

3. Consider a welfare program (such as SNAP) with benefits that decrease as an individual's income increases. Draw the individual's budget constraint with and without the subsidy. (Put hours of work on the horizontal axis, and income on the vertical axis.) Use the diagram to illustrate how work incentives are reduced and how a fixed dollar subsidy could lead the individual to the same level of utility at lower dollar cost.

4. There have been proposals for more extensive use of government subsidies to help poor individuals purchase private housing (just as the government's SNAP assistance helps them purchase food). Discuss the merits of private versus public provision of housing.

5. Several different proposals have been put forward concerning how housing subsidies should be provided. Discuss the merits of:

   a. The government's paying a given fraction of the family's housing expenditures, up to some maximum, with the percentage depending on the family's income.

   b. The government's paying a fixed dollar amount of housing allowance, the amount depending on the family's income.

   In both cases, discuss the consequences for the family's expenditures on medicine; on food.

6. Assume you were particularly concerned with the welfare of children. How would this affect the kind of welfare programs you might support or how you might design your welfare programs?

7. a. Draw the budget constraint of an individual facing a negative income tax with a constant marginal tax rate of, say, 30 percent, assuming the individual receives $1000 if he or she does not work (and has no other source of income). Assume that the basic "grant" is increased to $2000 and the marginal tax rate is increased to 45 percent. Draw the new budget constraint. Explain why (1) very poor individuals are likely to work less under the second regime, and (2) middle- and upper-income individuals may work more or less.

   If the government were particularly concerned about the lack of work incentives of the poor, but still wanted to provide a high level of basic support, it could have a marginal tax rate of, say, 30 percent, up to $30,000 of income, and then impose, say, a tax rate of 55 percent for higher incomes. Compare work effort in these two regimes. Who is likely to work harder? Less hard? Why might you be more concerned with the reduced work effort

of upper-income individuals compared to the increased work effort of lower-income individuals?

b. The size of the EITC depends on family size. Explain why it might make sense for the increase in benefits with a third child to be less than the increase in benefits from a second child. (Families with no children get no benefits.) What incentives might such a structure have for families breaking up?

c. Explain why, even taking into account Social Security taxes, two poor people can be made better off if they hire each other to look after their children. Would it pay them to hire each other to do nothing?

8. Several proposals have been made to encourage employers to hire former welfare recipients, but there has been concern that such subsidies will lead to welfare recipients' simply displacing other unskilled workers from their jobs.

a. Using standard demand and supply diagrams, assume that former welfare recipients are identical to other unskilled labor, and that work requirements simply shift the supply curve of unskilled labor. What happens to employment and wages of unskilled workers?

b. Now assume that there is a minimum wage that is above the equilibrium level of wages for unskilled workers. What happens to employment and wages of unskilled workers? What happens to the unemployment rate?

c. Now assume that the government provides tax subsidies to employers who hire former welfare recipients. Will this affect the total number of unskilled workers hired? Will it affect *which* unskilled workers get hired?

d. Now assume the government provides tax subsidies for the employment of *all* unskilled workers. What happens to employment, wages, and the unemployment rate?

e. Now assume that welfare recipients are slightly less productive than other unskilled workers, say, because they have been out of the labor market; if they return to the labor market for a short period of time, say, six months, their productivity becomes equal to that of other unskilled workers. If there is a minimum wage, but no special tax provisions for welfare recipients, what will happen to employment of welfare recipients? What will be the consequences of a short-term subsidy for hiring former welfare recipients?

9. Calculate the slopes of different segments of the stylized budget constraint of Figure 15.2C. What other shapes might the budget constraint take?

16

SOCIAL INSURANCE

Modern governments have long taken some responsibility for providing for the needy and providing social protection to the aged, the unemployed, and the disabled, but during the past fifty years this has come to be viewed as one of the primary functions of government. In the United States in 2009, social insurance and public assistance expenditures represented more than 40 percent of total government spending. Social Security, however, differs from most government programs in that it has its own earmarked payroll tax. The revenues from this tax go into special trust funds that finance benefit payments. Social Security was originally designed to be self-financing—that is, revenues from the payroll tax were intended to cover outlays.

Of the major social insurance programs, by far the largest is Old-Age, Survivors, and Disability Insurance (OASDI), enacted in 1935. This is usually referred to as **Social Security** and is intended to provide a basic standard of living to the aged, the disabled, and their survivors. As Table 16.1 indicates, over the past four decades this program has more than quadrupled in real terms (in 2005 dollars), from $138.6 billion in 1970 to $620.8 billion in 2009. Disability—technically, also included under Social Security—provides money to those no longer able to work. The second largest program,

TABLE 16.1 EXPENDITURES IN MAJOR SOCIAL INSURANCE PROGRAMS (BILLIONS OF 2005 DOLLARS)

	1970	1990	2009
Social Security (OASDI)	138.6	349.2	620.8
Old-Age and Survivors Insurance*	124.8	313.9	510.8
Disability Insurance	13.8	35.3	110.0
Unemployment Insurance	13.1	23.7	108.6
Medicare	24.2	129.7	388.9

*Most of the expenditures on Old-Age and Survivors Insurance go to provide retirement income, but there are also benefits for widows and children.

SOURCES: Executive Office of the President, Office of Management and Budget, *Budget of the United States Government, Fiscal Year 2013, Historical Tables*, Table 8.6; and *2010 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds*, August 9, 2010, Tables VI.A2, VI.A3, and VI.A4.

FOCUS QUESTIONS

1. What is social insurance, and why does government provide it? What market failures in particular provide the rationale for Social Security?

2. What are the financial problems facing Social Security programs, and why do they seem so hard to remedy? What are the merits of various proposals for financial reforms?

3. What are major inequities and inefficiencies associated with the design of the Social Security system, and how might they be remedied?

4. How does the U.S. Social Security system compare with government pension programs in other countries?

Medicare, provides medical care for the aged; it was discussed in detail in Chapter 13. Unemployment insurance, also enacted in 1935, is intended to provide income to individuals during short-term spells of unemployment (as its name suggests). Other social insurance programs include workers' compensation, which provides money to individuals who are injured on the job; disability benefits for veterans; retirement benefits for railroad workers; and funds for coal miners suffering from black lung disease.[1]

In many ways, social insurance programs provide insurance to individuals against particular risks that they face, just as private insurance does. Thus, Medicare covers medical expenditures of the aged, just as a private health policy would. Social Security is designed to replace a part of the income lost due to retirement or disability; private insurance policies exist that meet the same need.

There is one important difference: with private insurance there is a close relationship among individuals' payments, the risks they face, and what they receive. Thus, the premium for a private health insurance policy depends on factors affecting the individual's health condition, such as age. The amount that an individual receives back from an **annuity** (a private insurance policy providing a certain income every year after the individual reaches, say, age 65), on average, is effectively just what he or she puts in (plus accumulated interest). This is not true of social insurance. Social insurance programs provide insurance *and* redistribute income. Confusion between these two roles has been a major impediment in the evaluation and reform of social insurance programs.

[1] Like many private employers, the federal and state governments provide additional retirement benefits to their civilian employees and military personnel, which should probably be viewed as deferred compensation rather than as social insurance programs.

Supplemental security income (1.0%)
Other retirement (6.7%)
Miscellaneous (1.2%)
Veterans benefits (2.6%)
Housing assistance (1.0%)
Medicaid (6.9%)
Medicare (35.8%)
Social security (44.8%)

**Federal Government Spending for the
Elderly, 2010 (Projected): $1.1 trillion**

The Social Security system has been expanded greatly in the 75 years since its enactment. Originally, it covered only a fraction of the working population, with agricultural workers, the self-employed, government employees, and employees of nonprofit institutions excluded. Today, the only employees not covered are federal employees hired before 1984 and a few categories of employees of state and local governments. As a result, the number of beneficiaries of OASDI increased approximately 2.5 times from 1965 to 2009, from 20.2 to 51.9 million; in real terms (in 2005 dollars), the benefits have increased over more than sevenfold, from $86.0 billion in 1965 to $620.8 billion in 2009.[2] Note that a very large fraction of our social insurance system, including health, is aimed at the elderly, and that the "Social Security" program itself is only 45 percent of total federal outlays for the elderly (see Figure 16.1).

This chapter discusses the major issues facing the Social Security (OASDI) program, but focuses exclusively on the retirement (as opposed to the disability) part of this program.

THE SOCIAL SECURITY SYSTEM

OASDI is financed by a payroll tax that is paid partly by employees and partly by their employers. In 2009, the combined tax rate was 12.4 percent on the first $106,800 of income. In addition, there is a 2.9 percent tax to finance

---

[2] See *2010 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds*, August 9, 2010, Table IV.B2; and Executive Office of the President, Office of Management and Budget, *Budget of the United States Government, Fiscal Year 2013, Historical Tables*, Table 8.6.

Medicare on all income. (The inclusion of all income for the purposes of the Medicare tax was one of the changes instituted in 1993.) Both the maximum base and the tax rate have increased over time. Beginning at $3000 and a 2 percent combined employer/employee tax in 1937, the base more than doubled by 1966, to $6600 with an 8.4 percent combined rate; doubled again by 1974 to a $13,200 base and an 11.7 percent combined rate; and doubled again in the next seven years to $29,700 with a 13.3 percent rate.

According to the law, half the tax is paid by employees and half by their employers—but most economists believe that this is simply a legal fiction. The consequences of the tax are essentially the same as they would be if the individual were responsible for paying all of it. What difference should it make who mails the check to the government?[3]

A pension system in which each age group's pension is supported by its own contributions is called a **fully funded system**. Private pension systems normally are fully funded; while they are working, individuals contribute to a fund that is used to provide for their pensions in retirement. By contrast, the Social Security system is organized on a **modified pay-as-you-go basis**. In a pure pay-as-you-go system, the payroll taxes of those working today pay for the benefits received by the elderly today. The U.S. system is called a *modified* pay-as-you-go system because revenues and expenditures are supposed to balance out not each year, but over a seventy-five-year horizon. The balance between receipts and expenditures is added to or subtracted from the Social Security trust fund.

The benefits an individual receives are linked to his or her contributions by a complicated formula. The more an individual contributes, the more he or she gets back—just as in the case of private insurance. In the case of private insurance, however, an individual who spends $10,000 to buy an annuity (a retirement policy that pays, say, a fixed amount every year) will get approximately ten times the benefit of someone who spends $1000. This is not the case with Social Security. Social Security is not just a pension program; it is a redistribution program. Poorer individuals (or those who have made smaller contributions) get back proportionately more, and married couples in which only one individual worked get back proportionately more than do couples in which both worked and earned similar incomes. Moreover, because women live longer than men, women receive back more per dollar contributed than do men. This is true under private retirement programs as well.

---

[3] Employers are concerned only with their total labor costs, including any employment taxes; employees are concerned only with their net after-tax income. The government's revenue is simply the difference between the two, and it makes little difference who the government says is paying the tax. However, for very–low-paid workers who are receiving the minimum wage, the employer share of Social Security does have the effect of increasing their cost of labor; the impact would be different if workers bore all the costs. For a more extensive discussion, see Chapter 18.

NET LIFETIME SOCIAL SECURITY TRANSFER

In the initial years of Social Security, net transfers (the difference between benefits and contributions) increased significantly, but in recent years they have declined, and that trend is expected to continue. Net transfers have always been highest for one-earner couples. Net transfers used to be largest for high-wage workers, but this is no longer true.
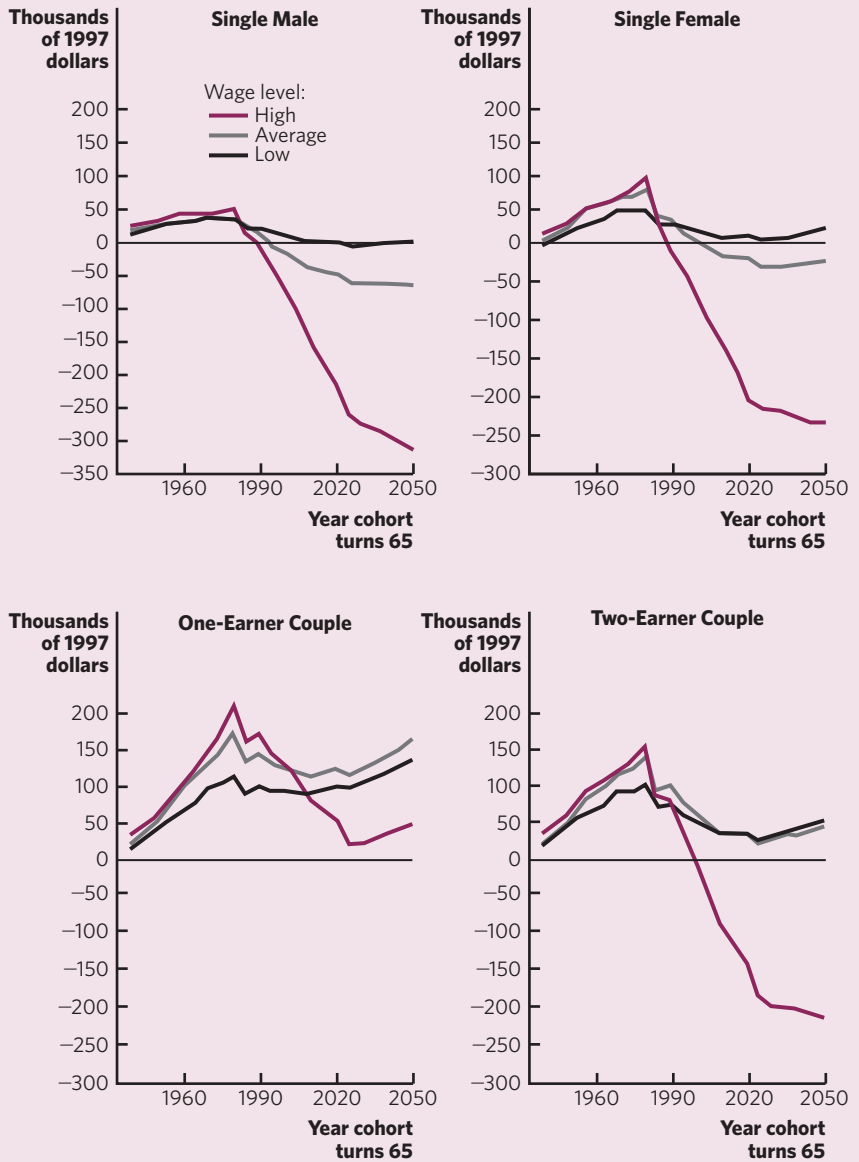
The system has changed dramatically over time. Figure 16.2 illustrates how the difference between the present discounted value of payments and receipts has changed and is projected to change in coming years.[4] Until 1983, essentially all groups received more than they contributed—this was just a transfer from the younger generation to the older. Interestingly, high-wage workers used to receive a larger transfer than did low-wage workers: although the *ratio* of their benefits to payments was smaller, the absolute amount of the transfer was larger. This is still true for one-earner families, but for high-wage single individuals and high-wage two-earner families, benefits are less than receipts. For two-earner families, it will be a couple of decades before the transfer to low-wage workers equals that to middle-income workers.

It also used to be that the present discounted value of benefits increased if one retired early; today, those who retire early have their benefits reduced sufficiently so that there is no incentive to retire early.[5]

Thus, the Social Security system redistributes income toward the poor, to one-earner families, and, on average, from the current working generation to present retirees.

SOCIAL SECURITY, PRIVATE INSURANCE, AND MARKET FAILURES

Before 1935, private markets provided life insurance but not retirement insurance. Few firms provided much in the way of pensions.

The Great Depression caused a crisis: there were many aged individuals who were thrown out of work and had little prospect of being rehired and no means of support. The Social Security system was intended to ensure that all the aged had at least a minimal level of support.

In subsequent years, however, there were improvements in private markets. Pension coverage expanded rapidly in the 1950s when large manufacturers adopted pension plans. Whereas in 1950 only 25 percent of nonagricultural workers in the private sector were covered by private

---

[4] The figures depict the net transfer to someone *retiring* at age 65 in that year. They illustrate the present discounted value (PDV) of benefits minus the present discounted value of contributions, using a 2 percent real interest rate. They take into account chances of death and "typical" family structures. Following the Great Recession of 2008, real interest rates fell to 2 percent; using these numbers, obviously, the PDV of benefits is much higher. As we discuss below, the Social Security system has insulated the elderly from such fluctuations in market interest rates.

[5] Matters are slightly more complicated: the life expectancy of the poor is shorter than that of the rich, so the expected benefits of a poor person must be adjusted to reflect these differences. By the same token, someone who foresees a low life expectancy has an incentive to retire earlier, and someone who foresees a long life expectancy has an incentive to retire later.

pensions, by 1979, 55 percent were covered. In the early 1980s, pension coverage declined to 50 percent as the number of jobs in manufacturing declined and the poorly covered service sector grew, and coverage has remained at that low level. Today, 51 percent of all workers participate in employer-provided retirement plans, with coverage ranging from 68 percent for white-collar occupations to 25 percent for service-sector occupations.[6] The government has taken steps to ensure the financial soundness of private pension programs[7]; however, these private insurance policies are deficient in several respects.

HIGH TRANSACTIONS COSTS

To provide for their retirement, individuals can purchase annuities from private firms. Annuities pay a fixed amount every month from some age (usually 65 or 70) until the individual dies, no matter how long he or she lives. Under most private annuity programs, however, the expected rate of return does not appear to be very good—far lower than market rates of interest. This is partly because of high administrative costs, including, in many cases, substantial commissions for the salespeople. Whereas administrative costs for Social Security are less than 1 percent of benefits paid, those of private pensions are almost 6 percent.[8]

One of the reasons for the high transactions costs is that insurance companies are always trying to "skim the cream"—recruit the best risks and identify the bad risks, which they will either not insure or insure but only at a high premium. This is expensive.

Moreover, many insurance companies spend a great deal on salesmanship, partially in an attempt to "fish for fools"—to find those individuals who do not understand risk and who are willing to overpay for the insurance they get. The insurers have discovered that many individuals will pay a great deal for a small risk—an amount well in excess of the actuarial value of the loss—which they could easily manage.

---

[6] U.S. Bureau of Labor Statistics, *Employee Benefits in Private Industry in the United States*, March 2008.
[7] The government not only regulates private pensions, but also provides insurance to ensure that workers receive the promised pensions. A number of large defaults on the private pension schemes have threatened the financial viability of the federal insurance program. The program run by the U.S. Pension Guaranty Corporation had a deficit of $2.7 billion in 1992. In response, Congress passed the Retirement Protection Act of 1994. In 2009, the program covered 44 million workers and retirees, and had a deficit of $22.5 billion. See House Committee on Ways and Means, "The Pension Benefit Guaranty Corporation," Chapter 12 in *2011 Green Book*.
[8] See Department of Labor, *Pension and Health Benefits of American Workers,* Table A15; and *2012 OASDI Trustees Report*, Table III.A6. Interestingly, when Social Security is compared to private insurance companies, including annuities, government looks even more favorable. Private insurance companies spend one dollar in administrative costs, dividends (profits), and taxes for every two dollars in benefits paid. See Charles T. Goodsell, *The Case for Bureaucracy* (Chatham, NJ: Chatham House Publishers, 1983), p. 52.

RISK MITIGATION

The purpose of insurance is reducing risk, but private markets often do not do a very good job of risk mitigation. An important risk facing those saving for retirement is what the cost of living will be in twenty or thirty years, yet no private insurance policies address this risk. In many ways, changes in retirement programs in the past thirty years have exacerbated these problems. It used to be that most private retirement programs were defined benefits: the individual received a certain amount, typically adjusted for inflation, depending on his or her salary and years of service. In the past quarter century, however, there has been a move toward defined contribution programs, in which the employer provides a fixed amount (sometimes matching employee contributions) to a retirement account. Individuals can decide on how to invest that money—in stocks or bonds—but then the income that the retiree gets depends entirely on how well the investments do.

Those who put their money into stocks saw their retirement accounts devastated in the stock market crash of 2008. Those who put their money in U.S. Treasury bills saw their income dwindle after 2008, as the Fed drove interest rates down to zero. Retirement programs that were meant to provide security in old age thus did nothing of the kind. For many, the only security was that provided by their Social Security payments.

LACK OF INDEXING: THE INABILITY OF PRIVATE MARKETS TO INSURE SOCIAL RISKS

A major difference between private insurance policies and the Social Security program is that Social security benefits are **indexed**—they increase with inflation. The closest private policies have come to indexed benefits are annuities whose benefits are linked to the performance of the stock market. When these insurance policies were introduced, it was thought that they would provide a hedge against inflation; the stock market would go up with prices in general. However, there have been long periods of time when stock prices have not kept up with inflation, and, in general, it is now recognized that stocks provide relatively poor protection against inflation.

The risks of inflation are an example of an important class of risks referred to as **social risks**. These are risks faced by society as a whole. It is difficult for any private insurance firm to bear such risks. Aside from exceptional circumstances, such as war, the deaths of different

individuals are "independent" events. A firm that insures a large number of individuals can predict fairly accurately the number of individuals who will die each year. However, if there is a war, the number may be much larger. Thus, most insurance policies exclude the coverage of death in a war. Similarly, if a firm insured against inflation, it would find that if the inflation rate increased much faster than it had expected, it would bear a loss on all its policies; it might well find that it was not able to meet all these commitments at the same time. This is perhaps part of the reason why there is no market for insurance against inflation; in fact, there is no way that individuals can fully hedge against inflation by buying any private security or mix of securities.

Although the market cannot, or at least has not, provided insurance against inflation, the government can provide insurance against inflation. It can sell—and beginning in January 1997, actually has sold—bonds, the returns for which are guaranteed in real terms. Thus, one of the great advantages of Social Security is that it provides retirees with some protection against inflation.

The government has obvious advantages over private firms in providing insurance for social risks: it is in a position to meet its obligations by raising taxes (of course, if it imposed taxes on the same people who were suppose to receive benefits, it wouldn't be of much help), and it can engage in risk sharing across generations. The costs of a war, for instance, can be shared by the current generation and future generations: through reducing investment during the period of war and through subsequently imposing taxes on the young for the benefit of the old, the costs of the war can effectively be shared by the generation that is working during the period of war and by subsequent generations. If the economy experiences a particularly bad episode of inflation in a given decade, it can transfer some of the burden of that onto younger, working generations.

ADVERSE SELECTION, DIFFERENTIAL RISKS, AND THE COST OF INSURANCE

A third major problem with private insurance arises from the fact that different individuals have different life expectancies. Consider life insurance, which provides a fixed payment to the insured person's survivors after his or her death. (Thus, life insurance is actually death or survivors' insurance.) A firm selling a life insurance policy does not want to insure people who are likely to die; if it knows that they are likely to die, it will insist on charging a high premium. For someone over age 65 or someone

with a heart condition, the premium may be particularly high. On the other hand, for private insurance firms selling an annuity, the concern is just the opposite: they want to insure only people who are unhealthy, who are likely to die soon. Because women live longer than men, insurance companies in the past charged women lower life insurance premiums but higher premiums for annuities.

To the extent that differences in individuals' life expectancies can easily be identified, economic efficiency requires that private insurance firms charge premiums reflecting those differences. Some believe that this is unfair: someone who has bad health is unlucky enough, to charge that individual a higher premium for life insurance adds insult to injury.

In practice, it is costly to identify good risks, and insurance firms spend considerable resources attempting to do so. Indeed, an insurance company's profits can be increased as much through better ways of identifying good risks from bad risks as they can be by improving the efficiency and quality of the overall service provided. If private insurance firms cannot discriminate among individuals by levels of risk, quite another problem arises. In competitive equilibrium, the premiums must reflect the *average* risk of those who purchase the policy (for life insurance or an annuity, this corresponds to the average life expectancy). However, this means that good risks are, in effect, subsidizing poor risks. With annuities, those who die young subsidize those who live a long time; with life insurance, those who live a long time subsidize those who die young. This means, in turn, that good risks, on average, get back less from the insurance company than they put in. To them, insurance is a bad gamble. However, if some individuals know that they are low risk, but cannot demonstrate this to the insurance company, they will not buy insurance if they are not very risk averse. When the best risks no longer purchase insurance, the premiums must increase. This process, by which only the worst risks purchase private insurance, is called *adverse selection*.[9] Adverse selection may provide part of the explanation for high premiums charged for annuities. The government, however, can force all individuals to purchase the insurance, and thus avoid the problem of adverse selection. In doing so, it is engaging in some redistribution: good risks are paying more, and bad risks are paying less than they "should." But all (or almost all) can benefit from the reduction in risk that results from social insurance.

---

[9] For a more extended analysis of the effects of adverse selection in insurance markets, see M. Rothschild and J. E. Stiglitz, "Equilibrium in Competitive Insurance Markets," *Quarterly Journal of Economics* 90 (1976): 629–650.

MORAL HAZARD AND SOCIAL SECURITY

There is another reason why private insurance firms often offer only limited insurance. Insurance may reduce the individual's incentive to avoid the insured-for event; we referred to this as *moral hazard* in Chapter 13.

Individuals, in contemplating making provisions for their eventual retirement, face two important sources of risk. The first is that they do not know how long they might live after retirement. Individuals who did not buy annuities would have to husband their resources carefully; they would have to worry about the possibility that they will live longer than average. In insuring this risk, no significant moral hazard problem arises. However, a moral hazard problem does arise in the second risk, for which Social Security provides insurance: individuals do not know how well they will be able to work at age 62 or 65 or 70. Some individuals are healthy and their skills have not become obsolete, so they continue to work well beyond age 70. Others are incapable of working due to a disability. However, many individuals are in an in-between state at age 62 or 65—they are not medically disabled, but they are finding it increasingly difficult, less enjoyable, or less productive to work. When they are younger, individuals like to purchase insurance against the possibility that they will fall into this gray area—not so disabled as to qualify for a medical disability, but not so well that they can easily continue working. Social Security provides that insurance: it enables an individual who wishes to retire at 62 to do so. However, the better the "insurance"—the larger the fraction of working income that Social Security replaces—the weaker the incentive to work. With full replacement, even an individual who is in perfect health and highly productive might be induced to retire. This is the central moral hazard problem associated with Social Security.

The failure of the private market to provide complete insurance should not be viewed as a capricious consequence of rapacious insurance companies trying to exploit hapless consumers, but rather as a rational response to a critical economic problem, of providing at least some incentives to the insured. To the extent that this explains the limitations of insurance provided by the private market, there is no reason to believe that the government can do any better: the trade-offs between risk reduction and incentives remain the same. In other words, concerns about the moral hazard problem provide a limitation on the extent of insurance that can or should be provided, privately or publicly.[10]

---

[10] This is a slight simplification. The trade-offs as viewed by a private insurance company may look markedly different from those as viewed by the government. For instance, the government should be aware that early retirement may result in a loss of tax revenues, a consideration that the private insurer would not care about.

RETIREMENT INSURANCE AS A MERIT GOOD

Even when there are good insurance markets, there remains a rationale for government action: if society believes that it cannot countenance an older individual suffering because he or she has failed to make adequate provision for the retirement years, and if a number of individuals fail to make adequate provision for their retirement on their own, there is an argument for *compelling* individuals to do so. Those who do make provision for their retirement may feel that it is unfair for them to bear the burden of those who could have made adequate provision for their retirement but lacked the foresight to do so. In this view, retirement insurance (and life insurance) are merit goods that a paternalistic government forces on the individual for his or her own good. However, they are different from many other merit goods in that much of the cost of the individual's failure to purchase the good is borne by others. To the extent that this provides the rationale for social insurance, however, it suggests only that the government should require individuals to obtain insurance, not that the government should require individuals to purchase the insurance from the government itself. Furthermore, it provides a rationale only for a certain "base level" of Social Security, just high enough that government feels no necessity to raise their consumption further, should they do no additional saving. It is the high transactions costs, the problems of adverse selection, and the failure of private insurance to insure against inflation and other social risks that provide the rationale for public provision.

> ## MARKET FAILURES
>
> - High transactions costs
> - Incomplete insurance
>   - Inability of private markets to insure social risks: lack of indexing
>   - Adverse selection: possible explanation of imperfections in annuity markets
>   - Moral hazard
> - Retirement insurance as a merit good: moral hazard problem created in supporting those who fail to provide for themselves

SOCIAL SECURITY, FORCED SAVINGS, AND INDIVIDUAL CHOICE

One of the reasons that Social Security has such low transactions costs and can avoid the adverse selection problem is that everyone participates. Two concerns are raised: (1) Does Social Security "force" individuals to oversave? (2) Does the conservative investment of Social Security limit the extent to which individuals can voluntarily expose themselves to risk? The trade-offs between reducing costs and increasing the scope for individual choice are clear: it is administratively less expensive to provide a uniform retirement program for all individuals than to have a large number of competing

programs available, among which the individuals can choose. However, as long as the level of retirement benefits is relatively low, few individuals are being forced to save more for retirement than they would like; hence, there is no significant welfare loss from the provision of a reasonably low level of benefits. Moreover, if individuals want to undertake more risk, they can do so in their own retirement programs, going beyond Social Security.

IS THERE A NEED TO REFORM SOCIAL SECURITY?

The previous discussion has clarified the many different roles that Social Security plays. It is a forced savings plan, ensuring that individuals put aside sufficient amounts so that they do not become public charges in their old age; it is a "community-rated" inflation-indexed annuity plan, in which all individuals who make the same contributions, regardless of their health conditions and life expectancy, get the same benefits, with the level of those benefits insured against inflation; and it is a redistribution program.[11]

To its advocates, this mixture is one of its strengths: it may have done more to reduce poverty among the elderly than might have been politically acceptable in an *explicit* redistribution program. Today, poverty among the elderly is lower than it has been since data began to be collected in the late 1950s.

To its critics, this mixture is one of its weaknesses: it leads to a lack of *transparency,* which should be the hallmark of democratic government. Just as there has been increasing demands for "truth in advertising"— that consumers should have the information that allows them to know what they are purchasing—so too in the public sector. Taxpayers should know, for instance, to what extent their tax dollars or Social Security contributions are being used to redistribute income from one group to another. Lack of transparency may contribute to the difficulties in reforming Social Security, as myths and reality get confused. For instance, the aged resist cuts in benefits, believing that they have already paid for those benefits through their Social Security contributions, even if the value of the benefits significantly exceeds the value of their contributions.

Reform of Social Security is so difficult both because it performs this multiplicity of roles and because it addresses a multiplicity of market failures. For instance, *if* Social Security were solely a "forced savings plan,"

---

[11] It used to provide insurance against "early obsolescence": individuals who retired earlier—either because of health or because their job skills were no longer highly valued in the labor market—were, in effect, compensated by receiving a higher present discounted value of benefits (though lower annual benefits). In the future, this will no longer be the case, as early retirees will have monthly checks reduced accordingly.
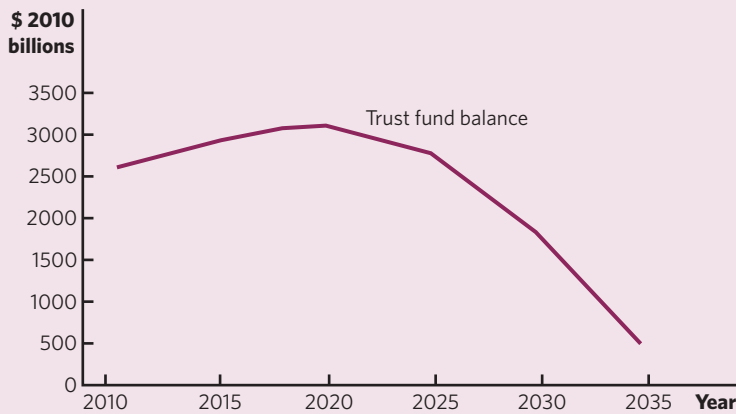
it is not obvious that individuals should be required to save through the government; a private savings mandate might do as well. Before turning to these reform issues, however, we first discuss the complaints against Social Security: why reform is needed.

From many perspectives, Social Security has been a success: whereas in the past, poverty among the elderly was viewed as a real social problem, today, not only are poverty rates among the elderly far lower, but they are also lower than those in the population as a whole. The beneficiaries are not only the recipients but also their children, who otherwise would have helped to support them or felt guilty about not doing so. In the past, the elderly frequently had no choice but to move in with their children; now they have a choice, and increasingly, they exercise that choice by maintaining independent households.[12]

There are five major complaints against Social Security. First, it is contributing to the country's long-run fiscal crisis. Beginning around 2021, there will be a shortfall of revenues relative to expenditures, and the trust fund that has been set aside to meet the needs of the aging population will be drawn upon; but this too will be exhausted by 2037.[13] (See Figure 16.3.) As a result, confidence in the Social Security system has eroded.

---

[12] Indeed, as a result, the household poverty data probably understate the improvement in the economic status of the elderly. When they have their own households, the household income may fall below the poverty threshold, whereas when they are in their children's households, they may be above the poverty threshold. To the extent that they *choose* to keep their own households, they are nonetheless better off, in spite of the fact that the poverty statistics would show an increase in the number of elderly living in households below the poverty line.

[13] The precise estimates vary from year to year. These numbers are from *The 2010 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds*, August 9, 2010, Table IV.F7 (intermediate projections).

For instance, a poll of 18- to 34-year-olds showed that more believe in UFOs (46 percent) than that Social Security will exist by the time they retire (28 percent).[14]

Second, critics allege that Social Security discourages savings, and thus slows down the growth of the economy. Third, critics argue that it discourages work effort. Fourth, they complain that it gives a low rate of return—it is a bad investment. Fifth, there are a number of inequities in its design.

It is worth noting that there are *not* major complaints about either the efficiency of the Social Security Administration or its responsiveness to its "clients." Miraculously, this appears to be the one major government department that has managed a successful computerization. Its overall administrative costs are low compared with either public or private agencies engaged in similar activities. Furthermore, a consumer satisfaction survey focusing on telephone service ranked the Social Security Administration among the top four "companies."

THE NATURE OF THE FISCAL CRISIS

The essential fiscal problem facing the Social Security system, in its modified pay-as-you-go basis, is that the system's financial viability depends on the ratio of the number of workers to the number of retirees. As Figure 16.4 shows, that number has decreased markedly, from more than 16 to 1 in 1950, to slightly less than 3 to 1 today—and it is anticipated to decrease further to 2 to 1 in the next twenty-five years. There are three reasons for the change: earlier retirement ages,[15] increased longevity,[16] and slower population growth.[17] The first two reasons translate into a larger population of retired people, and a slower population growth rate means that the ratio of the young to the old is lower than it used to be. The increased labor force participation of women during the past three decades has somewhat

---

[14] P. G. Peterson, *Will America Grow Up before It Grows Old?* (New York: Random House, 1996).

[15] In 1940, the average age at which retirees began receiving Social Security was 68.8 for men and 68.1 for women. By 1991, those numbers had declined to 63.7 for both men and women. In 1950, 46 percent of males age 65 and over worked; in 1991, only 16 percent did. C. E. Steuerle and J. M. Bakija, *Retooling Social Security for the 21st Century* (Washington, DC: Urban Institute Press, 1994), p. 43. For further information about retirement ages, see House Committee on Ways and Means, *2011 Green Book*.

[16] In 1940, the average remaining life expectancy for those surviving to age 65 was 12.7 years for men and 14.7 years for women. By 1990, these figures had increased to 15.3 for men and 19.6 for women. By 2050, they are expected to increase still further to 18.0 for men and 22.4 for women. Steuerle and Bakija (note 15), p. 41.

[17] Total fertility rates (the average number of children born to a woman over her lifetime), after reaching a peak of 3.7 in 1957, have fallen to about 2. At these low fertility rates, population will (apart from immigration) start to decline. See U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, "Births: Final Data for 2009," *National Vital Statistics Reports* 60, no. 1 (November 3, 2011): Table 4.
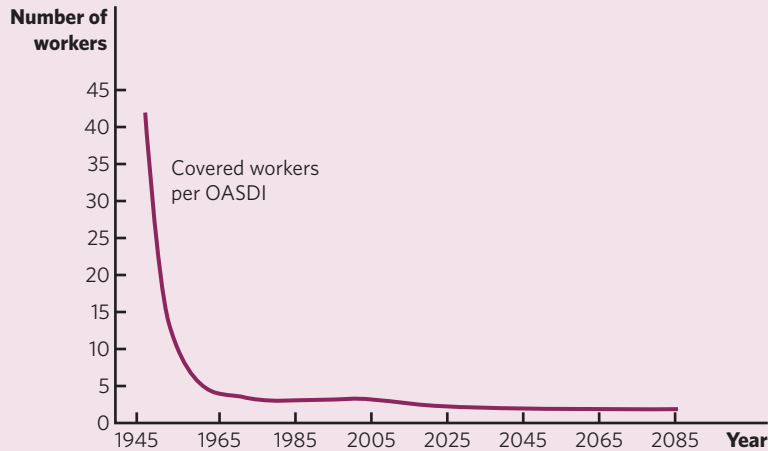
offset these demographic trends, but in the decades ahead the underlying trends will dominate.

The second factor determining the viability of the system is the rate of increase of productivity. What matters is the income of those working relative to the benefits of the retirees. If incomes are rising rapidly, due to increased productivity, then there is more income available to support retirees. However, the rate of productivity increase has declined markedly—by more than half—since 1973, and even more so has the associated rate of increase in incomes, so wage incomes have not kept pace with productivity,[18] exacerbating the fiscal problems.[19] Any upturn in is likely to be too small to restore financial viability.

Social Security is only one of several programs directed at the aged; others (Medicare, and the long-term-care part of Medicaid) were discussed in Chapter 13. They share in the problems we have just discussed: providing these programs for a growing aged population will impose increasing burdens on the working population. In the United States, the problem will be crystallized with the retirement of the baby boomers, which began sometime after the turn of the twenty-first century. Most of the advanced

[18] For the decades prior to 1973, the annual rate of increase of productivity averaged 2.9 percent; from 1973 through 1993, it averaged 1.1 percent. Since then, different statistics show either a slight increase or a slight decrease. See the *Economic Report of the President, 2012*.
[19] The problem has been made even worse by the increased inequality of incomes. Thus, even though average productivity and wages have increased, *median* real wages—that are much more relevant to the Social Security program—have not kept pace. This problem would be reduced if the ceiling on Social Security contribution were substantially increased.
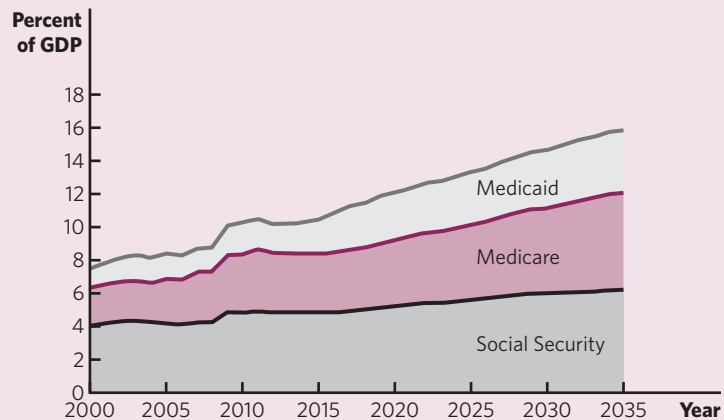
industrialized countries face problems that are even more severe; and many of them—including Japan—will face the problem earlier.

Taken as a whole before 2012, these entitlement programs appeared to present an immense budgetary problem. Expenditures on them, as a percentage of GDP, were projected to more than double between 2000 and 2035, from 7.5 to 15.9 percent (see Figure 16.5). Because, historically, the federal share of taxes in GDP has averaged around 19 to 20 percent, taxes would have to be increased substantially and/or other government programs would have to be cut back drastically to meet these expenditures.

As we noted in Chapter 13, the major contributor to these soaring expenditures were those for health care—Medicare and Medicaid. The costs were as much the fault of the high-cost U.S. health care system as of anything else: if the United States had a health care system that was as efficient and cost effective as several in Europe, there would have been no serious financial problem facing these programs. Coincidental with the passage of the Affordable Care Act, the cost curve for medical care in the United States began to bend down—the deficits that seemed so large suddenly looked manageable.

Social Security itself contributes only a small part of the problem: its expenditures are projected to grow from 4.1 percent of GDP to 6.2 percent. The Social Security problem is both smaller and easier to address than Medicare and Medicaid. It is *just* a financial problem: either receipts (contributions) must be increased or benefits reduced. The shortfall could be corrected permanently by a tax increase of around 3 percentage points, or a

benefit decrease, such as a modest increase in the retirement age, of a corresponding magnitude.[20]

However, even the magnitude of the shortfall is a subject of some controversy and depends on variables that are hard to project. The central forecasts on which projections of deficits are based are predicated on immigration slowing, relative to the labor force. Should the pace of immigration pick up, that, by itself, might have a major impact on Social Security's financial position.

> ## SOURCES OF FISCAL IMBALANCE
>
> - Decrease in population growth
> - Decrease in productivity growth
> - Decrease in retirement age
> - Increase in longevity

There is particular concern that these programs directed at the aged are crowding out expenditures on children—investments in the country's future. As it is, the federal government now spends more than eleven times as much on each elderly person as it spends on each child,[21] in spite of the fact that children are far more likely to be in poverty than are the elderly.

SAVINGS

The provision of Social Security reduces individuals' need to save for retirement. There is concern that this leads to lower aggregate savings, and thus lower investment and growth in productivity. The magnitude of the effects on savings and its implications for growth and policy have been a subject of debate.

Whereas Harvard professor and former chair of the Council of Economic Advisers, Martin Feldstein claims that Social Security may have led to a reduction in private savings in the United States of as much as 60 percent,[22] others argue that the effects are much smaller and that the system may have actually encouraged savings.[23] It allowed

---

[20] There is some uncertainty about these numbers. The Social Security trustees estimate alternative scenarios, involving different rates of growth of production, population, and so on. In the pessimistic scenario, the gap is twice as big. In the optimistic scenario, there is even a surplus under current arrangements.

[21] The statistic, though widely quoted, may be somewhat misleading, as it includes federal Social Security payments, which can be thought of as benefits to which the individual has contributed, in the same way that he or she would have by purchasing a retirement policy; only the *transfer* component should actually be included.

[22] The original source for these arguments is Martin Feldstein, "Social Security, Induced Retirement, and Aggregate Capital Accumulation," *Journal of Political Economy* 82, no. 5 (1974): 905–926. Subsequently, Feldstein attempted to answer his critics and show that his results were robust, see "Social Security and Saving: New Time Series Evidence," *National Tax Journal* 49, no. 2 (June 1996): 151–164.

[23] There is considerable controversy concerning Feldstein's results, involving a number of technical issues. S. Danziger, R. Haveman, and R. Plotnick, in "How Income Transfers Affect Work, Savings, and Income Distribution," *Journal of Economic Literature* (September 1981), conclude in their survey that the transfer programs, in which the effect of Social Security is dominant, "have depressed annual private savings by 0 to 20 percent relative to their value without these programs, with the most likely estimate lying near the lower end of this range." Robert Barro has taken a more extreme view and argued on theoretical grounds that the deficits in Social Security program should have no effect, as parents simply adjust the bequests that they leave to their children in response (e.g., providing larger bequests in response to the larger deficits). He has also attempted to provide empirical support for this proposition, although most economists remain skeptical.

individuals to retire early, and as they thought about earlier retirement, they realized that they would need more than the amount provided by Social Security.

This perspective is consistent with the growth of private pension plans, which occurred after Social Security began. In 2010, although 46 percent of aggregate income for those aged 65 or older came from Social Security or other public pensions, more than one-quarter of this elderly population supplemented their retirement income with income from private pensions or annuities, totaling 9 percent of aggregate income for this age cohort.[24]

In an open economy, in which the country can borrow from abroad, reduced savings need not be directly translated into reduced investment and thereby reduced productivity.[25] But still, it appears as if a $100 savings reduction leads to an investment reduction of around $40 to $60. Thus, negative impacts on savings remain a source of concern.

The role of Social Security in contributing to the low household savings rate of Americans remains, however, a subject of controversy. The dramatic fall in savings rates—to near zero in the years before the Great Recession—had little if anything to do with changes in Social Security.

Recent changes in taxation—subjecting retirees with higher income to taxation on 85 percent of Social Security benefits—have the effect of reducing the net returns to savings, and thus the level of savings, for some individuals. However, for lower-income individuals, the tax rate is sufficiently small, and they could take advantage of tax-free or tax-deferred retirement accounts. Moreover, there is little evidence of a high interest elasticity of savings—that is, as savings do not respond much to changes in the interest rate in some cases, lowering the net return to savings may induce more savings.

One reform that would probably increase national savings would be fully funding Social Security, so that each generation puts aside funds for its own benefit. A more controversial reform, which some argue would increase savings, is privatization of Social Security. This topic is discussed later in the chapter.

LABOR SUPPLY

Some have argued that Social Security also has an effect on labor supply: it induces individuals to retire earlier than they otherwise would. The large decline during recent decades in labor force participation

---

[24] Social Security Administration, *Income of the Population 55 or Older, 2010*, March 2012, Tables 2.A1 and 10.1.
[25] Even then, however, it leads to increased indebtedness to foreigners and thereby lower future standards of living.

of those over 65, noted earlier, has occurred at exactly the same time—between 1968 and 1976—that there has been a large increase in Social Security benefits per recipient—real benefits were increased by 39 percent. Michael Hurd of the State University of New York at Stony Brook and Michael Boskin of Stanford University argue that the decline in labor force participation between 1968 to 1996 was, indeed, due largely to the real increase in Social Security benefits.

Government programs have both income effects and substitution effects. Inefficiencies are associated with substitution effects. Both these effects arise in the case of Social Security. The large transfer of resources to the elderly has an income effect: the elderly take some of this increased income in the form of extra leisure—early retirement. In addition, there is a substitution effect, because Social Security changes the return from working.

There is, however, considerable controversy about the size, and even direction, of the substitution effect. As individuals work longer, their total contributions to Social Security increase, as they are taxed on their additional income, but their benefits per year are also increased. The question is: Do the benefits increase *enough* to compensate for the increased payments? The adjustment process is gradual, but as a result of the reform act of 1983, by the year 2008, on average, there was full adjustment. Furthermore, in recent years, there has been an effect encouraging workers to stay on: the magnitude of Social Security benefits depends on individuals' contributions in previous years. In 1937, their contributions were limited to the first $3000 of their income; in 2010, they were based on earnings up to $106,800. Thus, the way Social Security benefits are calculated means that an individual with, say, a $40,000 income may now experience a significant increase in benefits by staying at work for an additional year or two.

Eventually, however, for most individuals, the Social Security system will not act as much of a tax. Their increased contributions "buy" commensurately higher retirement income. The system acts as a tax only to the extent that individuals are forced to save more than they otherwise might, or that returns are lower than they might otherwise have obtained. Because most individuals do save beyond Social Security, it is not *forcing* them to save more than they would like. For poor individuals, the benefit formula generates returns to contributions that are beyond those that could be obtained elsewhere, and hence there is a positive substitution effect, whereas for rich individuals the opposite is true.

EFFECTS OF SOCIAL SECURITY

Social Security may have an adverse effect on savings:

- Reduces need for private savings for retirement.
- Reduced savings may result in lower investment and lower growth in productivity.

Social Security may have an adverse effect on labor supply:

- The effect in general is probably small.
- The effect may be larger among older workers.

Note that there is a distortion of the labor–leisure choice whether individuals are subsidized or taxed. The fact that some individuals are subsidized and some are taxed does not result in the distortions' netting out: the total distortion is not simply related to the average value of the marginal subsidy or tax.

Although for most individuals the substitution effects are probably small, there is one group—the elderly—for whom it is probably large, for two reasons. First, for those between 65 and 69 on Social Security, for every three dollars they earn above a certain amount, they lose one dollar of benefits—in effect, a 33⅓ percent tax on earnings—besides having to pay income taxes on their additional earnings. As a result, a 66-year-old with a high income from investments could face an effective marginal tax rate of more than 60 percent on wage income—and this does not include state and local income taxes.[26] Moreover, Social Security benefits are calculated on the basis of the individual's highest-earning thirty-five years. As a result, for many elderly individuals, especially those working part-time, benefits are completely unaffected by their additional earnings; for them, Social Security contributions are simply taxes.

THE RATE OF RETURN

The fourth complaint is that Social Security does not provide a rate of return comparable to that obtained in private retirement accounts. The complaint was not heard in the earliest years of Social Security, because then benefits far exceeded contributions, as we have noted. However, it has been heard since the 1983 reforms, as a semblance of fiscal responsibility was striven for.

The complaint is, in one sense, not quite accurate. The return should be contrasted with a security with comparable risk properties—a U.S. government bond that is indexed against inflation. Although these bonds have not been around long enough to be sure what they will yield in the long run, especially if the United States encounters a period of high and variable inflation, it appears as if the returns are comparable (perhaps slightly higher) than would be obtained if the funds had been invested in such securities. The returns are substantially lower than those obtained on the stock market when it is having a boom, as it did beginning in 1995, but are substantially higher than those obtained on the stock market when it is going through a slump—as it has periodically,

---

[26] Assume this 66-year-old earns $100 more and is in the 40 percent marginal tax bracket; after tax, he or she gets to keep only $60 and also loses $33 in benefits. With 85 percent of Social Security benefits taxable, the "net" cost of this loss is only $22. Netting this from the $60, however, leaves this individual with only $38; in effect, he or she faces a 62 percent marginal tax rate.

as with the stock market crashes of 1929, 1987, and 2008. Investments in equity thus have greater returns, but also greater risks. Social Security is about providing *security* for retirement, and thus it is appropriate that the money be managed safely.[27]

However, the question is really misplaced. The United States has a pay-as-you-go system, in which the long-run return (in equilibrium) is related to the rate of growth of the economy, which typically will be lower than the rate of return on capital, although the rate of growth

> ## REASONS FOR REFORM
>
> - Fiscal imbalance
> - Adverse effect on savings
> - Adverse effect on labor supply
> - Low (perceived) rate of return
> - Inequities—different actuarial benefits among those making similar contributions; redistribution not based on overall assessment of need

of the economy has, in fact, been consistently higher than the rate of return on government securities.[28] The low rate of return is not the result of a failure to invest well or of high transactions costs—indeed, transactions costs are remarkably low—but simply the result of having a pay-as-you-go system. Returns could be increased, but only by switching out of the system. Doing so, as we shall see later, would entail a sacrifice in consumption on the part of the generations involved in the switch. Even then, though, if the funds are invested safely, the returns are likely to remain low.

INEQUITIES

It is in the nature of any insurance program that some individuals get back more than they contribute in premiums and some less. For example, buyers of fire insurance who do not have a fire "lose," and those whose houses burn down "gain." In Social Security, those who live long get back far more than they contribute, whereas those who die before retirement get back less. This, in itself, is not inequitable.

Within Social Security, however, those who make similar contributions get back, *on average,* different amounts. That is, their *expected* paybacks differ because of differences in family status. To some, these

---

[27] There is actually some controversy over this issue: a widely diversified portfolio of stock held for more than twenty-five years invested in *any* year since 1930 would have done better than a comparable investment in government securities. However, there is no guarantee that this will be true in the future. Indeed, as investors have become aware of these differential returns, they have bid up the price of equities, reducing their returns.

[28] This can be seen in a stylized overlapping generations model, in which each generation lives for two periods, working in the first. In the absence of growth, a tax on the young used to pay retirement benefits for the old generates a "return" of zero: each generation gets back exactly the amount that it has contributed. If population is growing at the rate n and wages at the rate p, then for each dollar of contributions made when they were young, workers receive back $(1 + n)(1 + p)$ when they are old.

differences are viewed as inequitable: Why should someone simply get more because he or she *chooses* to get married?[29]

Of course, because there is a redistributive aspect of Social Security, for some people benefits relative to contributions may be large. To most observers, this is not inequitable, as long as those receiving the redistribution are, in some sense, "deserving." However, in determining the magnitude of the redistribution, Social Security looks only at the individual's Social Security contributions, not his or her wealth or nonwage income.

REFORMING SOCIAL SECURITY

There are two sets of proposals for reforming Social Security. A set of modest proposals entails reducing expenditures and increasing revenues to bring the program into fiscal balance, with only modest changes in its structure. A set of more drastic reforms entails significant structural changes, among them, privatizing Social Security.

REDUCING EXPENDITURES

A number of proposals have been put forward for reducing expenditures without undermining the overall effectiveness of the program.

ADJUSTMENTS OF BENEFIT FORMULAE   The formula for determining benefits could be adjusted in ways that would slightly reduce expenditures, and might even enhance efficiency and equity. For instance, as was noted earlier, currently, benefits are based on the highest thirty-five years of contributions. Individuals who work longer, but whose income in those later working years is not among their best, receive no increase in benefits in return for their additional contributions. The formula could be adjusted in ways that do not affect benefits of most workers who have been full-time participants in the work force all of their lives, but that at the same time provide greater incentives for work for those who are currently being encouraged to retire earlier because of the peculiar way that benefits are calculated.

---

[29] Because married men tend to live longer, even if there were no spousal benefits, the actuarial benefits of married men would exceed those of single men making comparable contributions.

There are other examples of possible "inequities": for comparable contributions, women receive back more than men, because women live longer than men; for comparable contributions, nonsmokers receive back more, because they live longer than smokers. Private insurance firms would adjust payments to reflect differences in life expectancies (when they are allowed to do so). Similar inequalities are in the Medicare program.

ADJUSTING THE NORMAL RETIREMENT AGE The 1983 reforms increased the normal retirement age at a very gradual pace, to age 67 by 2022. The argument for increasing the retirement age is that better overall health has led to increased longevity, contributing to the system's financial problem. On average, individuals should be able to remain in the labor force longer. (There is a certain irony that, at the same time, the actual retirement age has been coming down.) However, there is a wide range of health conditions: a substantial number of elderly have sufficiently poor health that they must retire at 65, or even at 62. Increasing the "normal" retirement age still further—say, to 68—will reduce the benefits for these individuals, by 10 percent or more. This could impose severe hardship on many of these people, who often have few assets or other sources of income. Advocates of increasing the retirement age give two answers: If such individuals truly have to retire, then they should be covered by disability; and if reducing benefits imposes real hardship, they should become eligible for supplementary income benefits. Changing the retirement age may necessitate concomitant adjustments in the disability and supplementary income programs, but this is no reason not to make an adjustment in retirement age in line with changes in current conditions.

ADJUSTMENT OF THE COST OF LIVING INDEX Once an individual retires, his or her benefits increase with the rate of inflation, as measured by the consumer price index (CPI). A commission appointed by the Senate Finance Committee and headed by Michael Boskin of Stanford University, who had been chair of the Council of Economic Advisers under President Bush, reported in December 1996 that it believed that the CPI overstated inflation by 1.1 percent per year. Even a half-percent correction in the CPI would reduce the long-term Social Security deficit substantially, by as much as a third. Over the subsequent years, a number of adjustments were made to the CPI, so that the magnitude of the problem has been greatly reduced. Nonetheless, some believe that inflation continues to be overstated, and in 2013, President Obama proposed making further adjustments.

Some worry that the CPI, which is based on the cost of living, on average, understates increases in the true cost of living of the elderly, especially because the elderly have higher medical costs. However, the rise in medical costs recently has appeared to be contained, and those who have looked carefully at the quality improvements in medical care believe that medical costs represent one of the major areas of overstatement—adjusting for quality, health care costs may be increasing more slowly than prices in general, and might even be decreasing.

As in the case of changing the retirement age, it may be desirable to make concomitant adjustments in the disability and supplementary

income programs at the same time that a CPI adjustment is made; to the extent that the CPI adjustment accurately reflects changes in cost of living, however, the CPI adjustment is intended merely to ensure that inflation itself does not impose additional hardship on the aged. Inflation also should not be the basis for providing unintended increases in real benefits. If there is a desire to increase real benefits, then there should be an explicit decision to do so.

MEANS-TESTING BENEFITS   Social Security is an unusual program: it redistributes income, but in doing so, looks only at the individual's (Social Security) wage income, not the overall well-being of the recipient. A billionaire who inherited his or her fortune and therefore worked little during life—and so had low wages—would receive back far more than his or her Social Security contributions, an amount identical to that received by an unskilled worker with the same history. Thus, subjecting at least the redistributive component of Social Security to means testing would increase equity and reduce expenditures.

Opposition to means testing, however, has been strong, for it would make explicit the redistribution role of Social Security. Today, many, if not most, individuals think of Social Security more as a government-run retirement program, to which they make contributions (pay premiums) and receive benefits that commensurate with the contributions. Making Social Security means tested would convert it into a welfare program—and Social Security advocates worry that, in the long run, this will undermine support for the program.

INCREASING REVENUES

The major proposal for increasing revenues is to subject Social Security payments to taxation. We noted earlier the anomalous nature of Social Security redistributions. There is a similar anomaly in the tax treatment of Social Security. The general principle of taxation is that taxes should be based on "ability to pay"; individuals with higher income should pay higher taxes. An important exception (discussed more extensively in later chapters) is that much of retirement savings is tax exempt; this is so partly to encourage individuals to save more for their retirement, or at least not to discourage them from saving. Individuals' own Social Security contributions are included in their income, and are subject to tax; however, the contributions that employers make on behalf of their workers are not included in individuals' incomes, and accordingly are not taxed. At the same time, in the past, none of Social Security benefits were taxed;

today, up to 50 percent of benefits are taxable for most people, but up to 85 percent for upper-income individuals. However, there is a strong theoretical rationale for imposing tax as based just on income regardless of source. Two elderly individuals could have exactly the same income—one derived exclusively from Social Security, the other derived largely from other sources—but pay different taxes. From this perspective, subjecting Social Security income to treatment comparable to other forms of retirement savings would increase equity at the same time as it raised revenue.

SUMMARY   These reforms could make a significant impact on Social Security's fiscal situation. By one calculation, simply raising the Social Security retirement age, adjusting the cost-of-living increases, instituting means testing, and making more of Social Security benefits taxable could restore fiscal balance, at least through the year 2030.

STRUCTURAL REFORMS

Several minor structural reforms that would increase equity, increase efficiency, and improve Social Security's financial position have already been noted: changes in the benefit formula to make benefits increase with contributions; means testing, at least for the redistributive component of Social Security; and subjecting Social Security at least to taxation comparable to other forms of retirement savings.

The dependence of Social Security benefits (relative to contributions) on family status has also been strongly criticized. Currently, nonworking spouses receive substantial benefits, so the actuarial benefits of a married worker whose spouse does not work outside the home are substantially higher than those of a single worker who has made the same contributions. Advocates of this preferential treatment sometimes argue that spouses who remain at home to raise children work too, and therefore should receive retirement benefits. However, that is not the issue; the issue is the magnitude of the benefits relative to Social Security contributions. Spousal benefits could be provided, but the monthly benefits would be adjusted to ensure that those who make comparable contributions receive comparable *actuarial* benefits. These inequities might have been overlooked in the past, when there was one predominant family structure, with a man working and his wife staying at home. Given today's varying patterns of family structures and work, there is increasing concern about these inequities.

Two major proposed changes go beyond these structural reforms: one entails Social Security investing its trust funds in equities, and the other entails privatization of Social Security.

SOCIAL SECURITY ABROAD

Many countries have faced problems with their Social Security systems. For instance, in many European countries, expenditures on Social Security not only are much larger than in the United States, but have grown more rapidly (see figure below). As a result, payroll taxes (earmarked, as in the United States, for Social Security) are high: 31 percent in the Netherlands and 33 percent in Italy. There has been increasing concern within these countries about the incentive effects of such high tax rates (particularly when combined with high income tax rates), although to the extent that benefits increase commensurately with contributions, incentive effects are likely to be very limited.

In some countries, poorly designed Social Security systems (at least before subsequent reforms) gave rise to far larger distortions, as well as budgetary problems. In Uruguay, benefits were especially related to the last few years of an individual's working life; not surprisingly, large incomes were reported during those years.

A few countries have privatized their Social Security systems. For example, Chile introduced a system of privately managed individual accounts in 1981, and since 1990, ten other Latin American countries have adopted the Chilean model. In its earliest days, Chile's privatized system seemed a success, for it coincided with a period of high stock
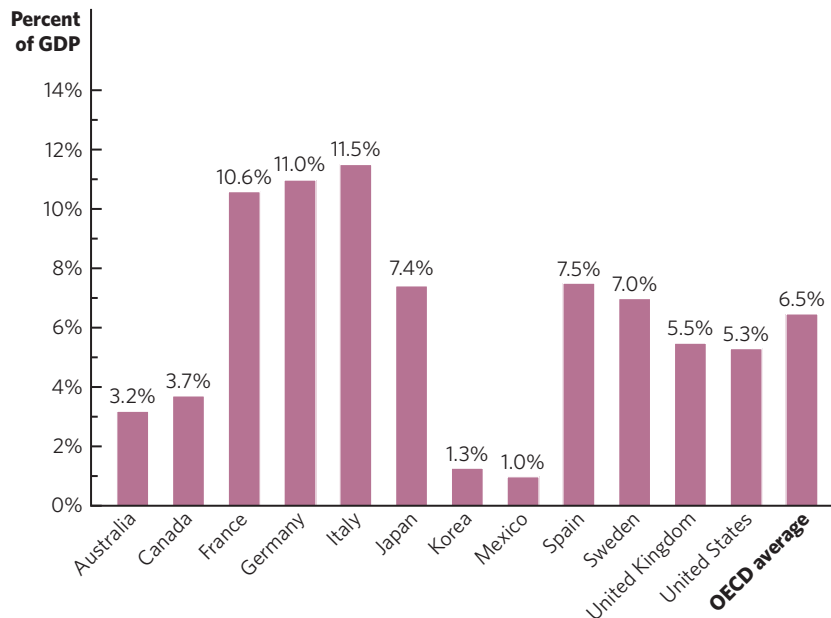
OECD PUBLIC EXPENDITURES ON PENSIONS IN 2005   Public spending on pensions in OECD countries varies widely, with the United States slightly below the OECD average of 6.5 percent of GDP.

SOURCE: "Public and Private Expenditure on Pension," *OECD Factbook 2010: Economic, Environmental and Social Statistics.*

returns. When the returns fell, though, enthusiasm diminished considerably, as investors began to comprehend more fully the high transactions costs and the high risk. There were also concerns over the large number of people not covered by the new system. Over the years, Chile made several system modifications to address these issues, culminating in a comprehensive reform package in 2008 that introduced a noncontributory pension (*Pensión Básica Solidaria*), initially for pensioners in the poorest 40 percent of the population and rising to 60 percent when fully phased in; rectified gender inequities; encouraged greater competition among pension fund management companies (AFPs); facilitated improved AFP management of financial risk; and strengthened financial literacy programs.

Even were Chilean privatization viewed as a success, however, there are some reasons why Chile's experience cannot be extrapolated to the U.S. economy. Chile, like the United States, had a pay-as-you-go system; privatization requires funding the unfunded liability. For the United States, these unfunded liabilities amount to more than $8 trillion. Chile used the funds from privatization of one of its major industries, its copper mines. The United States has no such windfall available, and it would have to impose taxes to fund the conversion.

Concerns about privatization of Social Security systems heightened during the financial crisis afflicting less-developed countries (LDCs) in 1997 and 1998, which saw stock prices in these countries collapse 30 percent, 50 percent, or more. Had workers invested a large part of their retirement savings in the stock market, they could have been left destitute. Government bailouts might follow, so that at least part of the improvement in the fiscal position as a result of privatization would prove to be a mirage. Advocates of privatization say there is an easy answer: individuals in LDCs should invest in stock markets in the more developed countries. However, we have seen periodically, for example, during the 2008–2009 Great Recession, that stock markets in high-income countries can also collapse. Moreover, unless there are offsetting inflows from other countries, one of the other major advantages of privatization—an increase of funds available for productive investments—will also prove to be a mirage. And in any case, most LDCs may find it difficult to justify a government program that, in effect, provides capital from the poor countries to help the rich.

SOURCES: N. Barr, P. Diamond, and E. Engel, "Reforming Pensions: Lessons from Economic Theory and Some Policy Directions, with Comment," *Economía* 11 (Fall 2010): 1–23; B. E. Kritzer, "Note: Chile's Next Generation Pension Reform," *Social Security Bulletin* 68 (2008): 1–22; and S. Borzutzky, "Reforming the Reform: Attempting Social Solidarity and Equity in Chile's Privatized Social Security System," *Journal of Policy Practice* 11 (2012): 77–91.

INVESTING TRUST FUNDS IN EQUITIES For at least the past seventy-five years, anyone who invested broadly in equities (stocks) and held them for twenty years or more would have done better than if he or she had invested a comparable amount in government bonds. For these long holding periods, stocks have earned substantially higher expected returns—some estimates put the difference of average returns at 4 percent per year or higher—with no additional risk. If a fraction of the Social Security trust fund were invested in a broad base of equities, the financial position of Social Security would, accordingly, be improved. Simulation exercises based on the distribution of returns of equities over the past seventy-five years show that, even if no other reforms were undertaken, the life of the trust fund could be substantially extended. Advocates of this proposal suggest further that the increased demand for equities will increase the price of equities, and hence lead to increased investment.

Critics, however, raise three objections. First, they worry about risk. Another major stock market crash, such as that of 1929 or 2008, is always possible. (This objection can be raised even more forcefully against privatization proposals; at least with the trust fund, there is the possibility of smoothing any losses that occur over an extended period of time.)

Second, critics worry that because today much of the deficit is "financed" by borrowing from Social Security (the Social Security trust fund invests its holdings today solely in government debt), government borrowing costs will increase. Thus, even though Social Security's financial position may be improved, that of the rest of the government will become worse. There are no reliable estimates of the extent to which interest rates would rise, and therefore of the adverse budgetary impacts. Many economists, however, believe that the worldwide demand for U.S. government securities is sufficiently strong, so that the interest rate effects of Social Security investing, say, 30 percent of its trust funds in equity would be modest, and far smaller than the increased revenues that the Social Security Administration would be receiving. Moreover, any adverse macroeconomic effect could be mitigated by the Federal Reserve lowering interest rates.

Third, there is concern that the resulting government ownership of equities would affect the private equities market. Some worry about how the shares would be voted. Even after the trust fund had invested 30 percent of its funds in equities, however, studies suggest that government ownership would be under 5 percent of the market. If these were treated as effectively nonvoting shares, the effect on the functioning of the stock market would be insignificant.

PRIVATIZATION   The boldest set of reforms is privatization. President Bush unsuccessfully advocated partial privatization, but in the aftermath of the 2008 crisis, enthusiasm for privatization has waned, as most Americans realized that had Social Security been privatized, many of the elderly would have found themselves bereft of funds for their old age. It was only Social Security that kept poverty at bay.

It became increasingly clear that Social Security privatization had been pushed largely by those in the financial sector, who saw the massive fees from running retirement accounts as a major new source of income. However, these high transactions costs would have lowered the income of retirees commensurately. The Advisory Council on Social Security, in its December 1996 report, suggested that the transactions costs would be substantial, upward of 100 basis points per year—that is, for every $100,000 investment, $1000 would go to Wall Street brokers and other investment agents—substantially higher than the costs that the Social Security Administration would face even if it invested in equities. One percent extra fees a year, every year, means that retirement income would be lowered by 40 percent or more.

A major motivation for the drive for privatization is that private investments provide a higher return. However, as was noted earlier, the low return on Social Security is not because of high transactions costs or poor investment strategies, it is mainly because of the pay-as-you-go structure. By switching to a fully funded system, later generations might achieve a higher return (though not necessarily a higher risk-adjusted return), but the temporary (seventy-five years!) tax to finance the transition would lower the welfare (returns) of the intervening generations. If there were a willingness to pay higher taxes of this magnitude, then Social Security would face no significant financial problems, and no drastic remedies such as privatization would need to be contemplated.

We have also noted that the trust fund could be invested in equities, generating higher returns, and because of the greater potential for smoothing returns over longer periods of time, the associated risk might be considerably lower than the risks associated with private investments in equities—risks that became so evident in the 2008 crisis.

Some argue that privatization is desirable because it would increase the national savings rate. However, it is essential to realize that *the increased savings arises from switching from pay-as-you-go to a fully funded system; it does not arise from the privatization of the system itself.* If, instead of making a transition to a privatized system, one imposed the same 2 percent tax and used the proceeds to create a larger trust fund (government savings), then national savings in the short run would be increased by a comparable amount.

In the long run, the savings rate under a fully funded Social Security or retirement system would be higher than under a pay-as-you-go system, although the differences are often exaggerated. This is because under a fully funded Social Security or retirement system, the elderly are dissaving, at the same time that the young are saving. Net national savings is only the difference between the two. If the economy is growing very slowly, then this difference is likely to be small, so the contribution to net national savings from retirement savings will be small.

PRIVATIZATION OF SOCIAL SECURITY

ADVANTAGES	DISADVANTAGES
Increase savings	Same impact can be had by switching to a fully funded Social Security system.
Enforce hard budget constraint	Restricts ability to engage in redistributions and risk sharing within and across generations.
Increase returns	Higher transactions costs may actually lower returns—probably signficantly.
	Higher returns are obtained only at the expense of higher risk, putting retirement security into jeopoardy.
	Transition requires taxes to fund unfunded liability, lowering welfare (returns) for transition generations.
	If individuals invest in excessively risky investments, they may become public charges in old age; avoiding this was one reason for introducing Social Security in first place.

Privatization faces one more problem: private markets for insurance are subject to the problems discussed earlier and in Chapter 13 for health insurance. In particular, there are strong incentives for cream skimming. By contrast, our public social insurance system provides an additional opportunity for limited redistributions within and across generations.

Given the proclivity of the financial sector to engage in exploitation of those who are financially less informed; given the incapacity of the sector to manage risk, demonstrated so forcefully in the 2008 crisis; and given the much lower transactions cost of Social Security, it is not surprising that today there is little interest even in partial privatization of Social Security.

REVIEW AND PRACTICE

SUMMARY

1. Governments provide social insurance because markets failed to provide insurance against many of the most important risks facing individuals. Private insurance markets face problems of adverse selection, moral hazard, and high transactions costs, and of insuring against social risks.

2. Another important motivation for Social Security is a moral hazard problem: if the government provides substantial assistance for the aged who are destitute because they have made no provision for their old age, then individuals will have insufficient incentive to provide for their own old age.

3. The Social Security retirement program is a forced savings program, an insurance program, and a transfer program.

4. The Social Security program has an effect on labor supply (most notably through its effect on early retirements) and capital formation (through its effect on savings). There is dispute about the significance of these effects.

5. Changes in birth rates and life expectancy, in labor force participation among the aged, and in the rate of growth of productivity all contributed to the financial difficulties facing the Social Security system.

6. Although in recent years several of the inequities in the Social Security program and features that contribute to inefficiencies have been corrected, several remain, including the differential treatment of families in different circumstances.

7. Restoring financial viability requires either changing benefits, such as by correcting the adjustment for increases in the cost of living, adjusting the benefit formulae, or means testing benefits; or increasing revenues, such as by subjecting Social Security benefits to taxation. Many of these reforms would increase equity and/or efficiency.

8. More drastic reforms include investing a portion of the Social Security trust fund in equities. Since the 2008 financial crisis, however, proposals to privatize have lost favor.

KEY CONCEPTS

Annuity

Fully funded system

Indexing

Modified pay-as-you-go basis

Social risks

Social Security

QUESTIONS AND PROBLEMS

1. For each of the major aspects of the Social Security program (retirement insurance, survivors' insurance, disability insurance) describe the market failures that gave rise to the program or that might be used to justify its continuation. Assume you were asked to design a program that was to address only one of the market failures. For as many market failures as you can, describe an alternative program to the present system and explain its advantages and disadvantages over the present one.

2. List the risks against which the Social Security program provides insurance. In which of these instances do you think providing insurance affects the likelihood of the insured against events occurring?

3. What are the theoretical reasons why Social Security might be expected to decrease savings? Are there any theoretical reasons why Social Security might be expected to increase savings? Why might a tax on interest income lead to later retirements? Under what circumstances might such a tax be desirable?

4. Discuss the equity and efficiency effects of the following recent and proposed changes in the Social Security system and, where appropriate, provide alternative reforms directed at the same objective:

   a. Changing the formula by which Social Security benefits are calculated to reflect total contributions, not just the contributions in a limited number of years.

   b. Tougher eligibility standards for disability payments.

   c. Increasing the eligibility age for Social Security benefits.

   d. Increasing benefits for those who retire later so that they receive the same expected present discounted value of benefits, regardless of age of retirement.

   e. Exempting those over the age of 65 from paying Social Security taxes.

   f. Lowering the cost of living adjustment for Social Security benefits.

5. To what extent could the purposes of the Social Security program be served by a law requiring individuals to purchase retirement insurance from a private firm? Discuss the difficulties with such a proposal, and what kinds of regulations might be required to avoid these difficulties.

6. Do you think Social Security benefits, unemployment benefits, or disability payments should be treated like ordinary income for purposes of the income tax?

7. Smokers do not live as long as nonsmokers. This means that the expected present discounted value of Social Security benefits for smokers is much smaller than that for nonsmokers. Is this unfair? What differences in individual or family circumstances should the Social Security program recognize?

8. If individuals are risk averse, why do they care more about large losses—that is, why are they willing to pay a larger amount in excess of the actuarial value of the losses for insurance? What does this imply for the design of social insurance?

9. Consider John, a 68-year-old individual facing a 39.6 percent marginal tax rate, with no wage income. Assume that if he earns a dollar, he loses 50 cents in Social Security benefits, and that 85 percent of his Social Security benefits are included in his taxable income.

   a. Calculate his total marginal tax rate.

   b. If he lives in California and faces an 11 percent state income tax, what is his marginal tax rate?

c. Assume he is self-employed, and must pay an additional 15 percent self-employment tax (a Social Security tax), but that his Social Security payments do not increase at all. Half the tax is deductible from his income for purposes of his income tax. What is his marginal tax rate now?

10. Consider a simple life-cycle model, in which individuals live for two periods; they work in the first, saving for the second. Let C_1 denote consumption in the first period and C_2 in the second; r, the interest rate, and w, the wage. Then

$$C_2 = (w - C_1)(1 + r). \qquad (1)$$

Assume there is no growth, and savings are used to buy capital goods, which last for one year, so

$$K = w - C_1, \qquad (2)$$

where $K =$ capital stock. The gross output of the economy, Q, increases with the capital stock,

$$Q = F(K), \qquad (3)$$

the interest rate is the marginal product of capital,

$$1 + r = F'(K), \qquad (4)$$

and workers receive what is left over,

$$wL = F - K F'(K), \qquad (5)$$

where L is the labor supply, which for convenience, we set equal to 1.

Individuals supply labor inelastically, but decide how much to consume by maximizing $U(C_1, C_2)$, the two-period utility function, subject to equation 1.

Describe the market equilibrium. You may find the following graphical interpretation useful. Put C_2 on the vertical axis and C_1 on the horizontal. Plot the set of feasible combinations, using equations 1 through 5. For each K there is a value of the wage; for each value of consumption in the two periods there is a value of the marginal rate of substitution, the amount of consumption while working that individuals are willing to trade off for more consumption in retirement. In equilibrium, the marginal rate of substitution equals the interest rate. Through any point on the feasibility locus, you can draw a line with slope equal to $1 + r$ (this just depends on K), and you can draw the indifference curve. The equilibrium point along the feasibility locus is the point where the two are tangent *at that point*.

Now, explore how a pay-as-you-go Social Security system changes the equilibrium. Let T be the transfer from the younger generation to the older. Now $C_2 = (w - T - C_1)(1 + r) 1 T$. Show how the equilibrium may be affected, and why the capital stock may be smaller.

TAXATION IN THEORY

The next two parts are concerned with taxation. Part Five develops the general theory of taxation. Part Six examines in some detail both the design and implementation of taxation in the United States, as well as analyzes the U.S. tax system from an international comparative perspective.

First, Chapter 17 sets out the general principles of taxation. Chapter 18 discusses who bears the burden of taxation. Chapter 19 explores the effects of taxation on economic efficiency, and Chapter 20 shows how equity and efficiency considerations may be balanced off against each other. Chapter 21 considers one of the most important problems of tax design: how to tax the returns to capital.

17

INTRODUCTION TO TAXATION

FOCUS QUESTIONS

1. What are the five key attributes that a good tax system should have?

2. What are the various ways in which tax systems affect economic efficiency?

3. What are the difficulties in determining what is a "fair" tax system?

Unlike most transfers of money from one individual to another, which are entered into voluntarily, taxation is compulsory. Chapter 5 showed that contributions to support public services need to be compulsory because of the free rider problem: unless support for public goods is made compulsory, no one will have an incentive to contribute. The discussion showed how all individuals may be made better off by being compelled to contribute to the support of public goods.

Government's ability to compel individuals' support of public goods, however, may also enable it to compel individuals' support of special interest groups: it can force one group to give up its resources to another group. Such forced transfers have been likened to theft, with one major difference: transfers through the government wear the mantle of legality and respectability conferred upon them by the political process. When the political process in a country becomes detached from the citizenry and is used to transfer resources to the groups in power, the distinction between taxation and theft becomes blurred at best.

These issues were of critical concern to the founders of the United States. The rebellion that became the Revolutionary War is often dated to

the Boston Tea Party, which was motivated by the conviction that unjust taxes were being levied on the colonies. The slogan "Taxation without representation is tyranny" provided one of the central motifs of the revolution. Distinguishing between the legitimate and illegitimate uses of the power of taxation is a matter of continual contention.

BACKGROUND

Taxation is as old as organized government. The Bible says that a tithe (one-tenth) of the crops should be set aside for purposes of redistribution and for the support of the priesthood. It is not clear what the enforcement mechanism was then, and the Bible does not report on the extent of tax evasion. Chinese emperors dating back to 300 BC generated a significant share of their revenues by taxing salt. In the Middle Ages, individuals provided services directly to their feudal lords; these were effectively taxes but they were not monetized. The fact that vassals were forced to provide these services meant that they were, to some extent, slaves. Some have argued that the fact that modern taxes are **monetized**—individuals are compelled to provide not services (except in the special case of the military draft), but money—should not obscure the underlying relationships. An individual who must give the government, say, one-fourth of his or her earnings is effectively working for the government one-fourth of the time.

There are, however, two critical distinctions between feudal levies and modern taxes. In the former case, individuals were not allowed to leave their manor without the permission of their lord; today individuals can choose where they wish to live, and therefore the jurisdiction that will impose taxes on them. Second, whereas under the manorial system individuals were compelled to work, in modern taxation individuals are compelled only to share with the government what they receive from working (or what they receive from investing, or what they spend). They can choose to pay less if they are willing to work less and receive less for themselves.

In the United States, concern about the possible abuse of the power to tax led to certain constitutional restrictions on the kinds of taxes that could be imposed. For instance, because the agricultural South—then the major exporting region of the country—was afraid that the more populous North would impose export levies, forcing southerners to bear a disproportionate share of the costs of government, such levies were explicitly barred by the Constitution. Other provisions of the Constitution attempted to ensure that taxes would not be imposed in a discriminatory manner.

For example, the uniformity clause says that taxes must be imposed in a uniform way, and the apportionment clause says that direct taxes must be apportioned among the states on the basis of population. Such constitutional restrictions were interpreted to imply that the government could not impose an income tax, and it was not until a constitutional amendment was passed in 1913 that the federal government could impose such a tax.

The restrictions on taxation reflected the experiences of the American colonies with discriminatory taxes levied by the British government. The writers of the Constitution did not anticipate—and probably could not have anticipated—all of the forms of discriminatory taxation. Thus, despite the safeguards that the founders of the republic attempted to provide through the Constitution, issues of taxation have been among the most divisive facing the country. For instance, in the early nineteenth century there was considerable controversy over tariffs.[1] Although they raised revenues, tariffs on industrial goods also served to protect the industrial North, and the South suffered by having to pay higher prices for the protected goods.

FORMS OF TAXATION

The variety of taxes governments have levied has been huge. At various times, there have been taxes on windows, luxury boats, sales of securities, dividends, capital gains, and many more items. Taxes can be divided into two broad categories: *direct taxes* on individuals and corporations; and *indirect taxes* on a variety of goods and services.

In the United States, the three principal direct taxes at the federal level are: the individual income tax; the payroll tax, a fixed percentage of wages up to some limit, used to finance Social Security; and the corporation income tax, a tax on the net income of corporations. Another important direct tax is the estate and gift tax, which is primarily a tax on bequests from one generation to the next. Because the individual income tax is levied both on wages and on capital income, it affects decisions about labor supply, retirement, education, and so forth, as well as decisions about savings and investment. At the state and local level, there is one other important direct tax, on property.

The principal indirect taxes at the federal level are customs duties, levied on imports of goods from abroad; and excise taxes, on goods such as

---

[1] Tariffs are taxes imposed on imported goods. By raising the prices of the imported goods, they enable domestic producers of similar goods to raise their prices as well. In this way, tariffs "protect" domestic producers.

Other (4.7%)
Excise taxes and customs duties (4.5%)
Taxes on corporate income (including federal reserve revenue) (12.7%)
Personal current taxes (36.7%)
Contributions for Social Security (41.4%)

Federal Government Current Receipts, 2010: $2386.6 billion
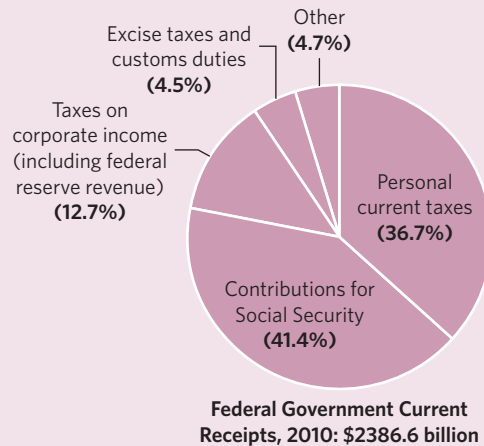
telephone service, air travel, and luxuries. Many states and local jurisdictions also impose a sales tax, a flat percentage tax on all retail sales of a broad category of goods. Some states exempt food from their sales tax, whereas others have a broader range of exemptions. Many foreign governments, rather than imposing a sales tax at the retail level, impose a value-added tax: at each stage of production, the value added is the difference between the value of the sales and the value of purchased (non-labor) inputs.

Figures 17.1 and 17.2 show the sources of revenues at the federal, state, and local levels in 2010. At the federal level, Social Security taxes and individual income taxes together comprise almost four-fifths of revenues (78.1 percent), and the corporate income tax accounts for 12.7 percent. All other taxes are small by comparison. However, at the state and local level, individual income and payroll taxes together make up just over one-seventh of revenues (14.8 percent); transfers from the federal government contribute a quarter of revenues, whereas property and sales taxes each account for another fifth.

CHANGING PATTERNS OF TAXATION IN THE UNITED STATES

The establishment of the income tax with the passage of the Sixteenth Amendment in 1913 marked a turning point in the structure of taxation

Personal income taxes
(12.4%)

Payroll taxes
(2.4%)

Taxes on
corporate income
(4.1%)

Other
(15.5%)

Sales taxes
(20.2%)

Federal
grants-in-aid
(24.9%)

Property
taxes
(20.5%)

**State and Local Receipts,
2010: 2128.1 billion**
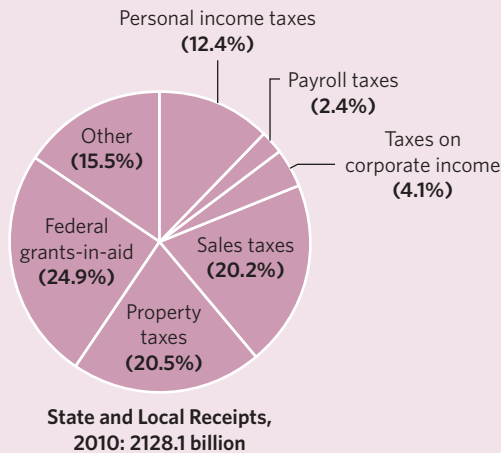
FIGURE 17.2

**SOURCES OF STATE AND
LOCAL REVENUES, 2010**
Sales taxes, property taxes, and
federal grants contribute about
equal amounts to state and
local revenues. Income taxes
are less important.

SOURCE: U.S. Department of
Commerce, Bureau of Economic
Analysis, *National Income and
Product Accounts*, Table 3.3.

in the United States. Prior to that, the principal sources of federal revenues were excise taxes and customs duties. During the twentieth century, these have dwindled in importance.

Figure 17.3 shows the changes in the relative importance of various taxes during the past century. In particular, we see (a) a marked increase in the relative importance of taxes imposed directly on individuals and a steep drop in the importance of indirect taxes, and (b) within direct taxes, a sharp decrease since 1950 in the role of the corporate income tax and a marked increase in the role of payroll taxes.

There have also been marked trends at the state and local level, reflected in Figure 17.4: an increased reliance on personal income taxes and federal grants, and a decreased reliance on property taxes.

COMPARISONS WITH OTHER COUNTRIES

During recent decades, most countries have switched to an increased reliance on value-added taxes. As is evident in Figure 17.5, the United States and other Anglophone countries rely much more heavily on the individual income tax than do most other nations, whereas middle-income countries such as Chile, Mexico, and Turkey are particularly dependent on the value-added tax.

FIGURE 17.3

SOURCES OF FEDERAL REVENUE, 1940–2010
Customs and excise taxes, and the corporate income tax, have diminished in importance over time, whereas the dependence on payroll taxes for social insurance has grown.

SOURCE: U.S. Department of Commerce, Bureau of Economic Analysis, *National Income and Product Accounts*, Table 3.2.
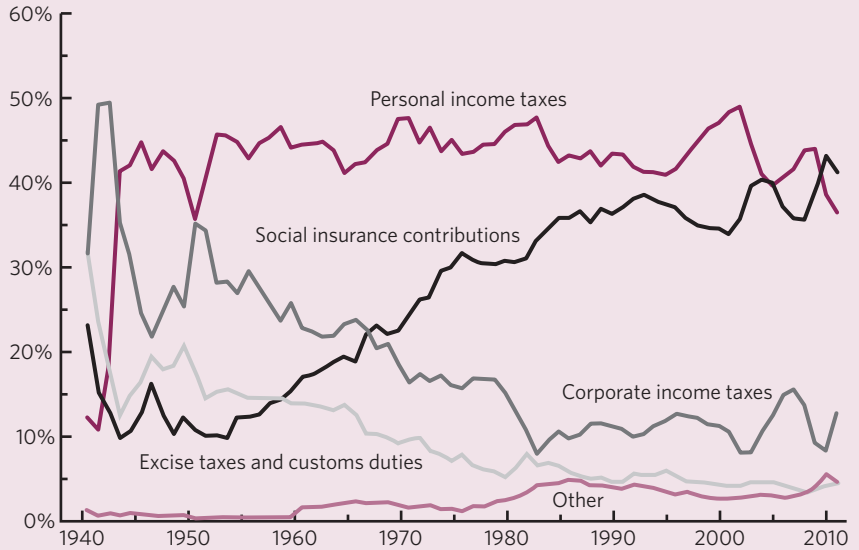
FIGURE 17.4

SOURCES OF STATE AND LOCAL REVENUES, 1940–2010
The contribution of property taxes to state and local revenues has fallen substantially over the past seventy years, whereas the importance of all other sources (particularly personal income taxes and federal grants) has grown.

SOURCE: U.S. Department of Commerce, Bureau of Economic Analysis, *National Income and Product Accounts,* Table 3.3.

*unweighted average

THE FIVE DESIRABLE CHARACTERISTICS OF ANY TAX SYSTEM

Taxes are inevitably painful. Not surprisingly, designing tax systems has always been a subject of considerable controversy. To put it most simply, most people would like to see their own taxes reduced. Quite ingenious arguments can be devised for why others should pay more.

Governments, in deciding how best to raise the revenue they require, have looked for general principles. There are five accepted properties of a "good" tax system:

1. *Economic efficiency*: the tax system should not interfere with the efficient allocation of resources.

2. *Administrative simplicity*: the tax system ought to be easy and relatively inexpensive to administer. Good tax systems rely on self-compliance, so the system should be designed to make compliance easy and voluntary.

PRINCIPLES OF TAXATION

- *Efficiency*: the tax system should not be distortionary; if possible, it should be used to enhance economic efficiency.
- *Administrative simplicity*: the tax system should have low costs of administration and compliance.
- *Flexibility*: the tax system should allow easy adaptation to changed circumstances.
- *Political responsibility*: the tax system should be transparent.
- *Fairness*: the tax system should be, and should be seen to be, fair—treating those in similar circumstances similarly, and imposing higher taxes on those who can better bear the burden of taxation.

3. *Flexibility*: the tax system ought to be able to respond easily, in some cases, automatically, to changed economic circumstances.

4. *Transparent political responsibility*: the tax system should be designed so that individuals can ascertain what they are paying, and evaluate how accurately the system reflects their preferences.

5. *Fairness*: the tax system ought to be fair in its relative treatment of different individuals.

In many developing countries (and even in some developed countries) there is an important sixth attribute: the tax system should be "**corruption resistant**." (See case study, "Corruption-Resistant Tax Systems," which describes what this means.)

ECONOMIC EFFICIENCY

Recall that in the absence of market failures, the economy would automatically allocate resources efficiently. Information conveyed by market prices would lead to production, exchange, and product mix efficiency, and no one could be made better off without making anyone worse off. Most taxes change relative prices. As a result, the price signals are distorted, and the allocation of resources is altered.

A persistent concern is the extent to which the tax system discourages savings and work, and distorts other decisions relating to consumption and production. For instance, the large number of Arabian and other very expensive breeds of horses in the United States has been attributed to a peculiar loophole in the tax structure. The special treatment of gas and oil may have led to excessive drilling. Railroad boxcars were used for a while as a tax shelter, until a glut of these developed.

The history of taxation is dotted with other examples of distortionary effects. The result of the window tax imposed in Britain during the seventeenth century was that houses were constructed without windows. Modern England provides other examples. Three-wheeled vehicles, though perhaps slightly less safe and not much less costly than four-wheeled vehicles, were taxed more lightly than the latter. Hence many individuals chose them in preference to the more conventional four-wheeled vehicles. Vans (station wagons without windows) were taxed

more lightly than station wagons with windows, and again, many individuals were motivated to purchase these vehicles, though not by a preference for darkness in the rear of their vehicle. In the United States, the favorable depreciation treatment for movable walls[2] has encouraged the building of commercial office buildings with movable walls—even if there is no intention of ever moving them. Although movable walls do provide greater flexibility, they typically provide less sound insulation. Thus, even though we no longer have a window tax, modern tax laws nonetheless do affect construction patterns.

BEHAVIORAL EFFECTS OF TAXATION   Most of the efficiency effects of taxation are far more subtle and difficult to assess. Income taxation may affect the length of time an individual stays in school by affecting the after-tax return to education, the choice of jobs (because for some jobs a larger fraction of the return comes in untaxed "benefits"), whether an individual enters the labor force or stays at home to take care of children, the number of hours a taxpayer works (when he or she has discretion over that), whether he or she takes a second job and the effort put into the job, the amount that the individual saves and the form savings take (the choice between bank accounts and the stock market), the age at which an individual retires, and whether he or she works part-time beyond age 65.

The effects of taxation are not limited to decisions concerning work, savings, education, and consumption. Although the extent to which the tax system affects individuals' decisions to marry or divorce is open to question, there is little doubt that it affects the timing of these decisions. For instance, the U.S. tax code considers a couple married for the entire calendar year even if the wedding is held on December 31. Thus, two working people who earn similar incomes, choosing between a December and January wedding date, are strongly encouraged by the income tax to choose January. The reverse is true for divorce.[3] Taxation affects risk taking, the allocation of resources to research and development, and the long-run rate of growth of the economy. It affects not only the level of investment in firms but also the form of the investment, including the durability of machines. It affects the fraction of national savings that is allocated to housing, to other structures, and to equipment. It affects the rate at which our natural resources are depleted. There is hardly an important resource allocation decision in our economy that is unaffected in some way or another by taxation.

---

[2] Taxpayers are allowed to deduct from their income an amount that reflects the wearing out (or depreciation) of their plant or equipment. Movable walls are treated as equipment, and thus can be depreciated much more rapidly than fixed walls. As a result, the present discount value of the depreciation allowances is much higher.

[3] Married couples are taxed on the basis of their combined income. Because the income tax is progressive, the income of the lower-earning spouse is taxed at a higher rate than it would be if the individual were single. (See Chapter 22.)

With tax rates at current levels, tax considerations are often of primary concern; it may seem more advantageous to allocate one's effort to reducing one's taxes than to designing better projects or producing more.

FINANCIAL EFFECTS OF TAXATION   Sometimes taxation affects a transaction's form more than its substance. For instance, apart from tax considerations, it may make little practical difference whether an employer gives an employee money to purchase a health insurance policy or purchases it for the employee. In tax terms, though, it makes a great deal of difference. In the former case, the individual receives "income" on which he or she is taxed; in the latter case, the "benefit" is not taxed. Similarly, in real terms, it makes little difference whether I save directly for my retirement or my employer takes some of my salary and invests it in a (fully funded) pension plan. However, the tax implications are quite different, and as a result, individuals are induced to save through the pension plan rather than directly. These financial effects may in turn, of course, have further real effects on the economy: employers, because of the restrictions imposed on pension plans, may invest their funds differently from the way an individual saving for retirement might invest them. Typically, they cannot, for instance, invest in highly risky "below investment grade" securities. Moreover, some individuals may be "forced" to save through their pension plans more than they would voluntarily save on their own.

Similarly, because dividends, capital gains (increases in the price of an asset), and interest are all treated differently, the tax structure may have a significant effect on the *financial* structure of U.S. corporations—for instance, on firms' decisions whether to finance additional investments by borrowing or by issuing new shares. These financial decisions, in turn, have real consequences. A firm with a heavy debt burden is likely to be less willing to undertake risky projects than a firm that has raised most of its funds by issuing stock, and, in the current economic downturn, it is now likely to go into bankruptcy.

ORGANIZATIONAL EFFECTS OF TAXATION   Taxes affect the way our economy is organized. Many of these organizational effects have real consequences for how resources get allocated—for instance, for how much risk taking occurs. Our tax laws differentiate between corporations, which have limited liability, and individuals and partnerships, which do not. Because with unlimited liability the losses that an investor can incur from a $100 investment are far greater than just the $100 he or she has invested, without limited liability enterprises often have a difficult time raising capital, and managers of enterprises with unlimited liability may act in a far more risk-averse manner. By tilting the

tax system either for or against corporations, the tax system can either encourage or discourage economic activity from taking place within corporations, thereby changing the degree and nature of risk taking in the economy.

Financial effects are often intertwined with organizational effects. The tax system may encourage or discourage banks relative to other financial institutions or arrangements; this may lead firms to raise more or less money through banks (rather than, say, on the stock or bond market). This can make a great deal of difference. Recent studies have shown, for instance, that firms that raise more of their money through banks exhibit less volatility in investment, partly because the bank is better able to monitor why the firm may be short of funds—whether, for instance, it is because of short-run cyclical effects or because of problems of mismanagement. Because banks can better identify the nature of the problem, they can respond effectively when there is a cyclical downturn, providing capital to good firms that are temporarily short of cash and withholding funds from firms that have more fundamental weaknesses. By contrast, firms that rely on capital markets (issuing bonds or new equities) to raise funds find life far more difficult in a downturn; they may find it virtually impossible or extremely expensive to raise funds at such times.

Another major economic organization within our society is the family. Tax laws affect both family formation and the distribution of well-being within a family. For instance, tax laws typically tax only payments from corporations to households (that is, to their employees). They do not tax consumption within corporations. Thus, tax laws encourage "inside the firm" consumption—large company cars, expensive company meals, and so forth. In those economies, such as the United States before World War II, where men typically worked outside the home and women within the home, such tax policies obviously had the effect of discriminating in favor of the spouse working outside the home.

GENERAL EQUILIBRIUM EFFECTS   The imposition of a tax such as that on wages or on the return to capital alters the equilibrium of the economy. A tax on interest may reduce the supply of savings and, eventually, the stock of capital; this in turn may reduce the productivity of workers and their wages. We refer to these indirect repercussions of the tax as its **general equilibrium effects**.

General equilibrium effects have important distributional consequences, sometimes in a direction quite opposite to the intent of the legislation. A tax on capital may reduce the supply of capital, thereby increasing the return to capital; in some instances, the degree of inequality may actually be increased by such a tax.

ECONOMIC IMPACTS OF TAXATION

Behavioral effects

- Work, education, retirement
- Savings, investment, risk taking
- Energies devoted to avoiding taxes instead of creating wealth
- Marriage and divorce

Financial effects

- Fringe benefits
- Financial structure of firms

Organizational effects

- Corporations versus unincorporated enterprises
- Intertwined with financial effects (banks versus insurance versus other forms of finance)

General equilibrium effects

- Often important indirect effects, especially with broad-based taxes, such as on wages or interest

Announcement effects and capitalization

- Future taxes on an asset reflected ("capitalized") in the price of the asset at the time the tax is announced

ANNOUNCEMENT EFFECTS AND CAPITALIZATION   The economy does not instantaneously adjust to a new tax. Often, the long-run distortions are much greater than the short-run distortions, as the economy is able to respond more fully to the new situation.

However, some effects of the tax may be felt even before it is imposed, simply upon its announcement. An announcement concerning the future tax treatment of an asset has an immediate impact on the value of the asset. For instance, if it is believed that a particular category of assets, say, housing, is about to be subjected to greater taxation (e.g., interest deduction on mortgages is about to be eliminated), then the price of that category of assets may fall markedly. Owners of those assets at the time the announcement is made will, perhaps unfairly, bear the major burden of the tax.

It is these **announcement effects**, or *impact effects*, which may be quite significant, that have given rise to the saying that "an old tax is a good tax." Not only may the announcement effect present serious equity problems, but anticipation of it can also affect the supply of assets. Discussions about eliminating the deduction of interest paid on home mortgages may lead individuals to anticipate significant capital loss were they to invest in housing, for instance, and hence the demand for new housing may be seriously reduced.

DISTORTIONARY AND NONDISTORTIONARY TAXATION   Any tax system influences behavior. After all, the government is taking money away from an individual, and we would expect that individual to respond, in some way, to this lower income. When we say that we want the tax system to be nondistortionary, clearly we do not mean that we want the individual not to react at all.

A tax is **nondistortionary** if, and only if, there is nothing an individual or firm can do to alter the tax liability. Economists call taxes that are nondistortionary **lump-sum taxes**. Distortions are associated with the individual's or firm's attempt to lower the tax liability. Virtually all taxes imposed in the United States are, in this sense, distortionary. A *head tax*—a tax one has to pay regardless of income or wealth—is a lump-sum tax. A tax that depends on nonalterable characteristics, such as age or sex, is also a lump-sum tax.

Because individuals and firms cannot avoid them, lump-sum taxes do not lead to changes in behavior or the reallocation of resources, other than the income effect of reduced after-tax income.

Any tax on commodities is **distortionary**: an individual can change his or her tax liability simply by reducing purchases of the commodity. Any tax on income is also distortionary: an individual can reduce his or her tax liability simply by working less or by saving less.

In Chapter 19, we shall see that distortionary taxes are inefficient in the sense that if the government could replace them with a lump-sum tax, it could raise more revenue, with the same effect on the welfare of individuals; or equivalently, the government could raise the same revenue and increase the welfare of individuals.

> ## TAXES AND ECONOMIC EFFICIENCY
>
> - All taxes affect behavior (reduce spending power); *distortions* are associated with actions an individual takes to avoid taxes—for example, by working less, one's tax liability is reduced.
> - Lump-sum taxes are taxes that are fixed and cannot be altered by any action that the individual can take.
> - Corrective taxation, by imposing taxes on activities that generate negative externalities, such as pollution, simultaneously raises revenues and improves economic efficiency.

CORRECTIVE TAXATION   So far, this discussion has emphasized the negative aspects of taxation: that a tax system should be designed so as not to interfere with economic efficiency. Recall, however, that in the presence of market failures, the allocation of resources will not, in general, be efficient. Taxation can sometimes be used in a positive way, to correct some market failure. As discussed in Chapter 6, taxation could sometimes be used to correct for externalities. **Corrective taxes** (as these taxes are called) both raise revenue and improve the efficiency of resource allocations. The United States has made limited use of corrective taxes. The tax imposed on the chemical industry to pay for the costs of cleaning up and disposing of toxic wastes can be thought of as a corrective tax. Those who view American energy consumption as excessive (because of the associated pollution) have argued for an energy tax, which would both raise revenue and reduce profligate energy consumption. These concerns have increased with the awareness of the dangers of global warming associated with the buildup of greenhouse gases (such as carbon dioxide) in the atmosphere, largely resulting from high energy usage.

ADMINISTRATIVE COSTS

Administering our tax system entails significant costs. There are direct costs—the cost of running the Internal Revenue Service—and indirect costs, borne by taxpayers. These indirect costs, called *compliance costs*,

CORRECTIVE TAXES AND THE DOUBLE DIVIDEND

Corrective taxes have gained increasing support in two areas. The first was smoking. The recognition that smokers impose costs on others, including higher health costs, underlay the idea that a tax on cigarettes should help finance health care reforms, such as the provision of health insurance for poor children.

The second has to do with air pollution. Greenhouse gases lead to global warming, and are generated by the burning of fossil fuels. A carbon tax would "correct" this externality. President Clinton proposed a modification of such a tax in his 1993 budget, called the BTU tax, because it was a tax based on the amount of energy in a fuel. (Australia adopted such a tax in 2011, but as this book goes to press, efforts are underway to repeal it by Australia's conservative Coalition government, Liberals and Nationals.) Some forms of energy like burning of coal give rise to far more greenhouse gases than others. (Hydroelectric power gives rise to none.) Hence, the BTU tax was only an imperfect substitute for a "carbon" tax. However, the industries that were high users of energy, as well as the oil, coal, and gas industries, mounted a successful campaign against the tax, and Congress passed instead a 4.3 cent increase in the gasoline tax. This, too, can be viewed as a corrective tax—it helps correct for the externalities associated with auto pollution and congestion.

There is still no nationwide carbon tax in the United States, although the tax has been introduced in a few subnational jurisdictions, such as the nine counties in the San Francisco Bay Area covered by the Bay Area Air Quality Management District. The EU has had similar problems introducing a comprehensive carbon tax, but several countries have imposed energy taxes based partly on carbon content.

Such taxes are attractive because they raise revenues at the same time that they correct a market failure. Thus, less revenues have to be raised through other, distortionary, methods. There is, accordingly, a "double dividend" associated with such a tax. The economy benefits from the reduced pollution *and* from the reduced reliance on taxes that distort production. The argument for corrective taxation is sometimes put another way: Why should society tax productive ("good") economic activities, such as savings and hard work, rather than bad economic activities, like pollution?

take on a variety of forms: the costs of time spent filling out tax forms, costs of record keeping, and the costs of services of accountants and tax lawyers. Joel Slemrod of the University of Michigan has estimated that the indirect costs are at least five times greater than the direct costs.

The administrative costs of running a tax system depend on a number of factors. First, they depend on what records would be kept in the absence of taxation. Businesses need to keep records for their own internal management purposes: the advent of high-speed computers has greatly reduced the costs of record keeping for large corporations. Thus, the tax system imposes a relatively small additional burden on large

corporations for reporting wage income of their employees. At the other extreme, many small businesses and most households that hire domestic help find the additional record keeping and filings required by the income tax very burdensome. In 1993, the government allowed those with household employees to file their tax reports on their employees with their own individual income tax, though they still have to make separate filings with the Social Security Administration for unemployment insurance, state income taxes, and sometimes local income taxes.

The record keeping required for capital gains taxation is particularly onerous because the records often have to be kept over a long period of time. Indeed, record keeping associated with taxation on owner-occupied housing was sufficiently onerous that few fully complied with the law; finally, in 1997, the tax laws were changed to exempt almost all capital gains from owner-occupied housing from taxation.

A second factor that determines the administrative costs of a tax system is its complexity. Much of the cost of administering the income tax system comes from special provisions. For instance, the deductibility of certain categories of expenditures (medical, charity, interest) requires that records be kept on these expenditures.

Differentiation of rates across individuals (with some individuals paying a much higher rate than others) and across categories of income gives rise to attempts to "shift" income to members of one's family with lower tax rates, or to categories of income that are more lightly taxed. Attempts to restrict this shifting also account for much of the complexity of the current tax structure.

Third, taxing some categories of income may be more expensive than taxing others. It is widely believed that the administrative costs associated with imposing taxes on capital are much larger than those associated with taxing labor, partly because of the difficulty of differentiating between income and the return of capital. For instance, payments to capital owners may be "income" (dividends) or "principal" (the repayment of previously invested funds); the dollars look the same. If the tax law treats these payments differently, taxpayers will be moved to characterize the dollars one way or the other. The government has had to write elaborate rules, which only partially address the problem.

Similarly, the administrative costs of raising taxes (per dollar of revenue raised) from small businesses may be much larger than those from large corporations. Thus, the administrative costs of the value-added tax, in which a large fraction of the revenue is raised from large corporations responsible for a significant fraction of the economy's value added, is lower than the administrative costs of a sales tax, which imposes taxes only at the final sale, in the myriad of retail outlets.

FLEXIBILITY

Changes in economic circumstances require changes in tax rates. For some tax structures, these adjustments are easy; for some, they require extensive political debate; still for others, they occur automatically.

AUTOMATIC STABILIZATION   For instance, as the economy goes into a recession, a reduction in tax revenues may be extremely desirable, to provide needed stimulus for the economy. When prices are stable, a progressive tax structure will provide "automatic" stabilization. When incomes drop, as a result of a recession, the average tax rate is reduced—individuals face lower tax rates because their incomes are lower. On the other hand, when income increases, the average tax rate increases. However, before tax brackets were indexed (i.e., adjusted to take account of inflation) in 1981, during periods of *stagflation*—when the economy was in a recession but there was still inflation—the average tax rate increased, although a lower rate was needed to move the economy out of the recession. Indexing thus contributes to stabilization when prices rise during recessions. During periods of expansion and inflation, indexing reduces the built-in stabilizing effects of the income tax.

POLITICAL DIFFICULTIES OF ADJUSTING RATES   When changing the tax rates is considered desirable, attempts to adjust the U.S. income tax often occasion intense political debate. Given the complexity of the tax code, which rates ought to be adjusted? Should all rates be increased proportionately, or are the rich or the poor already bearing a disproportionately large share of the tax burden, so that their taxes should be increased less than proportionately? Indeed, it is not even clear how to assess the fairness of a reform proposal. Is it fairer to reduce the taxes of individuals at different income levels by the same *dollar* amount, or by the same *percentage* amount? Should focus be on the average tax rates individuals pay, or on their marginal tax rates? Is a tax reform fair if it lowers the average rate faced by a one-earner family but increases the average rate on a two-earner family? Should the tax rate on capital be decreased, to encourage more savings, or increased, because capital owners are in a better position to bear the tax?

The political difficulty of adjusting the income tax rate should be contrasted, for instance, with that of the property tax. The property tax is beset by a number of administrative problems, not least of which is the difficulty of assessing the value of various pieces of property. Still, it has one advantage: adjustments in the tax rates are made annually in a simple manner as the revenues required for the provision of local public services change.

SPEED OF ADJUSTMENTS   Finally, an important aspect of the "flexibility" of a tax system for purposes of stabilizing the economy is timing: the speed with which changes in the tax code (once enacted) can be implemented, and the lags in the collection of funds. If fluctuations in the economy are rapid, the lags may limit the efficacy of, say, the income tax, in stabilizing the economy. There is always the danger that with sufficiently long lags, taxes will be increased just when the economy needs a tax reduction, and vice versa.

TRANSPARENT POLITICAL RESPONSIBILITY

A widely embraced political value is that government should not try to take advantage of uninformed citizens. In the context of taxation, this view recommends taxes for which the burden of payment is clear. Such taxes are known as **transparent taxes**, and transparency has increasingly been recognized as an important characteristic of good government. Government policies are said to be *transparent*[4] when they are subject to daylight—when it is clear who is benefiting and who is paying. In this view, the individual income tax is a good tax.

Sometimes it seems as though the government deliberately misrepresents the true costs of the services it provides or who bears the costs. For instance, there is widespread agreement that there is no meaningful distinction between the part of the Social Security tax that is paid by the employer and the part paid by the employee. (According to law, half is paid by each.) The employer is concerned only with the total costs of the employee, the employee only with his or her take-home pay. No one's economic behavior should be affected if it were announced that the entire tax was to be borne by the employee, were employers to give an equivalent pay raise to their employees to cover the increased tax. Would workers' attitudes toward Social Security be altered if they thought they had to bear the entire costs?

In some cases there is an almost deliberate attempt to persuade individuals that the cost of government is less than it is. Just as businesses find that they can sell cars more easily if they describe the cost as "only $340 a month for a short 40 months" than if they describe it as "$13,600 paid over three and a half years," so, too, governments sometimes show a preference for tax systems in which individuals never fully reckon the cost of government. One of the arguments put forward for sales taxes is that they

---

[4] The term has taken on a particular meaning in some recent discussions. The nongovernmental organization (NGO) Transparency International focuses on identifying corrupt practices and governments. Its view is that it is lack of transparency that gives rise to much of the political corruption one sees.

are less noticed than other taxes, such as income taxes. Individuals never calculate the *total* amount they pay to the government—but the sales tax is at least usually listed separately on receipts in the United States, whereas the value-added tax in Europe is traditionally included in the purchase price, making the tax burden even less transparent.

Jean-Baptiste Colbert, a finance minister to Louis XIV, wrote: "The art of taxation consists in so plucking the goose as to obtain the largest amount of feathers with the least possible amount of hissing." From this perspective, the corporation tax may thus be viewed to be a good tax. Politicians can claim that it is anonymous corporations that pay it. From the perspective of transparency, though, the corporation tax is one of the worst, because who really pays the tax is not apparent. Taxes are paid by people, not by institutions: it is the shareholders, workers, and customers who ultimately bear the burden of the tax.

However, there is a good reason for levying the corporate income tax: in the absence of such a tax, the profits of a corporation would not be taxed until they were distributed. A corporation thus becomes like an individual retirement account (IRA)—a tax free way of accumulating savings. That is inequitable, and often leads to economic inefficiencies—inducing money to stay within the corporation, even if the managers of the corporation do not invest it as well as it could be. In some cases, this problem can be resolved by integrating the corporation and individual income tax: attributing the income of the corporation to the shareholders, and requiring them to pay a tax on the corporation's income as if it were distributed, whether it is or not.

In practice, as globalization has proceeded, corporations have become skilled at avoiding taxes; for instance, claiming that their profits are earned in low-taxed jurisdictions. As a result, there has been a steady erosion in the amounts raised by the corporation tax (see Figure 17.3). For example, from 2009 to 2012, Apple Operations International, registered in Cork, Ireland, generated $30 billion in overseas profits without paying any corporate income tax on this to the Irish, United States, or any other national government.

A politically responsible tax structure is also one in which changes in taxes come about as a result of legislation, and with which the government must repeatedly come back to the citizenry for an appraisal of whether it is spending too much or too little. Steeply progressive tax rates (rates that rise as incomes rise) combined with a tax system that does not adjust for inflation result in government's tax revenues in *real* terms (as a share of, say, national income) rising in inflationary times, as they did between 1975 and 1980. These increases in taxes were never directly legislated: indeed, many would argue that Congress would have been unlikely to impose

directly, say, a 10 percent increase in taxes in 1980, although inflation had exactly this effect.

FAIRNESS

Most criticisms of tax systems begin with their unfairness. However, as we shall see, it is difficult to define precisely what is or is not fair. There are two distinct concepts of fairness: horizontal equity and vertical equity.

HORIZONTAL EQUITY   A tax system is said to be *horizontally equitable* if individuals who are the same in all relevant respects are treated the same. The principle of **horizontal equity** is so important that it is, in effect, enshrined in the Constitution as the Fourteenth Amendment (the Equal Protection Clause). Thus, a tax system that discriminates on the basis of race, color, or creed would, in the United States, generally be viewed to be horizontally inequitable (and unconstitutional). Although the underlying idea is clear enough, there are two fuzzy notions in our definition: What does it mean for two individuals to be identical in all relevant respects? And what does it mean for two individuals to be treated the same?

Consider twins who are identical in every respect except that one likes chocolate ice cream and only chocolate ice cream, while the other likes vanilla ice cream and only vanilla. For simplicity, we assume that chocolate and vanilla ice cream cost exactly the same amount. Is the tax system treating the two individuals in a horizontally equitable manner if it taxes vanilla and chocolate ice cream at different rates? One ends up paying more in taxes than the other, and in this sense, the tax system appears to be unfair. But the twins faced the same "opportunity set." The chocolate lover could have bought vanilla ice cream if he or she had wanted to (or vice versa). The tax system did not discriminate; it did not differentiate between individuals. This example is contrived so we could have two commodities that are "essentially" identical. In practice, though, there are many examples in which the tax system gives different treatment to individuals who differ in tastes—the higher taxes on hard liquor, for instance, discriminate against those who prefer scotch relative to those who prefer wine or beer. Individuals who prefer to spend their vacations in their own vacation homes are treated preferentially compared to those who prefer to travel during their vacation.

If we say that the differences in taste are an important economic difference that the tax system may well take into account, then we can say that the principle of horizontal equity does not apply here. The twins are not identical in all relevant respects. Carried to this extreme, the principle

soon becomes vacuous: no two individuals are ever identical. What are to be acceptable distinctions? Unfortunately, the principle of horizontal equity gives us little guidance on how to answer this question.

One's first intuition might be that all distinctions are inadmissible: age, sex, and marital status should all be irrelevant. In fact, at present we make distinctions on the basis of age (those over 65 are allowed an extra exemption) and marital status (two individuals with the same income who marry pay more in taxes than they did before marriage). Congress has felt that those distinctions are relevant.

Perhaps age and marital status are relevant because they affect individuals' ability to pay. If these are admissible bases for differentiation, however, are there others? For instance, does variation in the economic costs associated with taxing different groups provide legitimate grounds for differentiation? In a later chapter we shall see that the inefficiencies arising out of a tax system depend on the magnitude of the responses to the tax. In households with two workers, the worker with the lower wage displays much more sensitivity to the wage rate than the higher-earning worker: income taxes may have large effects on the secondary worker, although they have almost no effect on the amount of labor supplied by the primary worker. Thus, if the government were concerned with minimizing the inefficiencies arising out of the tax system, it would impose a lower tax on the secondary workers. Is this fair? Another example illustrating the difficulties is provided by health care expenditures. Should two individuals with the same income but different health care expenditures be treated the same? Does it make a difference whether the health care expenditures are "voluntary" (e.g., face lift) or "necessary" (e.g., heart bypass surgery)? Can the government tell which is which?

The following example illustrates the difficulty of even defining the meaning of equal treatment. Assume we could agree that a man and a woman who had received the same income over their working lives should be treated equally for purposes of Social Security. Should the total expected benefits be the same for the man as for the woman, or should the annual benefit be the same? On average, women live significantly longer than men, so these two rules give different results. If the woman receives the same annual benefit as the man (as is the case at present), the total expected value of her benefits will be much greater than the man's. Many would view this to be unfair.

VERTICAL EQUITY   While the principle of horizontal equity says that individuals who are essentially identical should be treated the same, the principle of **vertical equity** says that some individuals are in a position to pay higher taxes than others, and that these individuals should do so.

There are three problems: (1) determining who, in principle, should pay at the higher rate; (2) implementing this principle—that is, writing tax rules corresponding to this principle; and (3) deciding, if someone is in a position to pay the higher rate, how much more that individual should pay than others.

INCOME AS A BASIS OF TAXATION   Income is the most widely used basis of taxation; it is widely viewed by governments and policy makers as a good measure of ability to pay. Those who have a higher income have greater ability to pay and should therefore pay higher taxes. How much more is, as we have said, a more difficult question. There is a widely held view that those with a higher income not only should pay more taxes, but should also pay a higher fraction of their income in taxes—that is, taxes should be progressive. Note, however, that the rich can pay a smaller fraction of their income in taxes—taxes can be regressive—but still pay more in absolute terms.

Until the twentieth century, governments relied on indirect taxes—tariffs and customs duties and taxes on certain luxuries—to raise revenues. It was only when governments took on a wider role, with a greater need for income, they resorted to broad-based taxes, especially the income tax. In addition, the income tax seemed able to introduce a high degree of progressivity and to avoid the distortions associated with having a large number of taxes on different commodities. Today, in much of the world, the income tax has become a less important source of revenue. It has been replaced, or supplemented, by the value-added tax, which is designed to tax only consumption, not savings or investment, and which is typically not progressive.[5]

There is a further concern about taxing only consumption, even if one could impose a progressive consumption tax: much of consumption, especially of the very rich, would likely escape taxation, for instance, if the consumption occurs abroad.

CONSUMPTION AS A BASIS OF TAXATION   One of the most forceful arguments against income as a fair basis of taxation is that income corresponds to the individual's contribution to society—the value of his or her economic output. Is it not fairer to tax individuals on the basis of what they take out of society rather than what they contribute—that is, on the basis of consumption, rather than income?

---

[5] Sometimes countries introduce some degree of progressivity by exempting, or taxing at lower rates, food and other commodities that play a larger role in the expenditure patterns of the poor; at the same time, some expenditures like foreign travel, which loom larger in the expenditures of the rich, escape taxation as well.

Income and consumption differ by savings.[6] That is, income (Y) is either consumed (C) or saved (S):

$$C + S = Y, \quad \text{or}$$
$$C = Y - S.$$

Thus, a major issue is whether savings ought to be exempt from taxation. It can be shown that this is equivalent to the question of whether the return to savings (interest, dividends, and capital gains) ought to be exempt from taxation. The following example illustrates again the conflicting views of equity.

Consider another pair of identical twins, to whom we shall refer as Prudence and Imprudence. They both earn the same wages during their lifetimes. Prudence, however, saves 20 percent of her wages during her lifetime, accumulating a sizable nest egg for her retirement. Imprudence, on the other hand, always spends what she receives and, when she reaches retirement, applies for welfare. Under the present tax system, Prudence pays considerably higher taxes than Imprudence (because Prudence must pay taxes on the interest that she earns on her savings), while she receives fewer government benefits.

Prudence views the present tax system as unfair, as their economic opportunity sets were, in fact, identical. Because their opportunity sets were identical, she believes they really have the same ability to pay, and should pay the same taxes. She asks, "Should the government force me to be my sister's keeper, if my sister does not choose to help herself?" Is it fair to punish Prudence with additional taxation and reward her high-living sister? Her sister replies that the past makes no difference: as they approach retirement, their incomes differ. The fact is that Prudence's income is considerably in excess of Imprudence's and Prudence is therefore better able to pay for the support of the government (and her sister).

LIFETIME INCOME AS THE BASIS OF TAXATION   The contrast between consumption and income as a basis of taxation may not be as stark as it has sometimes been portrayed. The real issue may be the appropriate time unit to use as the basis of taxation. Under a view that is growing in support, the appropriate basis of taxation should be lifetime income, not income in one year. Lifetime income is defined as the *present discounted value* of the individual's wage income.

Recall from Chapter 11 the discussion of the problem of adding up the benefits (and costs) of a project that occurred at different dates. We argued there that $1.00 in the next period was worth less than $1.00 in this period. If we receive $1.00 in this period, we could put it in the bank, and have $(1 + r)$

---

[6] And by bequests and inheritances, which we could view as special forms of consumption and income.

dollars in the next period, where r is the rate of interest. If r is 10 percent, we would have \$1.10 in the next period. Thus, we should be indifferent between receiving \$1.00 today or \$1.10 in the next period. We say that the present discounted value of \$1.10 in the next period is \$1.00. That is, we discount future receipts because they are less valuable. If an individual lives for two periods, and receives a wage of w_0 in the first period and w_1 in the second, the present discounted value of his or her income, Y^*, is

$$Y^* = w_0 + \frac{w_1}{1 + r}.$$

Of course, the present discounted value of an individual's consumption over his or her lifetime must be equal to the present discounted value of his or her income (ignoring inheritance and bequests). That is, if c_0 is the individual's consumption in the first period, and c_1 is the consumption in the second,[7]

$$Y^* = c_0 + \frac{c_1}{1 + r}.$$

It thus becomes clear that if we believe that the correct basis of taxation is the individual's lifetime income, this is equivalent to believing that the correct basis of taxation is the individual's lifetime consumption.[8]

To say that taxes should be based on lifetime income or consumption means that if two individuals have the same lifetime income or consumption, then they should pay the same (present discounted value of) tax, regardless of the pattern of that income or consumption over their lifetime. *When* they pay that tax depends on how the tax is implemented.[9]

---

[7] To confirm this, assume that the individual consumed an amount that is less than his or her wage income in the first period. This individual's savings would then be $(w_0 - c_0)$. In the next period, the individual would have his or her wage income plus savings to consume, that is,

$$c_1 = w_1 + (w_0 - c_0)(1 + r).$$

We can rearrange terms to write

$$c_1 + c_0(1 + r) = w_1 + w_0(1 + r).$$

Divide by $(1 + r)$ to obtain the desired result.

[8] There are a couple of qualifications to this analysis. First, with bequests and inheritances, the present discounted value of an individual's consumption may either exceed or be less than his or her lifetime income. Although, for most individuals, bequests and inheritances are relatively small—and thus could safely be ignored—for the very rich, they loom large. How they "should be" and are treated is discussed in Chapter 21.

This analysis also ignores uncertainty. Whereas by definition—ignoring bequests and inheritances—lifetime income equals lifetime consumption, an individual who faces larger uninsured risks, say, associated with a variable wage, is likely to be worse off than an individual with the same expected income but facing no risk; yet, under the consumption (or lifetime income) tax, they both pay the same taxes to the government. Is this fair? Is there any "fair" way to reflect such risks in the design of taxation? This issue has not received much attention so far.

[9] This is most easily seen in the case of proportional taxation. The tax can be imposed on wages, in which case it is paid as individuals earn their wage income, or on consumption, in which case it is paid as individuals consume goods.

The analysis has ignored the problems that arise if there is an imperfect capital market, such that there is no single interest rate at which individuals can borrow or lend.

Note the one, strong implication of using lifetime income as a basis of taxation: interest income should be exempt from taxation. A consumption tax is equivalent to a lifetime income tax, which in turn is equivalent to a tax on wages alone. Although many see the first two as plausible bases for taxes, they find the third hard to accept—even after they see its equivalence to the others. Why should those who earn interest income be exempt from taxation?[10]

CRITICISMS OF INCOME AS A BASIS OF TAXATION   Some have criticized the use of income as a basis for taxation, believing that neither income—lifetime or annual—nor consumption provides a fair basis of taxation. Their reasoning is illustrated by the following example. Consider Joe Smith and his twin brother, Jim, who have identical abilities and education. Joe decides to take a job as a high school teacher of economics. He teaches six hours a day and the rest of the time he spends fishing, swimming, and sailing. He is very happy. Not surprisingly, his pay is very low. Jim becomes an economic consultant. He works seventy hours a week and has no time for fishing, swimming, or sailing. Their economic opportunity sets—what they could have done—are identical (i.e., Jim and Joe have the same earning ability), yet they have made different choices. One has a high income, one a low income. Is it fair that Jim should pay far higher taxes than Joe? Joe believes that it is not economic opportunities that provide the fair basis of taxation, but the extent to which individuals have seized advantage of whatever opportunities society has offered—in short, actual income provides the appropriate basis of taxation. Jim believes that it is not actual income that should be relevant, but earning ability.

The problem is that even if one accepted Jim's argument, the government has no way of accurately assessing the individual's opportunity set. Wage rates may provide a better indicator than income, but even wages are affected by individuals' choices (e.g., how hard to work or whether to accept a high-risk job). Moreover, in many jobs, wages are hard to measure and even harder to verify. We may know how much income an individual gets paid, but it is often difficult to know how many hours he or she has worked, especially in jobs where individuals do not punch a clock.

In practice, then, governments use income or consumption as the basis of taxation, even if they are flawed measures either of ability to pay or of an individual's well-being.

---

[10] We said that the equivalence held only if there are no inheritances (note 8). One might argue that interest received on inherited capital should receive different treatment. In fact, under current U.S. law, the return on a large fraction of an individual's savings over his or her lifetime—retirement savings, as well as investments in housing—are not taxed.

THE BENEFIT APPROACH   We noted earlier that one argument for the use of consumption as the appropriate basis for taxation is that it seems fairer to tax individuals on the basis of what they take out of the economic system. Some economists have gone further and argued that individuals should contribute to the support of the government in proportion to the benefit they receive from public services. The principles of charging for public services should be analogous to those used for private services—and taxes can be viewed as simply the "charge" for the provision of public services.

In a few cases, the **benefit approach** is explicitly adopted: fees (taxes) are charged for the use of bridges and some toll roads. Financing roads with gasoline taxes can be thought of as a simple mechanism for relating benefits (road usage, as measured by gasoline consumption) to taxes.

For the most part, economists have not been attracted to the benefit approach to taxation, largely because it is impossible to identify the magnitude of the benefits received by different individuals. We all receive some benefit from defense expenditures, but how are the relative benefits to be apportioned among different individuals? For many categories of expenditures, assessment of benefits is essentially impossible. A second objection raised against benefit taxes when they are related to usage is that they are distortionary. Basing taxes on usage of a public facility (such as a bridge) may discourage its use and thus lead to an inefficient allocation of resources.

There are often equity–efficiency trade-offs involved in levying benefit taxes (when it is possible to do so). In the absence of benefit taxes, it is impossible to make those who benefit from a public facility such as a bridge bear the cost; if the bridge is financed out of general revenues, those who do not use the bridge (but contribute to it through taxes) are made worse off. It seems unfair to them that they should subsidize those who use the bridge.

ALTERNATIVE BASES OF TAXATION   The principle of vertical equity says that those who are better off or have a greater ability to pay ought to contribute more to support the government than those less well off. The principle of horizontal equity says that those who are equally well off (who have equal ability to pay) should all contribute the same amount. Our discussion of both principles has focused on the difficulties of determining whether one individual is better off than another, or of determining whether one individual has a greater ability to pay than another. How should we adjust for the myriad differences in circumstances facing different individuals?

In each of the three following examples, present tax laws make some adjustments for differences in circumstances; there is, however, some controversy about whether the adjustments are appropriate.

The first example has to do with health. Clearly, an individual who is sick and has an income of $10,000 is different from an individual who is well and has the same income. Most of us would say that the individual who is sick is worse off (other things being equal) than the one who is well. Being sick or well, though, is not always readily observable. Accordingly, it is difficult for the tax code to make adjustments for health status.

There is a surrogate, though: medical expenditures. Those who spend more on hospital bills are, on average, worse off than those who have no hospital bills. The current tax law does allow for the deduction of medical expenses in excess of 10 percent of the individual's income.

The second example has to do with marriage. Individuals who are married differ from those who are not. Surveys by sociologists indicate that married men, for instance, are happier than unmarried men. Whether or not much credence should be placed in such evidence, the fact is that married men do live longer and are, on average, in better health. This would suggest that a married man with a given income is better off than an unmarried man with the same income. Does the principle of vertical equity imply that the married man should pay higher taxes? The present tax structure does discriminate against married individuals when the husband and the wife have similar incomes (though probably not for the reasons just given), whereas marriage may reduce the taxes of a man and woman who have very different incomes.

The third example has to do with the tax treatment of children. Consider two married couples with identical incomes. Both couples would like to have two children. One of the couples is infertile, while the other is blessed with two children. Clearly, the couple with the two children is better off than the infertile couple. Even taking account of the extra costs of raising children, the fertile couple would not change places with the other. The principle that those who are better off should pay more taxes would suggest that this couple should pay more taxes;

FAIRNESS

Horizontal equity: Individuals who are identical (or in essentially similar economic circumstances) should be treated the same, and pay the same taxes.

   Key question: What differences are relevant?

Vertical equity: Individuals who have greater ability to pay or who are better off or receive greater benefits from government services should pay more taxes.

   Key questions:

   What should be the basis of taxation?

   How is "ability to pay" or "benefits received" or "economic welfare" to be measured?

   How much more should those considered better off pay?

- Income is most often used as a basis of taxation, but it is an indirect and imperfect measure of both ability to pay and economic well-being.
- Consumption may be "fairer"—it measures what one takes out of society rather than what one contributes.
- Lifetime consumption is equivalent to lifetime income. Lifetime income is a fairer basis than annual income (a better measure of overall ability to pay, or welfare).
- Lifetime consumption/income is a flawed measure of ability to pay: it unfairly disadvantages individuals who choose to work hard rather than enjoy leisure, and it is not a real measure of one's opportunity set.
- Benefit taxation is hindered by difficulties of measuring benefits, especially for pure public goods.
- What are fair adjustments to income as the basis of taxation, taking into account differences in health, marital status, children?

CORRUPTION-RESISTANT TAX SYSTEMS

We have described several of the key criteria for evaluating alternative tax systems. In many developing countries trying to create an effective tax system, there is something else that is of concern: corruption.

This takes many forms. Government officials can be bribed, for instance, to categorize an imported good as belonging to a lower-taxed category, or to overlook certain income. Tax authorities can use their power to extort money from taxpayers, saying that unless they get a payoff they will claim that taxpayers have a higher income (perhaps even more than they really do). In the United States, corruption typically does not take the form of passing money in plain brown envelopes. It occurs in the legislative process, as lawmakers receive large campaign contributions in return for special legislation—SUVs may be treated as trucks, not cars; racetracks may get special benefits.

In recent years, economists have asked: Are there some tax systems that are more corruption resistant? Are there some that are more corruption prone?

A tax system that makes multiple distinctions—for instance, with different categories of goods (cars versus trucks) being treated differently—is more prone to corruption. A tax system in which the things being taxed are more easily observable, and verifiable by a third party, are more corruption resistant. A tax per car produced is more corruption resistant than a tax on the value of the car, since it is easy to count the number of cars produced but hard to check on the value of the car. Because most cars are purchased with a loan, the car company could sell the car at a cheaper price, but make up the difference with higher profits on the loan. Tax authorities must make judgments on whether there is cheating going on, and that opens up the door for corruption, which can take the form of campaign contributions to get the rules written in the right way. Tax rates are also important: a 50 percent tax on income or a 25 percent tax on sales increases the incentive to cheat—the gains might be worth the risk.

As these examples illustrate, there are often trade-offs. Complex, opaque tax systems that give tax officials considerable discretion and have high tax rates are incubators of corruption. Simple, transparent, rules-based tax systems with low tax rates are more corruption resistant, but may fall short on other objectives listed in the text.

Moreover, how tax laws are enforced can also affect the extent of corruption. Frequent rotation of tax agents from one task to another and one location to another makes it more difficult to form close links between particular taxpayers and particular government officials, so this has proven to be an effective tool for reducing corruption. The use of computer systems and related information technology makes it easier to monitor the behavior of taxpayers and tax officials and spot malfeasance. Requiring taxpayers to use the banking system to make their tax payments is another way to make a tax system more corruption resistant, as it prevents tax officials from handling cash payments from taxpayers. Adequate compensation for tax officials is a third antidote for corruption—it can help to limit systemic petty graft because tax officials are not compelled to supplement their income to support their families. However, the most powerful tax administration medicine for making a tax system corruption resistant is to offer taxpayers a balanced combination of a simple tax system well administered, so that it is easy to do the right thing, and a credible threat of sanctions if they do not comply with tax laws.

in fact, the tax law results in the couple with children paying lower taxes. The law seemingly looks not at their "well-being" but at their ability to pay, and recognizes that, having had the children, they face additional expenses that make them less able to pay taxes.

The analysis so far has shown that although the principles of vertical and horizontal equity seem, at first, to provide "reasonable" bases for designing a fair tax system, they are, in fact, of only limited help. The difficult questions are left unanswered: How can we tell which of two individuals is better off or which has a greater ability to pay? What do we mean by equality of treatment? Furthermore, the principle of vertical equity does not tell us how much more someone who is better off should contribute to the support of the government; all it tells us is that he or she should pay more.

Because of these difficulties, economists have looked for other principles by which to choose among alternative tax systems.

GENERAL FRAMEWORK FOR CHOOSING AMONG TAX SYSTEMS

The concerns of equity and efficiency that we have raised about different bases of taxation may be integrated into a general framework—essentially an application of standard welfare economics. We first look at efficiency (taking into account both the distortions and the resources used to implement a tax, the administrative and compliance costs). We identify **Pareto efficient tax systems**—tax structures such that, given the tools and information available to the government, no one can be made better off without making someone else worse off. Then we choose among the possible Pareto efficient tax structures using a social welfare function, which summarizes society's attitudes toward the welfare of different individuals.

The advantage of this approach is that it separates efficiency considerations from value judgments. Almost all would agree that if a tax structure could be found in which everyone was better off (or some better off and no one else worse off), it should be adopted. On the other hand, often none of the alternative tax systems available dominates the others. In one tax system the poor may be better off, the rich worse off; but are the gains to the poor sufficiently large to justify the losses to the rich? The answer depends on value judgments, over which reasonable people may differ.

Recall from Chapter 7 that economists have made use of two special social welfare functions: the utilitarian (social welfare equals the sum of

all individuals' utilities) and the Rawlsian (social welfare equals the utility of the worst-off individual). Either social welfare function makes it possible to say not only by how much taxes should increase with income but also, for instance, whether and under what circumstances a deduction for medical expenses should be allowed.[11] We now explore briefly what each of these two social welfare functions implies for tax design.

UTILITARIANISM

Traditionally, utilitarianism was thought to provide a rationale for progressive taxation, the taxation of rich individuals at higher rates than poor individuals. Under utilitarianism, taxes should be such that the marginal utility of income—the loss in utility from taking a dollar away from an individual—should be the same for all individuals.[12] If the marginal utility of income of Jim exceeds that of Joe, reducing Jim's tax by a dollar and increasing Joe's by a dollar increases total utility (social welfare), as the gain in utility to Jim exceeds the loss to Joe. Because taking a dollar away from a rich person causes him or her less loss of welfare than taking a dollar away from a poor person, utilitarianism seemed to provide a basis for progressive taxation.

However, this argument fails to take into account that individuals' income depends on their work (effort), and that raising taxes on those earning higher incomes may lead to a reduction in their work (effort). It is thus possible that raising the tax rate actually reduces the government's tax revenue, or that the marginal utility loss to the individual per dollar raised by the government may be very large. The earlier argument assumed, in other words, that income would not be affected by the imposition of taxes; it is now widely assumed that it generally will be. When it is, utilitarianism requires that we compare the loss in utility from an increase in a tax with the gain in revenue. We require that

$$\frac{\text{change in utility}}{\text{change in revenue}}$$

be the same for all individuals. If some group of individuals has a very elastic labor supply (that is, as tax rates are increased they greatly reduce their labor supply), an increase in the income tax rate on that group will yield relatively little revenue, so they should not be taxed heavily.

---

[11] To make utilitarianism (or Rawlsianism) operational, one must make additional assumptions, as we noted in Chapter 7. It is conventionally assumed that all individuals have the same utility function—at each level of income, all individuals benefit equally from an extra dollar—and that they exhibit diminishing returns—an extra dollar is worth less at progressively higher levels of income.

[12] Thus, under utilitarianism, taxes are not *directly* related to the benefits one receives from spending financed by the tax, or to the level of economic welfare, but to the *marginal* benefit of a dollar of additional income.

Utilitarianism was also once thought to provide a basis for the principle of horizontal equity. If everyone had the same utility function, individuals with the same income should be taxed the same. Assume that one individual faced a higher tax than another with the same income. Because of diminishing marginal utility, this individual's marginal utility of income would be higher than the other's. Raising the tax on the individual with the *low* tax rate would cause him or her less loss in utility than the gain in utility from lowering the tax on the individual with the high tax rate. Again, this argument would be correct if income were unaffected. It is affected, though, so the argument may no longer be valid.[13]

The argument that utilitarianism may imply horizontal *inequity* is perhaps best made by the story of the shipwrecked crew. The crew has enough food for all but one of its members to survive. Equality would thus imply that all individuals must die—clearly a worse situation, from a utilitarian point of view, than one in which only one dies.

RAWLSIAN SOCIAL WELFARE FUNCTION

Some economists and philosophers believe that the utilitarian approach is not sufficiently egalitarian—that it pays insufficient attention to inequality. In Chapter 7, we discussed the view of John Rawls that society should be concerned only with the welfare of the worst-off individual, and it ought to design the tax system (and other social policies) so as to maximize that individual's welfare. The Rawlsian social welfare function—maximizing the welfare of the worst-off individual—has some simple and direct implications for tax policy: increase the tax rates on all individuals (other than the worst-off individual) to the point at which the tax revenues from them are maximized. This does not necessarily imply that very rich individuals should be taxed at 80 or 90 percent of their income, or even that marginal tax rates should always increase with income. It could turn out that those with very high incomes have labor supplies that are more sensitive to tax rates than those of middle-income individuals. (Empirically, this appears not to be the case, and a Rawlsian social welfare fuction does suggest a very high tax rate at the top.)

There are those who argue that not even the Rawlsian criterion is sufficiently egalitarian. Consider a policy change that makes the worst-off individual just slightly better off, but makes the richest 5 percent of the population much better off. Under Rawls, this is a desirable change—Rawls pays

---

[13] It can be shown that under plausible conditions, utilitarianism requires that with distortionary taxes, individuals who appear to be essentially identical be treated differently. A formal exposition of the argument is presented in J. E. Stiglitz, "Utilitarianism and Horizontal Equity: The Case for Random Taxation," *Journal of Public Economics* 21 (1982): 257–294. See also D. L. Brito et al., "Randomization in Optimal Tax Schedules," *Journal of Public Economics* 56, no. 2 (February 1995): 189–223.

attention only to the worst-off individual. However, some would argue that inequality itself is a social evil or gives rise to social evils. Differences in levels of wealth may give rise, for instance, to social tensions. Inequality of goods leads, in many political situations, to inequality in political power, and this may be used, eventually, to the advantage of the well-off at the expense of the poor.

LIMITATIONS OF THE SOCIAL WELFARE FUNCTION APPROACH Even though economists have found the social welfare function extremely useful in thinking about the trade-offs in designing tax structures, the fundamental problems of the ability to pay and related principles remain (though often swept underneath the surface). If everyone were the same except for some attribute, such as their wage or inheritance, then it would be plausible to treat them the

> ## GENERAL FRAMEWORK FOR THINKING ABOUT TAXATION
>
> - *Pareto efficient taxation:* tax structures such that, given the revenue raised, no one can be made better off without making someone else worse off. Choice among Pareto efficient tax structures depends on values, reflected in the *social welfare function*.
> - *Utilitarian social welfare function:* chooses the Pareto efficient tax structure that maximizes the sum of utilities of individuals; marginal loss of utility per dollar of revenue raised must be the same for all individuals.
> - *Rawlsian social welfare function:* chooses the Pareto efficient tax structure that maximizes the utility of the worst-off individual.

same; we could, under utilitarianism, simply add up their "utilities." However, individuals *do* differ; some have a need for immediate gratification, whereas others get more enjoyment from taking a longer-run perspective. The social welfare function approach may tell us that in choosing a tax structure, we should equate the change in utility relative to the change in revenue for all individuals. However, it does not tell us how we can compare the utility of Spendthrift with that of his brother Scrooge,[14] and to actually design a tax structure requires making such judgments.

WHAT ECONOMISTS CAN CONTRIBUTE TO DISCUSSIONS OF FAIRNESS Although economists (or philosophers) have not resolved the basic issues involved in the choice of bases for judging fairness, still much can be said. It is important, for instance, to be able to describe the full consequences of any tax, and these are seldom simply described by the amounts of tax each person pays directly. We can attempt to describe how various groups in the population are affected by different tax programs. In all tax systems, certain groups seem to pay less than their fair share, given any reasonable concept of fairness. We then need to ask: Why are they treated differently? It may be that to treat them fairly would necessitate introducing other, even worse, inequities into the tax code. Tax systems must be based on *observable* variables, such as income or expenditures. As we noted earlier, many of the concepts, such as welfare, involved in our more general philosophical discussions are not directly measurable.

---

[14] Recall the discussion in Chapter 7 on the problems of interpersonal utility comparisons.

Even income is not as well defined as might seem to be the case at first. Thus, many apparent inequities in our tax system are simply consequences of the inherent difficulty of translating what seem like well-defined concepts into the precise language required by any tax law.

In other cases, by considering carefully how different provisions of the tax code and changes in those provisions affect different groups, we can obtain some insight into why one group may claim that a set of provisions is unfair while another group claims that to change them is unfair. We can attempt to distinguish cases in which the term "fairness" is used simply to cover up a group's pursuit of self-interest from cases in which some reasonable ethical or philosophical position underlies individuals' claims.

REVIEW AND PRACTICE

SUMMARY

1. The five attributes that a good tax system should have are:

    Economic efficiency

    Administrative simplicity

    Flexibility

    Political responsiveness

    Fairness

    The tax system should also be "corruption resistant."

2. Tax distortions arise when behavior is altered in an attempt to avoid or reduce taxes. With the exception of lump-sum taxes, all taxes create such distortions. Taxes affect decisions in all markets, including labor supply and savings decisions, and have impacts on financial and organization structures. Taxes on the future returns of an asset are typically capitalized in the value of the asset at the time the taxes are announced.

3. The two major aspects of fairness are horizontal equity and vertical equity.

4. A central question in applying the principle of horizontal equity, which requires that identical individuals pay identical taxes, is specifying the criteria for grouping individuals as identical (for purposes of taxation).

5. The principle of vertical equity says that those who are more able to pay, or who have a higher welfare, should pay higher taxes. Income is the most commonly used measure of either ability to pay or economic welfare, but it is a flawed measure. Some argue that consumption provides a better basis. Taxing lifetime consumption is equivalent to taxing lifetime income, and both are widely viewed as superior to basing taxes on annual income. Hard questions are posed by what adjustments should be taken into account—for instance, for differences in health, marital status, or children.

6. Pareto efficient tax structures are those such that, given the tools and information available to the government, no one can be made better off without someone else being made worse off.

7. The utilitarian approach argues that the tax system chosen should maximize the sum of utilities. The Rawlsian approach argues that the tax system chosen should maximize the welfare of the worst-off individual.

KEY CONCEPTS

Announcement effects

Benefit approach

Corrective taxes

Corruption resistant

Distortionary taxes

General equilibrium effects

Horizontal equity

Lump-sum taxes

Monetized taxes

Nondistortionary taxes

Pareto efficient tax systems

Transparent Taxes

Vertical equity

QUESTIONS AND PROBLEMS

1. Discuss how your views concerning the tax treatment of children might be affected by whether you (a) lived in a highly congested country or in an underpopulated country; (b) viewed children as a consumption good (for their parents), like other consumption goods. Discuss both efficiency and equity considerations.

2. With a progressive tax structure, it makes a great deal of difference whether husbands' and wives' incomes are added together and taxed, or taxed separately. Discuss some of the equity and efficiency considerations that bear on the tax treatment of the family.

3. Does utilitarianism necessarily imply that tax structures should be progressive?

4. Consider an individual who has lost a leg but, with a new artificial leg, has the same earning power as before. How should his or her taxes differ from a similar individual who has not lost a leg: (a) under utilitarianism; (b) under a Rawlsian social welfare function; or (c) if you believed that ability to pay provided the appropriate basis for taxation?

5. The government has passed a number of pieces of legislation aimed at ensuring that firms do not take advantage of consumers' limited information. What might be meant by a "truth in taxation" law? What might be the advantages of and problems with such a law?

6. "Because the needs, other than medical, of the aged are typically not as great as those of younger individuals who have children to support, if the government provides free medical care to the aged, it should simultaneously subject the aged to higher income tax rates." Discuss the equity and efficiency consequences of doing this (consider alternative views of equity).

7. Suppose that the labor supply of married women is very sensitive to the after-tax wage (that is, it is very elastic), whereas the labor supply of men is not. The government proposes to reduce the tax on income earned by married women by 5 percent, and raise the tax on earnings of married men by 15 percent. How would this tax change affect total tax revenues? How would it affect the distribution of income?

8. Consider a state debating how to finance emergency road and bridge improvements. Among the possibilities are increased fees on drivers' licenses, a personal property tax on motor vehicles, a tax on automobile parts (including tires), and higher taxes on cigarettes and liquor. Which of these taxes are benefit taxes, which are corrective taxes, which are both? Which of these taxes is least distortionary?

9. Issues of efficiency and fairness often get intertwined in complex ways. Use the perspectives on the principles of taxation provided in the text to discuss the appropriateness of:

   a. Taxes on gasoline.

   b. Subsidies to public transportation.

   c. Polluter-pays principle, in which those who cause pollution have to pay for its cleanup (e.g., dry-cleaning establishments having to clean up the toxic wastes resulting from the chemicals they use).

18

TAX INCIDENCE

When Congress or a state legislature enacts a new tax, the debate usually includes some opinions about who should pay for running the government or for the particular program being supported by the tax. For example, when Congress adopted the Social Security tax to pay for the Social Security system, it levied half the tax on the employer and half on the employee. It thought that both parties should share in the costs of the Social Security system.

Economic reality, however—for better or worse—does not always follow the laws passed by legislatures. Thus, economists distinguish between those who bear the **burden** of a tax and those on whom a tax is imposed or levied. The *tax burden* is the true economic weight of a tax. It is the difference between the individual's real income before and after the tax has been imposed, taking full account of how wages and prices may have adjusted. Economists use a more neutral word to describe the effects of taxation—they ask, what is the **incidence** of the tax? Who actually pays, in the sense that real income is lowered? This chapter studies the incidence of various taxes.

The actual incidence of the tax may differ markedly from the intended incidence. Consider two taxes that are imposed on firms: the employer-paid portion of the Social Security tax and the corporation income tax.

As a result of either tax, wages might fall or prices might rise. If wages fall, we say that the tax **shifted backward** (to a factor of production, labor). If wages fall by the full amount of the tax, we say that they have been *fully shifted*; if wages fall by less than the amount of the tax, we say they have been *partially shifted*. If prices rise, we say that the tax **shifted forward** (to consumers). Most economists believe that most of the employer-paid portion of the Social Security tax is shifted backward and that the effect of the tax officially levied on employers is essentially the same as that levied on workers. Thus, although the government has levied only half of the tax on employees, the employees bear the full (or almost the full) burden of the tax in the form of lower wages.

There is considerable controversy over the incidence of the corporation income tax. Although one reason the tax is popular is that, ostensibly, firms and their shareholders pay the tax, most economists believe that a substantial portion of the tax is shifted. If firms raise their prices as a result of the tax, the tax is borne by consumers. If, as a result of the tax, demand for labor falls and wages fall, the tax is partially borne by workers, not investors. If the tax makes investing in the corporate sector less attractive, capital will move out of the sector, driving down the return to capital in the unincorporated sector. Thus, part of the burden of the corporate tax is on capital as a whole, not just capital in the corporate sector.

The study of tax incidence is one of the most important and difficult topics in the economics of the public sector. In the preceding chapter, we saw that one of the principles of a desirable tax system is that it should be *fair*. Fairness, however, depends not on whom the tax is imposed, but on who actually pays the tax—on the incidence of the tax. If it were decided, for instance, that fairness dictated that owners of capital should pay higher taxes, but the tax was levied in such a way that the owners of capital could shift the tax onto consumers or workers, then the tax would not have achieved its goal. Economics, not Congress, often determines who actually bears the burden of a tax, although in designing the tax Congress can often affect the outcome: two taxes, both imposed on corporations but differently designed, can have markedly different consequences.

Just as two taxes that look similar—in that both are imposed on, say, corporations—can have markedly different effects, two taxes that look different, in that they are imposed in quite different ways, can have identical effects. Such taxes are said to be **equivalent**.

In Chapter 17 we saw that another principle of a desirable tax system, besides fairness, is transparency. This has two implications. First, it is preferable to impose taxes whose incidence is clear. Second, because most individuals do not understand incidence analysis, it is preferable to impose taxes in a manner that makes the apparent incidence of a tax correspond to the actual incidence. Thus, imposing half of the Social Security tax on

the employer contributes to the lack of transparency, as it makes workers believe that the employer actually bears half the burden of the tax.

The incidence of a tax depends on a number of factors—most important, on whether the economy is competitive; and if it is competitive, on the shape of the demand and supply curves. This chapter is divided into five sections. The first and second analyze incidence in perfectly competitive markets and in markets in which there is imperfect or no competition. The third analyzes some important equivalent tax structures. In the fourth section, some other important determinants of incidence are discussed, examining a tax on capital in the corporate sector. In the final section, we discuss briefly the implications of our analysis for the overall incidence of taxation in the United States.

Although this chapter focuses on the incidence of taxes, it should be clear that precisely the same issues arise in discussing subsidies and other benefits, such as those discussed in earlier chapters on government expenditures. If corn is subsidized, for instance, it may not be corn growers who really benefit: if the price of corn falls, the benefit is shifted forward to consumers; if the price of land on which corn is grown increases, the benefit is shifted backward to the owners of land. The principles elucidated here apply equally to the analysis of government benefit programs.

TAX INCIDENCE IN COMPETITIVE MARKETS

In this section, we will show that it makes no difference whether a tax on a commodity is legally imposed on the commodity's consumers or on its producers—it makes no difference whether producers of beer or its consumers "pay" the tax. What does make a difference is the shape of the demand and supply curves.

EFFECT OF TAX AT THE LEVEL OF A FIRM

Consider a commodity tax imposed at a fixed rate per unit of the good (e.g., so many cents per can of beer) that the firm must pay. Figure 18.1 illustrates the effect of the tax on the firm's production decision. In competitive markets, firms produce at the level at which price equals marginal costs.[1] If the firm has to pay the tax, then its effective cost of production has been increased, *by the amount of the tax*. Accordingly, the amount it is willing to supply at the price p_0 is reduced.

---

[1] At lower levels of output, increasing output increases revenues by more than the increased costs, so profits increase. The converse occurs at higher levels of output.
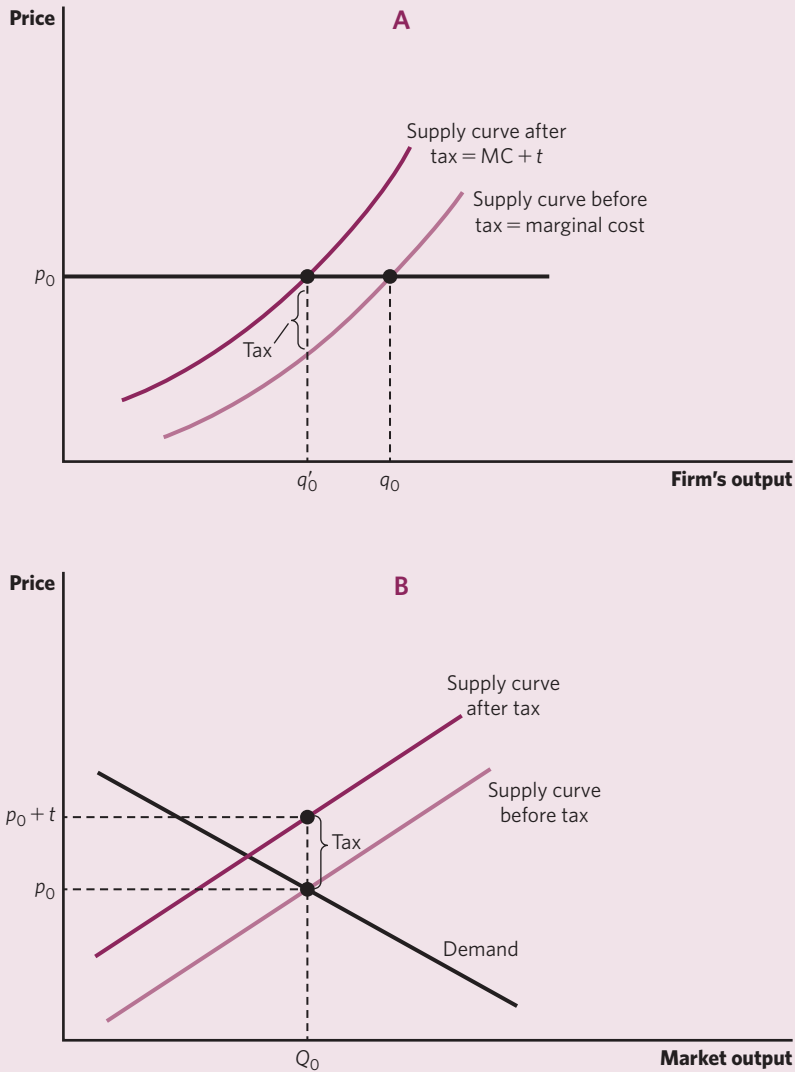
THE EFFECT OF A COMMODITY TAX ON SUPPLY

(A) Shows the effect of a commodity tax on the quantity supplied by a firm. At any price, p_0, the firm will supply a lower quantity. The tax can be thought of as increasing the marginal cost of production. Output supplied is reduced from q_0 to q_0'. (B) Shows the effect of a commodity tax on the market supply curve and equilibrium. At each price the market is willing to supply less (the supply curve shifts to the left), or, equivalently, the price required to elicit a given supply out of the market is higher, by an amount exactly equal to the tax.

The firm's supply curve gives the amount the firm is willing to supply at each price. Its supply curve is shifted, as illustrated in Figure 18.1A. This is, of course, true for *every* firm. The *market* supply curve gives the total amount that *all* firms are willing to supply at each price. It is simply the "sum" of the supply curves of each firm. Equivalently, we can think of the market supply curve as telling us what the market price must be for firms to be willing to produce a given level of output. The market supply curve, like the individual firm supply curves, is shifted, as illustrated in Figure 18.1B. The amount of the shift is easy to ascertain. If t is the tax rate, then the net

amount received by the firm when the price is $p_0 + t$ after the tax is the same as it would have received when the price was just p_0 before the tax; the quantity that each firm is thus willing to supply at the price $p_0 + t$ after the tax is the same as it would have been willing to supply at the price p_0 before the tax. In effect, the supply curve is shifted up by the amount of the tax.

IMPACT ON MARKET EQUILIBRIUM

We can now easily see the impact of the tax on prices and output. Figure 18.2 shows the equilibrium before taxes, at the intersection of the demand and supply curve, at which Q_0 bottles of beer are produced in equilibrium, at a price of $1 each.

Assume that the tax on each producer is 10 cents per bottle of beer. The supply curve shifts up by that amount, and the price rises. Although the tax was nominally imposed on producers, consumers are forced to pay a part of the increased cost, through higher prices. Notice, however, that in this example, the price rises by less than 10 cents, to $1.05. Producers cannot shift the entire cost of the tax to consumers because as the price rises, the quantity demanded falls.

Each firm now receives the higher price of $1.05, and faces additional costs of 10 cents per bottle. The firms in Figure 18.2 produce less than before the tax, but more than they would have if consumers did not bear part of the additional cost.

FIGURE 18.2

EFFECT OF TAX ON PRICES AND QUANTITIES
The tax shifts the supply curve up by the amount of tax. This lowers the quantity consumed and raises the price paid by consumers.
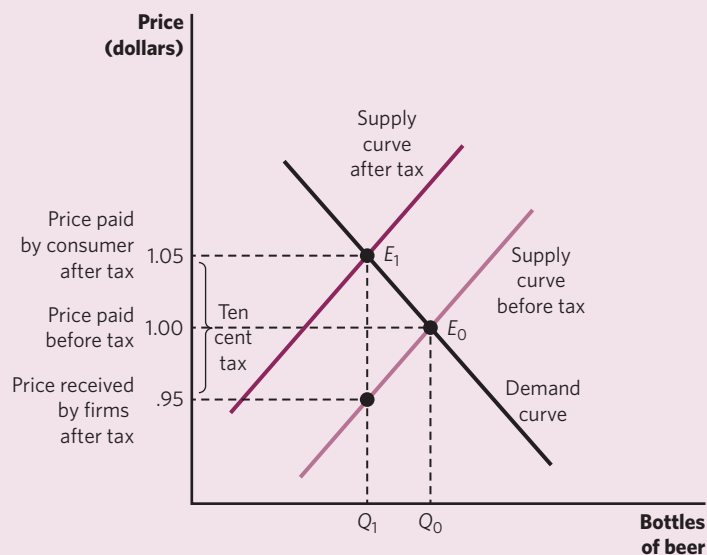
DOES IT MATTER WHETHER THE TAX IS LEVIED ON CONSUMERS OR ON PRODUCERS?

Consider now what would happen if Congress passed a beer tax, but this time said that consumers would have to pay the tax. For each bottle of beer purchased, consumers would have to pay a 10-cent tax. What consumers care about, of course, is not who receives the money they pay, but simply the total cost of the beer—just as what producers care about is how much they receive. Return to Figure 18.2, which showed the effect of a 10-cent tax imposed on producers. At the new equilibrium output Q_1, producers receive $0.95, after tax, and consumers pay $1.05. In that situation, the producers mail the government a check for 10 cents for every bottle of beer. However, nothing would change if consumers, or the retailers from whom they buy beer, had to send a check in for the same amount. Producers would then pay no *direct* attention to the tax. They would sell the beer to consumers for 95 cents, and at that price they would be willing to produce Q_1. Consumers would pay the producers 95 cents and pay the government 10 cents, for a total price of $1.05. At the total price of $1.05, they are willing to purchase Q_1, so at Q_1—a consumer price of $1.05 and a producer price of $0.95—demand equals supply.

This situation is depicted diagrammatically in Figure 18.3. If we now interpret the price on the vertical axis to be the price received by the
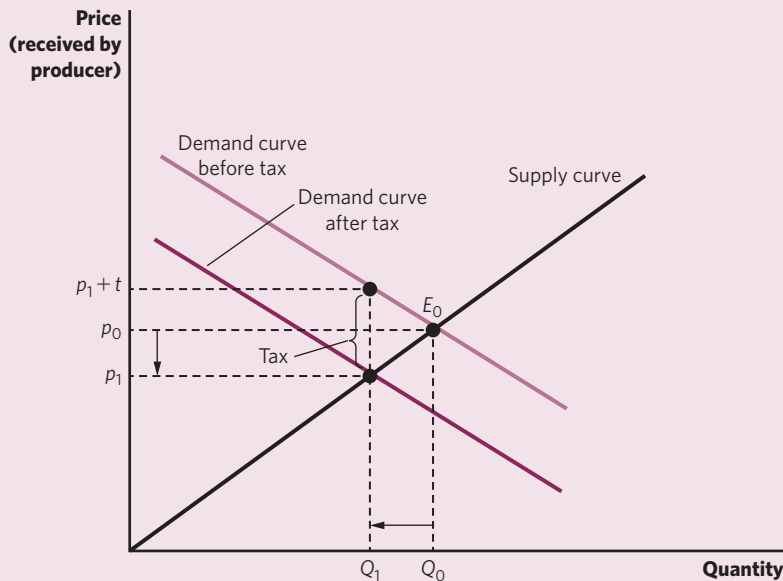
FIGURE 18.3

ALTERNATE VIEWS OF A TAX

The effects of a tax can be viewed as either a downward shift in the demand curve or an upward shift in the supply curve (compare with Figure 18.2).

THE INCIDENCE OF GOVERNMENT BENEFITS

The framework we have developed for analyzing the incidence of taxes can be used to analyze the incidence of a government program or subsidy. Consider a subsidy for beef. For simplicity, assume that the government subsidizes beef at $1 a pound. In the short run, the supply curve is relatively inelastic, as depicted in the top figure on the right. That means there is a small quantity response, but a large price response: in the short run much of the benefit does go to farmers.

In the long run, however, as entry occurs and producers can expand their facilities, the supply curve for beef becomes relatively flat; there is a large supply of acreage that can be used for pasture, and even though it takes time to breed cattle, they can be bred, and the costs of breeding and feeding are roughly fixed. The bottom figure on the right for the long-run beef market shows a horizontal supply curve combined with a downward-sloping demand curve, and the before-subsidy equilibrium at Q_0. The subsidy can be thought of as shifting the supply curve as depicted. The new equilibrium entails a larger quantity, but the price received by farmers remains unchanged. In the long run, all the benefit of the subsidy is received by meat consumers, none by farmers.

Beef Market in the Short Run

Beef Market in the Long Run
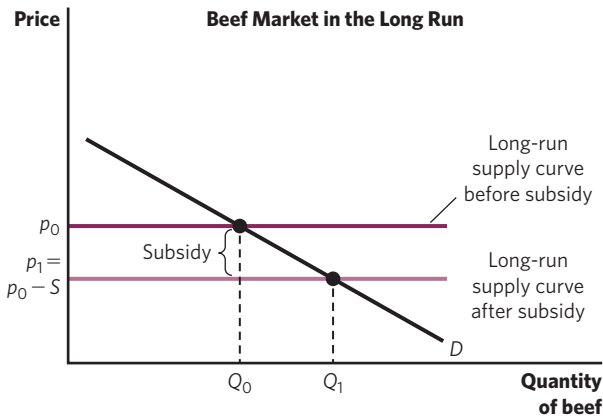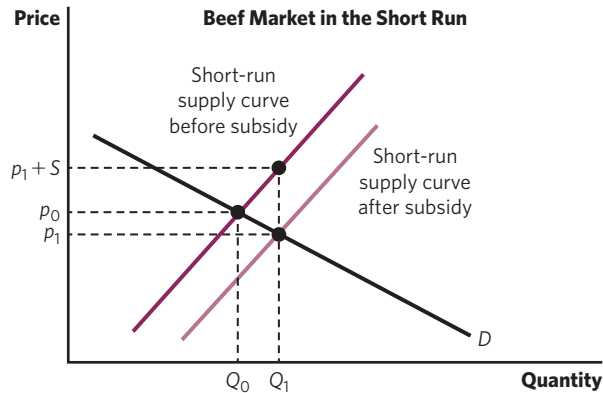
producer (rather than the price paid by the consumer), the tax on consumers can be represented by a downward shift in the demand curve, by the amount of the tax. That is, if the producer receives p_1, the consumer must pay $p_1 + t$, and the level of demand is Q_1, just as it would be if, in the before-tax situation, producers had charged $p_1 + t$. It should be apparent

that it makes no difference whether Congress imposes the tax on the producers of beer or on the consumers of beer.

AD VALOREM VERSUS SPECIFIC TAXES

Not only does it make no difference on whom the tax is levied, but it also makes no difference whether the tax is levied as a given percentage of the price or as a fixed amount per unit output. The former is called an **ad valorem tax**, the latter, a **specific tax**.

The ad valorem tax can be thought of as shifting down the demand curve, with the amount by which it is shifted down depending on the price, as illustrated in Figure 18.4. At a zero price, at which the demand curve intersects the horizontal axis, there is no tax. The manufacturer receives a fixed percentage of the price paid by the consumer, say, 95 percent, if the ad valorem tax rate is 5 percent. E_1 is the after-tax equilibrium, at the intersection of the after-tax demand curve D_1D_1 and the supply curve. In the figure, the after-tax demand curve is also drawn for the case of a specific tax *that is of the same magnitude at the equilibrium E_1*. With the tax at the same level at the equilibrium, the demand curve is shifted down by the same amount at that level of output, and thus the equilibrium output, tax revenues, prices paid by consumers, and prices received by manufacturers are all the same.

In practice the two taxes often differ, because tax authorities cannot adjust appropriately for differences in qualities of goods. When the government levies a specific tax—say, so many cents per pack of cigarettes—the tax is the same regardless of the quality of the product. Thus, the tax is a higher
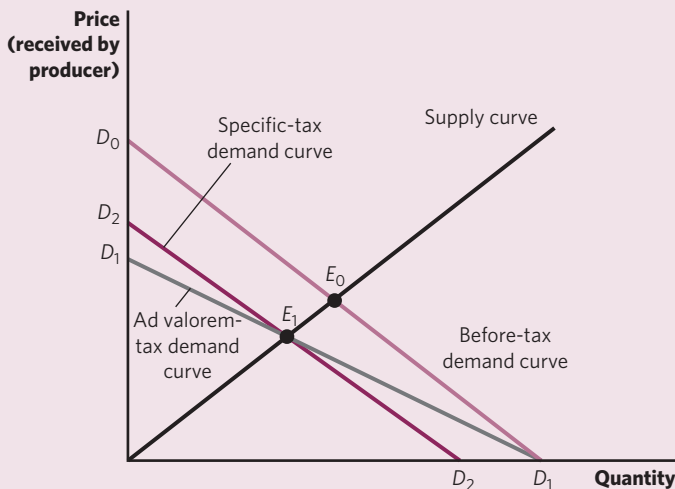
FIGURE 18.4

AD VALOREM AND SPECIFIC COMMODITY TAXES

In competitive markets, an ad valorem tax (a tax that is a fixed percentage of the price) and a specific tax (a tax that is a fixed amount per unit purchased) that raise the same revenue have the same effect on output.

TAX INCIDENCE

- The incidence of a tax describes who actually bears the tax. It does not depend on who writes the check to the government.
- It makes no difference whether a commodity tax is levied on producers or consumers.
- It makes no difference whether the Social Security tax (payroll tax) is paid half by the employer and half by the employee, or entirely paid by one or the other.
- In a competitive market, the incidences of an ad valorem tax and an equivalent specific tax are identical.

INCIDENCE IN COMPETITIVE MARKETS

- In competitive markets, incidence depends on the elasticity of demand and supply.
- A commodity tax is not borne at all by consumers if the demand curve is perfectly elastic, or by producers if the supply curve is perfectly elastic. It is borne completely by consumers if the demand curve is perfectly inelastic, or by producers if the supply curve is perfectly inelastic.

percentage of the price for low-quality goods than it is for higher-quality goods. In effect, the specific tax discriminates against lower-quality goods. Although, in principle, the government could adjust the specific tax rate to offset this bias, in fact it seldom does so.

On the other hand, it is often easier to monitor the quantity of a good sold than to monitor its price, particularly when firms sell more than one commodity. If these commodities are taxed at different ad valorem rates, there is an incentive to strike deals in which the higher-taxed commodity is underpriced on invoices, and the tax administrator may not be able to detect this. This kind of administrative problem has been the principal determinant of the form of taxation.

THE EFFECT OF ELASTICITY

The amount by which price rises—the extent to which consumers bear a tax—depends on the shape of the demand and supply curves, not on whom the tax is levied. In two limiting cases, the price rises by the full 10 cents, so the entire burden is borne by consumers. This occurs when the supply curve is perfectly horizontal, as in Figure 18.5A, or when the demand curve is perfectly vertical (individuals insist on consuming a fixed amount of beer, regardless of price), as in Figure 18.5B.

In two cases, the price paid by consumers does not rise at all; that is, the tax is borne entirely by producers, as shown in Figure 18.6. This occurs when the supply curve is perfectly vertical—the amount supplied does not depend at all on price— or when the demand curve is perfectly horizontal.

More generally, the steeper the demand curve or the flatter the supply curve, the more the tax will be borne by consumers; the flatter the demand curve or the steeper the supply curve, the more the tax will be borne by producers. We measure the steepness of a demand curve by the **elasticity of demand**; the elasticity of demand gives the percentage change in the quantity of the good consumed due to a

ELASTICITY OF SUPPLY AND DEMAND: TAX BORNE BY CONSUMERS
(A) With a perfectly elastic supply curve (horizontal supply curve), the price rises by the full amount of the tax; the entire burden of the tax is on consumers. (B) With a perfectly inelastic demand curve, the price rises by the full amount of the tax; the entire burden of the tax is on the consumers.

percentage change in its price. We thus say that the horizontal demand curve, on which a small reduction in the price results in an enormous increase in demand, is infinitely elastic; and the vertical demand curve, on which demand does not change at all with a reduction in price, has zero elasticity.

Similarly, we measure the steepness of a supply curve by the **elasticity of supply**; the elasticity of supply gives the percentage change in the quantity of the good supplied due to a percentage change in its price. We thus say that a vertical supply curve, on which the supply does not change at all with a change in price, has zero elasticity, whereas a horizontal supply curve has infinite elasticity.

FIGURE 18.6

ELASTICITY OF SUPPLY AND DEMAND: TAX BORNE BY PRODUCERS
(A) With a perfectly inelastic supply curve, the price does not rise at all; the full burden of the tax is on producers. (B) With a perfectly elastic demand curve (horizontal demand curve), the price does not rise at all; the entire burden of the tax is on producers.
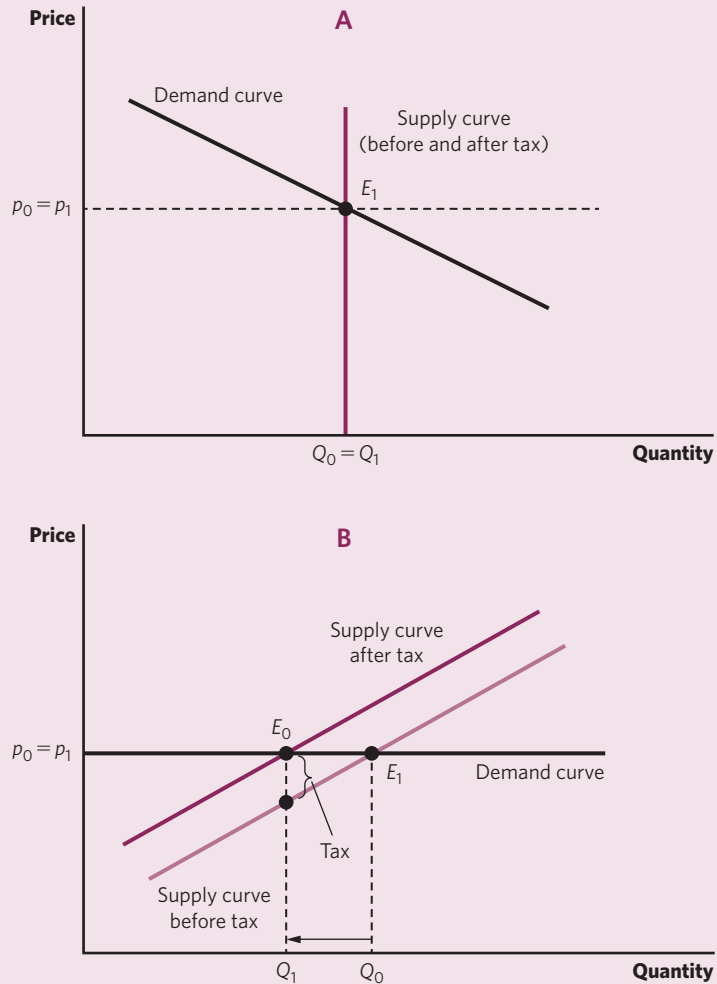
The more elastic the demand curve and the less elastic the supply curve, the more the tax is borne by producers; the less elastic the demand curve and the more elastic the supply curve, the more the tax will be borne by consumers.

TAXATION OF FACTORS

The basic principles we have just derived apply to all taxes in competitive markets, including taxes on factors of production.

THE PHILADELPHIA WAGE TAX

Many cities, including Philadelphia, levy a wage tax. A careful look at the incidence of the tax suggests that the burden of the tax is largely on landowners in Philadelphia. The supply curves for other factors, in particular for labor and capital, are relatively flat *in the long run*. Workers have a choice of working in Philadelphia or elsewhere. If their after-tax wage income is not commensurate with what they can receive elsewhere (taking into account the special amenities of Philadelphia), they will leave Philadelphia for employment elsewhere, and firms will not be able to recruit new workers. Thus, if a city such as Philadelphia imposes a wage tax, in the long run, wages must rise to fully offset the tax. Similarly, owners of capital have a choice of investing in Philadelphia or elsewhere. If their return is not commensurate with what they receive elsewhere, they will not invest in Philadelphia. Thus, after-tax wages and after-tax returns to capital are unaffected by the tax. Who, then, pays the tax? Only factors that are not mobile bear the brunt of the tax; land, in particular, is not mobile and is in inelastic supply.

Pittsburgh, across the state, takes a different route from Philadelphia, taxing unimproved land directly, and at a much higher rate than it taxes improvements. Pittsburgh is the only major U.S. city that uses a graded property tax—under which land and buildings are taxed at different rates. In 1979 and 1980, Pittsburgh restructured its property tax system so land was taxed at more than five times the rate on buildings (or improvements).*

---

*For more discussion on the property tax in Pittsburgh and the economic effects of this property tax experiment, see W. E. Oates and R. M. Schwab, "The Impact of Urban Land Taxation: The Pittsburgh Experience," *National Tax Journal* 50, no. 1 (March 1997): 1–21.

TAX INCIDENCE AND THE DEMAND AND SUPPLY FOR LABOR   Figure 18.7A depicts the market demand and supply curves for labor. It makes no difference whether a tax on labor is imposed on consumers (in this case, the firms that pay for the use of labor) or on producers (in this case, the individuals who are selling their labor services); the incidence of the tax is the same. The distinction made by Congress, that half of the Social Security tax should be paid by the employer and half by the employee, makes absolutely no difference for the effect of the tax. The consequences would have been the same had Congress said that firms must pay the entire tax or that individuals must pay the entire tax.[2]

---

[2] There may be a short-run difference. If Congress had imposed the entire tax on firms, it is unlikely that wages would have fallen immediately. In the short run, the labor market would not have been in equilibrium, and firms would have absorbed a large part of the Social Security tax.

There are also some differences arising out of the income tax. Whereas the employee's contribution to Social Security is included in his or her income (on which the employee must pay income tax), the employer's contribution to Social Security is not. Also, if the individual works for more than one employer and pays more than the maximum Social Security, the individual can claim a refund of the excess, but the employer is not entitled to any refund.

COMPARING THE EFFECTS OF A TAX ON THE DEMAND FOR LABOR

(A) The effect of a tax on labor is to shift the demand curve for labor down. A tax on labor will lead to a lower wage and a lower level of employment. (B) With a backward-bending supply schedule, the wage may fall by more than the amount of the tax.
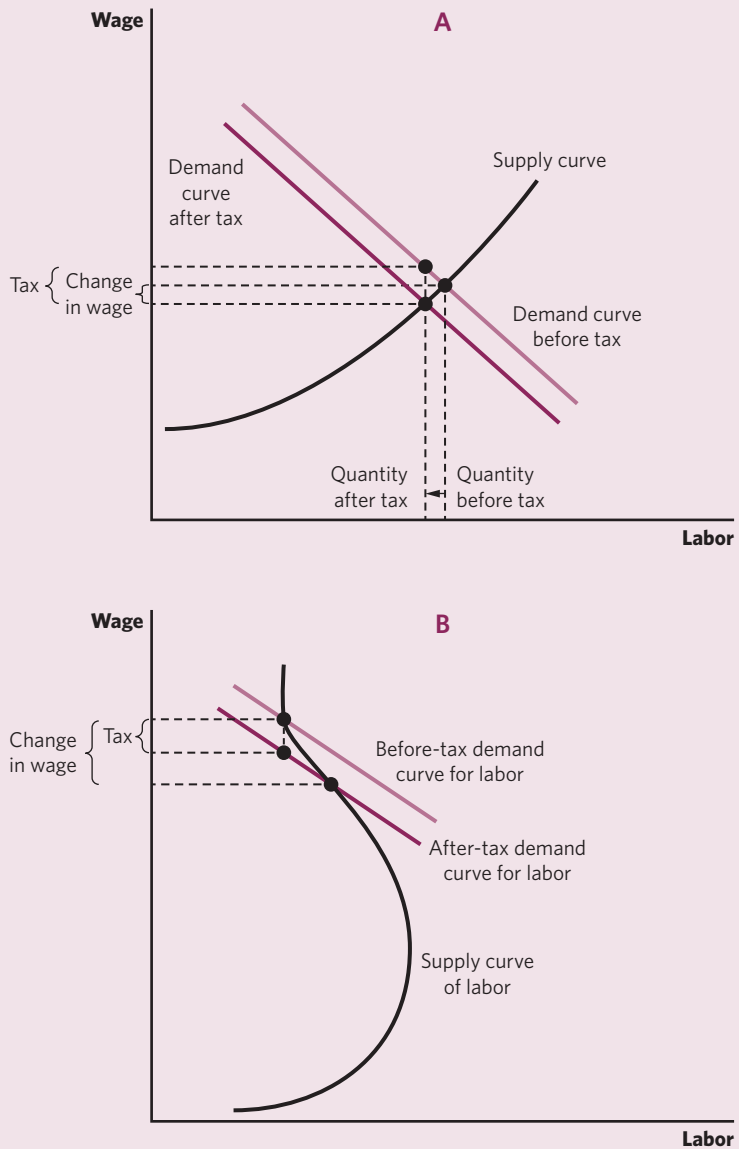
Who effectively pays the tax depends on the elasticity of demand and supply for labor. If, as is frequently claimed, the supply of labor is rela- tively inelastic—that is, almost vertical—most of the burden of the tax falls on workers, regardless of the legal imposition of the tax.

Some economists believe that the supply curve of labor actually is backward-bending, as illustrated in Figure 18.7B. As the wage rises above a certain level, the supply of labor actually decreases. Individuals decide that, at the higher standards of living that they can attain with the higher wages, they prefer to work less. Thus, higher wages reduce the supply of labor rather than increase it. In this case, a tax on labor may result in a reduction in the wage rate that is greater than the tax itself, as the decrease in wages induces a larger labor supply, which drives down the wage further.

TAXATION OF INELASTIC FACTORS   As we have noted, if the supply elasticity of labor or of a commodity is zero, the tax is borne fully by the supplier. The classic example of a commodity with a zero supply elasticity is unimproved land. The supply of land is fixed. Thus, if a tax is imposed on unimproved land, the total burden of the tax will fall on the landowners.

Unfortunately, it is difficult to distinguish the value of land from the value of improvements to it. In many parts of the United States, for instance, land in the wilderness, with no access to roads, sewers, or water, has almost no commercial value. It is difficult to ascertain how much of the value of land in urban areas is inherent in the land and how much is attributable to improvements. Because the supply elasticity of land improvements is large, a land tax may be largely shifted.

Another example of a factor in long-run inelastic supply is crude oil. Hence, a tax on oil is borne primarily by the owners of oil deposits. Because a disproportionate share of the world's oil is owned by those outside the major consuming nations, the consuming nations have strong incentives to impose taxes on oil. Of course, owners of oil wells in the United States actively resist these taxes, and they are a sufficiently powerful lobbying group to have done so quite successfully. In the United States, taxes on oil are far lower than those in most Western European nations.

TAXATION OF PERFECTLY ELASTIC FACTORS
Just as taxes imposed on perfectly inelastic factors of production are borne totally by the factor, taxes on perfectly elastic factors are not borne at all by the taxed factor; they are entirely shifted. This simple observation has important implications for tax policy. The supply of capital to a small country is usually thought of as being highly elastic: just as a small firm must take the price it pays for capital as given, so too does a small country in an open, global market. The country cannot induce capital to flow in if it pays less than the market rate of

TAXATION OF FACTORS

- The incidence of a tax on a factor in a competitive market depends on the elasticity of supply and demand for the factor.
- The incidence of a tax on a factor whose supply is perfectly inelastic is borne completely by the factor.
- A tax on a factor whose supply is perfectly elastic is completely shifted.

FIGURE 18.8

PERFECTLY ELASTICALLY SUPPLIED FACTOR

The incidence of a tax imposed on a perfectly elastically supplied factor is always fully shifted. The demand curve is shifted down by the amount of tax, leaving the price received by suppliers unchanged.
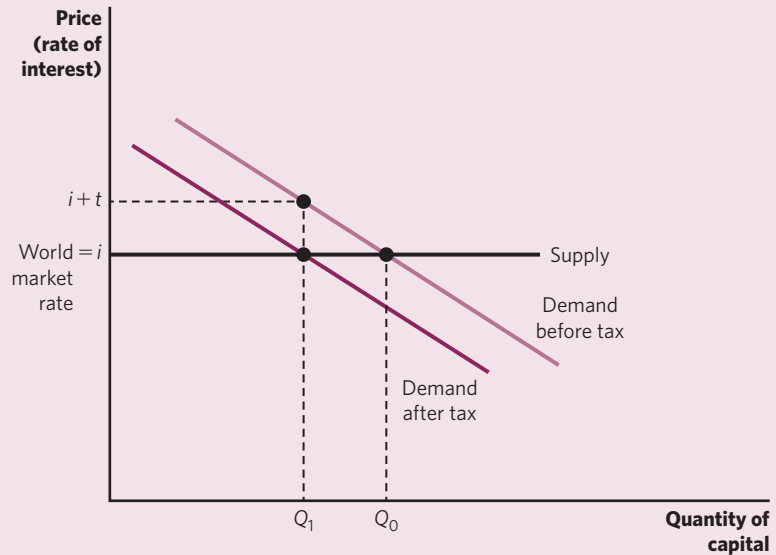
interest, but at the market rate, it can obtain all the capital it could possibly absorb. Figure 18.8 plots the demand and supply of capital depending on the interest rate. With a tax on interest, the interest paid differs from the interest received. The capital owner must receive the market rate, however, or he or she supplies nothing. The users of capital must make up the difference, paying $i + t$. In the figure, the vertical axis represents the interest rate received, so the supply curve remains unchanged. The tax shifts the demand curve for capital down. At the new equilibrium, the interest rate received is unchanged. A tax on interest in this situation is fully shifted from capital owners to capital users.

TAX INCIDENCE IN ENVIRONMENTS WITHOUT PERFECT COMPETITION

The effect of the imposition of a tax depends critically on the nature of the market. The analysis in the preceding section assumed that markets are competitive. However, if markets are less than fully competitive—if, for instance, the industry consists of a monopoly or of firms acting collusively, so that their combined behavior is similar to that of a monopoly—the effect of a tax could be markedly different.

Price of aluminum

A

Price change $\{$ p_1 p_0

Demand curve

Marginal cost after tax

Tax $\{$

Marginal cost before tax

Marginal revenue

Q_1 Q_0

Quantity of aluminum

Price

B

$\{$ p_1

Demand curve

Price change $\{$

p_0

Marginal cost after tax

Tax $\{$

Marginal cost before tax

Marginal revenue

Q_1 Q_0

Quantity
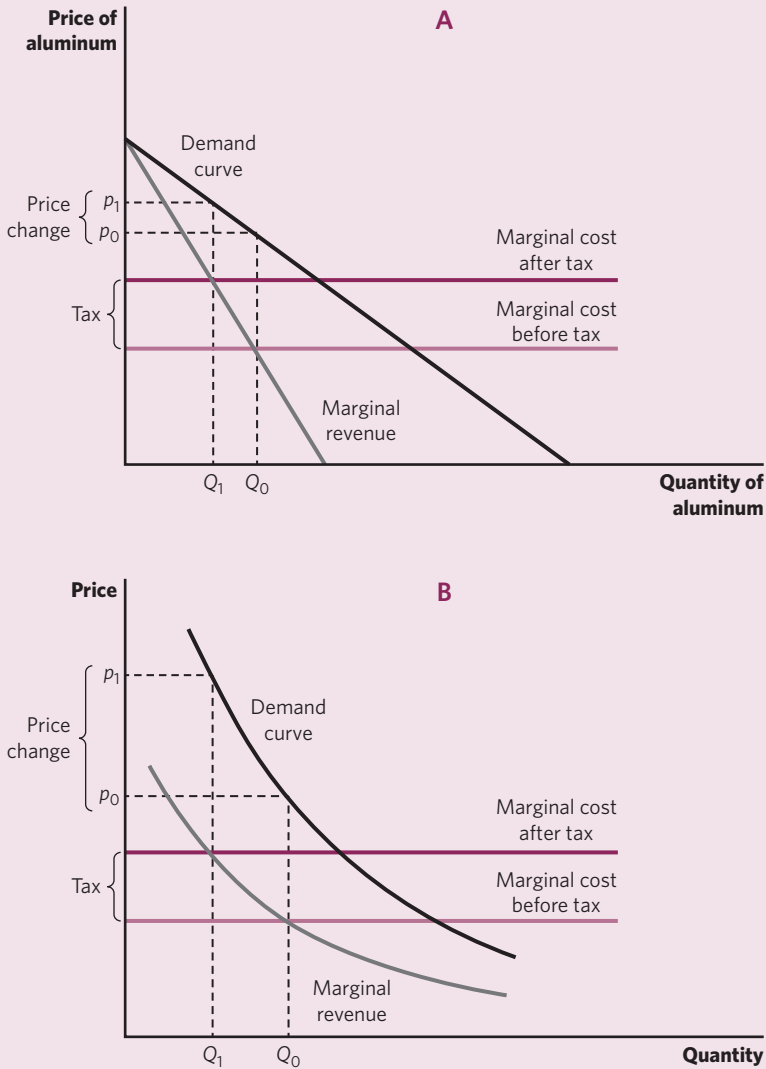
FIGURE 18.9

TAXING A MONOPOLY
(A) With linear demand and horizontal marginal cost curves, the price paid by consumers rises by exactly half the tax; consumers and producers share the burden of the tax. (B) With constant elasticity demand curves, the price rises by more than the tax.

In the absence of a tax, a monopolist will choose that level of output such that the cost of producing any additional output (the marginal cost) is just equal to the additional sales revenue it would receive (its marginal revenue). To maximize profits, the monopolist thus sets its **marginal cost** equal to its **marginal revenue**.

Figure 18.9 depicts the demand curve for aluminum, the marginal revenue curve, and the marginal cost of production. The marginal revenue curve lies below the demand curve. It represents the extra revenue the firm receives from selling an extra unit of output. The marginal revenue is the price the

firm receives for that extra unit, minus the loss it sustains on the other units it sells, because as it attempts to sell more, it must lower the price.[3] The monopolist chooses Q_0 as its level of output, the quantity at which the marginal cost and marginal revenue curves intersect. To find the price charged by the monopolist, we go up to the demand curve and locate price p_0.

A tax on aluminum can be viewed simply as an increase in the cost of production, which is to say, a shift upward in the marginal cost curve. This will reduce output to Q_1 and increase the price to p_1.

RELATIONSHIP BETWEEN THE CHANGE IN THE PRICE AND THE TAX

In the case of a competitive industry, we showed that the consumer price increased by an amount that normally was less than the tax, and that the magnitude of the price increase depended on the demand and supply elasticities. The results for a monopolist are more complicated.

First, the steeper the marginal cost curve, the smaller the change in output and hence the smaller the increase in price. With a perfectly vertical marginal cost schedule, there is no change in output and no change in price; the tax is borne by producers. A supply (or marginal cost) curve is perfectly vertical if no increase in price calls forth an increase in supply. This result parallels that for competitive markets.

On the other hand, with a horizontal marginal cost schedule, as in Figure 18.9, the extent to which producers or consumers bear the tax depends on the *shape* of the demand curve. (Contrast this to competitive markets, in which the consumer would bear the entire tax.) Figure 18.9 illustrates two possibilities. With a linear demand curve, as in Figure 18.9A, the price rises by exactly half the tax.[4] With a constant

---

[3] Recall that, in contrast, a perfectly competitive firm must take the market price as fixed, but can sell any amount of output at that price. Its marginal revenue is simply the market price.

[4] With a linear demand curve,

$$p = a - bQ,$$

price and output are related linearly.

Revenue, pQ, is given by

$$pQ = aQ - bQ^2,$$

so marginal revenue, MR, is

$$a - 2bQ.$$

This is set equal to marginal cost plus the tax:

$$a - 2bQ = \text{MC} + t$$

or

$$2a - 2bQ = 2p = a + \text{MC} + t$$

or

$$p = \frac{a + \text{MC} + t}{2}.$$

Hence, if t increases by, say, $2, p increases by $1.

elasticity demand curve, in which a 1 percent increase in the price results in, say, a 2 percent reduction in the demand, regardless of the price level, marginal revenue is a constant fraction of the price[5]:

$$MR = p\left(1 - \frac{1}{\eta^d}\right),$$

where η^d is the elasticity of demand (a constant).

Because the monopolist sets marginal revenue equal to marginal cost,

$$MR = MC \quad \text{or}$$

$$p\left(1 - \frac{1}{\eta^d}\right) = MC \quad \text{or}$$

$$p = \frac{MC}{1 - 1/\eta^d}.$$

A tax has the same effect as raising the marginal cost of production; that is,

$$p = \frac{MC + t}{1 - 1/\eta^d}.$$

Hence, the price increases by a multiple $1/(1 - 1/\eta^d)$ of the tax. If η^d is 2, then the increase in price is twice the tax.

---

[5] This formula is general. In the case of a constant elasticity demand curve, η^d is constant. To derive the formula, recall that marginal revenue is the extra revenue received from producing one more unit. Revenue is just the price received per unit, p, times the number of units sold, Q. Thus, when a firm sells one more unit, it receives p, but to sell the additional unit, it must have reduced its price from its previous level. Denote the change in price by Δp. The firm loses this amount on all sales, Q. Thus, the net gain is

$$MR = p + \Delta p \cdot Q = p\left(1 + \frac{\Delta p}{p} \cdot Q\right).$$

Recall, too, that

$$-\frac{\Delta Q/Q}{\Delta p/p} = -\frac{\text{change in } Q/Q}{\text{change in } p/p}$$

is just the percent change in quantity as a result of a percent change in price, which is just the elasticity of demand.

Here, the change in quantity is just 1; that is, $\Delta Q = 1$, so we can rewrite

$$-\frac{\Delta p}{p} \cdot Q = -\frac{\Delta p}{p} \cdot \frac{Q}{\Delta Q} = -\frac{\Delta p/p}{\Delta Q/Q} = \text{elasticity of demand} = \eta^d.$$

So,

$$MR = p\left(1 - \frac{1}{\eta^d}\right).$$

AD VALOREM VERSUS SPECIFIC TAXES

There is another important difference between the taxation of competitive and monopolistic industries. In the case of competitive industries, the form in which the tax is levied makes no difference. We can choose between a specific tax, which is specified as a fixed amount per unit of output, and an ad valorem tax, which is specified as a percentage of the value of the output. All that matters for determining the effect of the tax is the magnitude of the difference (in equilibrium) between the price received by producers and the price paid by consumers, what we refer to as the *wedge* between the two.

In the case of monopolistic industries, however, ad valorem and specific taxes have quite different effects. We show in the appendix to this chapter that for any given revenue raised by the government, the monopolist's output will be higher with an ad valorem tax than with a specific tax.

TAX INCIDENCE IN OLIGOPOLIES

Between the extremes of perfect competition and monopoly is the oligopoly market structure. In an **oligopoly**, such as the airline market and the rental car market, every producer interacts strategically with every other producer. If one producer changes its prices or output, the other producers may also change their prices or outputs, but these responses may be hard to predict.

There is no widely accepted theory of firm behavior in oligopoly, so it is impossible to make any definite predictions about the incidence of taxation in this case. Some economists believe that oligopolists are not likely to raise the prices they charge consumers when taxes change. Each oligopolist may believe that if it raises its price, other firms will steal its market share. An opposite conclusion follows if each oligopolist expects that its competitors will match its price increase after a tax is imposed. In this case, all will raise their prices and thereby shift the burden of the tax to consumers.

Although economists have explored the incidence of taxes in an oligopoly under different specific behavioral assumptions, until they gain a better understanding of oligopolistic behavior, there can be no general theory of the incidence of a tax in an oligopolistic market.

TAX INCIDENCE IN MONOPOLIES OR IMPERFECTLY COMPETITIVE MARKETS

- In monopolies or imperfectly competitive markets, tax incidence depends on the shape of the demand and supply curves; there may be more than 100 percent shifting.
- In a monopoly, with constant marginal cost, and with constant elasticity demand curves, there will always be more than 100 percent shifting of specific commodity taxes. With linear demand curves, price rises by half the tax.

EQUIVALENT TAXES

In the discussion thus far, several instances have been pointed out in which taxes appear to be different—a tax on employers to finance Social Security and a tax on employees; a tax on the producers of beer or a tax on beer consumers—but are really equivalent. Many other examples of taxes that appear to be very different (and that from an administrative point of view *are* different) are, from an economic point of view, equivalent.

INCOME TAX AND VALUE-ADDED TAX

An obvious example follows from the basic identity between national income (the total of what all the individuals in our society receive) and national output (the total of what they all produce). Because the value of income and the value of output must be the same, a uniform tax on income (a tax that taxes all sources of income at the same rate) and a uniform tax on output (a tax that taxes all outputs at the same rate) must be equivalent. A comprehensive uniform sales tax is a uniform tax on output and is thus equivalent to a uniform income tax.

The production of any commodity entails a large number of steps. The value of the final product represents the sum of the value added at each stage of production. We could impose the tax at the end of the production process, or at each stage along the way. A tax at the end of the production process is called a *sales tax*. A tax imposed at each stage of the production process is called a *value-added tax*. Thus, a uniform value-added tax and a comprehensive uniform sales tax are equivalent—and both are equivalent to a uniform income tax.

The value-added tax is used in most European countries, and there has been some discussion in the United States about introducing such a tax. Because a uniform value-added tax is equivalent to a uniform (proportional) income tax, replacing our current income tax system with a value-added tax would be equivalent to replacing it with a proportional income tax system.

The value-added tax in Europe typically exempts investment goods. It is imposed only on consumption. Thus, the European form of the value-added tax is equivalent to a tax on consumption. Because consumption is equal to income minus savings, a consumption tax is equivalent to a tax on income in which savings are exempted.

EQUIVALENCE OF CONSUMPTION AND WAGE TAXES

Suppose that individuals receive no inheritances and leave no bequests. Then a uniform tax on wages and a uniform tax on consumption are equivalent. To put it another way, a consumption tax is equivalent to an income tax in which interest and other returns to capital have been exempted. (Our present tax system, in which part of the return to capital is tax exempt, can be viewed as somewhere between a consumption tax and an income tax.)

The equivalence may be seen most clearly by looking at the lifetime budget constraint of an individual (with no inheritances or bequests). For simplicity, we divide the life of the individual into two periods: wage income is w_1 in the first period and w_2 in the second. The individual has to decide how much to consume the first period of life, while young, and how much when older. If the individual reduces consumption today by a dollar and invests it, next period he or she will have $1 + r$ dollars, where r is the rate of interest. With a 10 percent interest rate, he or she will have $1.10. The budget constraint is a straight line, depicted in Figure 18.10.

Consider what happens to the individual's budget constraint when a wage tax of 20 percent is imposed. The amount that he or she can consume shifts down. The slope of the budget constraint remains unchanged:

FIGURE 18.10

COMPARING THE EFFECTS OF A CONSUMPTION TAX AND WAGE TAX

A consumption tax and a wage tax have exactly the same effect on the individual's budget constraint.
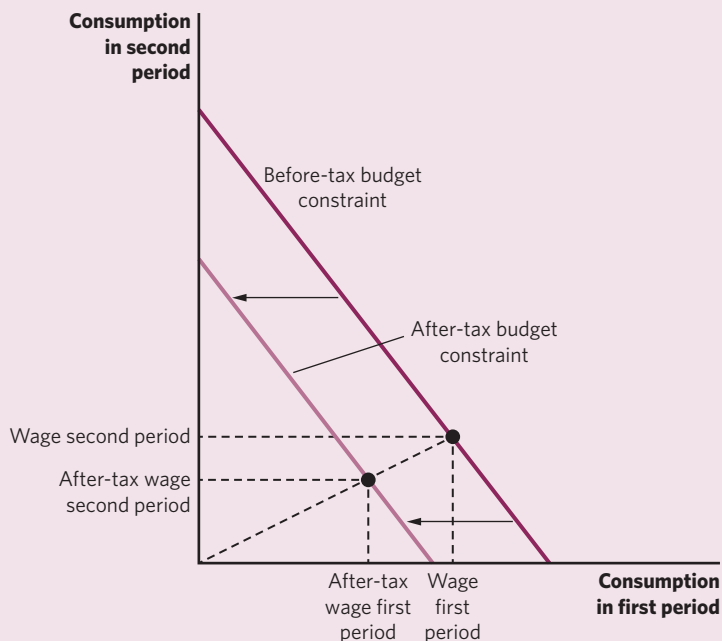
it is still the case that by giving up $1 of consumption in the first period, the individual can get $1.10 next period.

Now, consider what happens to the individual's budget constraint when a 20 percent consumption tax is imposed. Just as before, the amount that he or she can consume shifts down, and the slope of the budget constraint remains unchanged. If the individual spends $1 today, he or she gets 20 percent fewer goods because of the tax; however, when the individual spends $1 tomorrow, he or she also gets 20 percent fewer goods because of the tax. The trade-off between spending today and spending tomorrow remains unchanged. A wage tax and a consumption tax are equivalent.[6] Only the timing of the revenues to the government differs between the two taxes; this may be important if capital markets are imperfect.

There are, again, several ways that equivalent taxes can be imposed. We can impose a tax on wage income in each period, exempting all interest, dividends, and other returns on capital, or we can tax consumption in each period, which can be calculated by having the individual report his or her total income minus total savings.

> ### EQUIVALENT TAXES
>
> - Two sets of taxes are equivalent if their incidence is exactly the same.
> - Income taxes and value-added taxes (without exemption for investment) are equivalent.
> - Consumption taxes and value-added taxes with an exemption for investment are equivalent.
> - Consumption and wage taxes are equivalent.
> - Lifetime income and consumption taxes (in the absence of bequests and inheritances) are equivalent.

EQUIVALENCE OF LIFETIME CONSUMPTION AND LIFETIME INCOME TAXES

This analysis has one other important interpretation. Continuing with our example in which the life of an individual is divided into two periods, we can write the budget constraint[7] as

$$C_1 + \frac{C_2}{1 + r} = w_1 + \frac{w_2}{1 + r}.$$

---

[6] If there are bequests and inheritances, a wage-plus-inheritance tax is equivalent to a consumption-plus-bequest tax. These equivalency relations require a perfect capital market, but are true even if there is risk. See A. B. Atkinson and J. E. Stiglitz, *Lectures on Public Economics* (New York: McGraw-Hill, 1980), Lecture 3.

[7] This can be seen in a slightly different way. An individual's savings (borrowings) are the difference between wages and consumption in the first period:

$$w_1 - C_1.$$

Consumption in the second period is thus second-period wage income plus the savings with its interest minus borrowings with interest:

$$C_2 = w_2 + (1 + r)(w_1 - C_1).$$

Rearranging terms, we have

$$C_1(1 + r) + C_2 = (1 + r)w_1 + w_2.$$

Dividing by $(1 + r)$, we obtain the budget constraint in the form presented in the text.

The left-hand side of the equation is the present discounted value of the individual's consumption, and the right-hand side is the present discounted value of wage income. In the absence of bequests and inheritances, the present discounted value of consumption must equal the present discounted value of (wage) income. Thus, a lifetime consumption tax and a tax based on lifetime income are equivalent, as we saw in Chapter 17.

A CAVEAT ON EQUIVALENCE

The fact that two taxes are equivalent does not mean that there are no effects when one tax is switched to the other (or from some third tax to either of the two). Equivalence means only that the two taxes have exactly the same effects in the long run; in the short run—including the transition period as the tax is adopted—the effects may differ markedly. Take, for example, a switch from a lifetime income tax to a lifetime consumption tax. Leaving aside the problems of transition, the result would be that the elderly would face double taxation: in their youth, they paid taxes on wages, and in their retirement, they pay taxes on their consumption; or, say, a value-added tax is imposed. In the short run, prices consumers face rise, and more of the burden of the tax in the short run may be shifted to consumers than if the same revenues were raised by an income tax.

OTHER FACTORS AFFECTING TAX INCIDENCE

So far, we have shown that what determines who bears the burden of any tax is not who Congress says should bear it, but certain properties of demand and supply curves, and the nature of the market—whether it is competitive, monopolistic, or oligopolistic.

TAX INCIDENCE UNDER PARTIAL AND GENERAL EQUILIBRIUM

Several other important factors need to be taken into account in incidence analysis. First, there is an important distinction between a tax in a single industry and a tax affecting many industries. Earlier, we considered a tax

BEHAVIORAL ECONOMICS, MANAGERIAL CAPITALISM, AND TAX INCIDENCE

We have seen in this chapter how the incidence of a tax (who ultimately pays, and the full consequences of paying) depends on the nature of the market—whether markets are, for instance, highly competitive or more monopolistic. In recent years, economists have focused on ways in which economic behavior may not be well described by the standard models; and, if that is so, the analysis of tax incidence will have to be modified accordingly.

One important school of thought, called **behavioral economics**, focuses on behavior that is often not consistent with, or well explained by, the standard model of rational individuals maximizing their lifetime utilities that we have employed in earlier chapters. In some cases, behavioral responses may in fact be greater than predicted by the standard theory, sometimes smaller (see Chapter 10, "Private Sector Responses to Government Programs").

For instance, government programs to encourage savings through special programs such as IRAs (described in more detail in Chapter 21) may be more effective than standard theory predicts, because individuals like the idea of a "sale"—here, there is a special tax deal—*even though at the margin there is no benefit* (and standard theory says what really matters is how the tax system affects returns on the margin). The special program calls attention to the importance of savings, and people respond.

With respect to tax incidence, perceptions of fairness in distribution of the tax burden—whether, on average, high-income taxpayers pay more than low-income taxpayers—and their impact on tax compliance can be affected more by the mass media than by statistics. For example, despite progressivity in the income tax and the large share of total revenue coming from high-income taxpayers, newspaper articles identifying particular individuals or corporations that use the prowess of highly paid tax lawyers to reduce their tax burden create widespread popular resentment and potentially less willingness to pay one's own taxes. Behavioral economics studies how such perceptions of fairness are formed.

Another major strand of modern economics focuses on the behavior of corporations. Proponents of **managerial capitalism** argue that corporations are not really controlled by their owners, as claimed by the theory of **shareholder capitalism**, but by their managers, and if we are to understand how corporations respond to the incidence of taxes, we have to focus more on the decision making of managers and the incentives they face.

For example, how does preferential tax treatment of capital gains versus earned income affect the form and magnitude of executive compensation? Does this help explain why senior managers now receive a large portion of their pay in the form of corporate stock, and why their total compensation has grown so much? How does this, in turn, affect managerial decision making in the trade-off between actions that will increase share price in the short term but harm long-term corporate competitiveness?

on a small industry (beer). The presumption is that such a tax will not, for instance, have any significant effect on the wage rate. Although the reduction in the demand for beer will reduce the demand for labor in the beer industry, the assumption is that this industry is so small that workers released from their jobs can find employment elsewhere without any significant effect on the wage rate. We refer to this kind of analysis, in which we assume that all prices and wages (other than those on which attention is explicitly focused) remain constant, as **partial equilibrium analysis**.

Unfortunately, many taxes affect many industries simultaneously. The corporate income tax affects all incorporated businesses. If, as a result of the tax, incorporated businesses reduce their demand for capital, the capital released cannot be absorbed by the rest of the economy (the unincorporated sector) without reducing the return to capital there. Thus, we cannot assume that what the corporate sector must pay to obtain capital is independent of the tax imposed on that sector. To analyze the impact of the corporation tax requires analyzing its effect on the equilibrium of the entire economy, not just the businesses on which the tax is imposed. Such an analysis is called a **general equilibrium analysis**. In many instances, the general equilibrium impact of a tax may be markedly different from the partial equilibrium effect. For instance, if capital can be shifted relatively easily from the incorporated to the unincorporated sectors of the economy, the tax on corporate capital must be borne equally by capital in *both* sectors of the economy; they both must have the same after-tax return.

The overall incidence of the corporation income tax, like the tax on any factor, depends on the elasticity of demand and supply curves. Although we will postpone until Chapter 23 a fuller discussion of the incidence of the corporation income tax, we can see why the general equilibrium impact may be markedly different from the apparent effect by considering the limiting case in which the supply curve of capital is perfectly elastic.

Savers insist on a return r^*, as depicted in Figure 18.11. Below r^*, they supply no capital; at r^*, they are willing to supply an arbitrarily large amount. That means that the *after*-tax return to capital—in both the corporate and the unincorporated sector—must be r^*, so the before-tax return in the corporate sector must be $r^* + t$. The tax simply raises the before-tax cost of capital in the corporate sector. This has two effects. First, it raises the price of the products produced in the corporate sector, reducing demand for them; demand is shifted to the unincorporated sector. Second, within the corporate sector, firms use more labor and less capital. In general, some of the tax is shifted to workers and some is shifted forward to consumers of the goods the corporate sector produces. However, the magnitude of the effect on, say, workers, depends on, for instance, how easily firms in the corporate sector can substitute labor for capital and on the relative labor intensity of goods in the unincorporated and
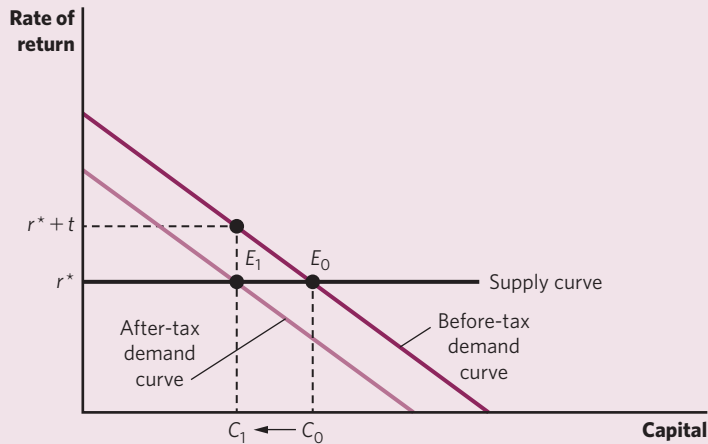
corporate sectors. If firms in the corporate sector can easily substitute labor
for the more costly capital, and if goods in the unincorporated sector are rel-
atively labor intensive, then the general equilibrium effects may lead to an
overall increase in the demand for labor, so wages actually increase, if labor
is inelastically supplied. In that case, the burden of the corporate income tax
lies on consumers of the goods produced by the corporate sector. Workers and
owners of capital may both be adversely affected by the price increase, but the
relative impact may depend as much on consumption patterns as on anything
else. If owners of capital largely consume services produced by the unincor-
porated sector, while workers consume more manufactured goods, then more
of the burden of the tax may be borne by workers.

   Three important points emerge from this analysis:

1. Corporations do not bear taxes, people do: shareholders, workers,
   consumers.

2. Because of general equilibrium responses, the impacts of corporation
   taxes are felt not just in the corporate sector but throughout the economy.

3. The effects may vary depending on the period of analysis and on var-
   ious assumptions about the structure of the economy. Can we assume
   that the overall stock of capital is fixed, or that capital can be shifted
   from one use to another with some degree of ease or difficulty? Can
   labor be easily substituted for capital? Can labor and capital move from
   one sector of the economy to another? The answers to such questions
   are critical to determining the effects of the tax.

SHORT-RUN VERSUS LONG-RUN EFFECTS

A distinction must also be made between the incidence of the tax in the long run and in the short run. Many things are fixed in the short run that can vary in the long run. Although capital currently being used in one industry (like steel) cannot easily be shifted for use into another, in the long run new investment can be shifted to other industries. Thus, a tax on the return to capital in the steel industry may have markedly different effects in the long run than in the short run.

If savings are taxed, the short-run effect may be minimal. In the long run, however, the tax may discourage savings, and this may reduce the capital stock. The reduction in the capital stock will reduce the demand for (and productivity of) labor, and this, in turn, will lead to a lowering of wages. As a result, the *long-run incidence* of a tax on savings (or capital) may be on workers, even if the *short-run incidence* is not.

The short run may differ from the long run also because of dynamics of adjustment. For instance, even in fairly competitive markets, firms frequently set prices initially by certain rules of thumb, which entail a given markup over variable costs. The long-run equilibrium in these industries has the markup adjust to the competitive level. In the short run, the market may be out of equilibrium.

The distinction between short-run and long-run effects is important, because governments and politicians are often shortsighted. They observe the immediate effect of a tax without realizing that the full consequences may not be those that they intended.

A number of factors affect the disparity between the short-run and long-run effects, and between the partial and general equilibrium effects.

OPEN VERSUS CLOSED ECONOMY

One of the most important factors is whether the economy is closed (does not trade with other countries) or open. If a small, open country like Switzerland imposed a tax on capital, the before-tax rate of return would have to adjust fully to offset the tax, otherwise investors would withdraw their funds from Switzerland and invest elsewhere; the tax would be borne by land and labor. Effectively, the supply schedule for capital is infinitely elastic. The same analysis applies, of course, to any state within the United States.

ASSOCIATED POLICY CHANGES

The final aspect of incidence analysis that needs to be discussed here is that it is almost never possible for the government to change only one policy at a time. There is a basic government budget constraint, which says that tax revenues plus the increase in the size of the deficit (increased borrowing) must equal government expenditures. If the government raises some tax rate, it must either lower another, reduce its borrowing, or increase its expenditure. Different combinations of policies will have different effects. We cannot simply ask the question: What would happen if the government increased income taxes? We need to specify whether the income tax is to be accompanied by a reduction in some

> ## FACTORS AFFECTING INCIDENCE
>
> Time span: short run versus long run
> - Demand and supply curves are likely to be more elastic in the long run than in the short run.
>
> Open versus closed economy
> - Supply curves of factors are more elastic in an open economy.
>
> Mix of policy changes
> - Differential tax analysis: one tax is substituted for another, keeping revenue constant.
> - Balanced budget analysis: expenditure is changed as tax revenues change.
> - Balanced growth analysis: a mix of policies which leaves capital accumulation unaffected.

other tax, by an increase in government expenditure, or by a reduction in government borrowing. (Often the accompanying change is taken to be understood but not made explicit; for example, if taxes are raised, there will be a smaller deficit.)

We call the analysis of a tax increase accompanied by a decrease in some other tax **differential tax incidence analysis**; we call the analysis of a tax increase accompanied by an increase in government expenditure a **balanced budget tax incidence analysis**. Such exercises have become particularly relevant in recent years as budgetary processes in the United States have been reformed in an attempt to control the deficit. Under what are called "Pay-As-You-Go" (PAYGO) rules, any increase in expenditure has to be matched by a decrease somewhere else, or by a new source of tax revenue.[8]

Sometimes we are interested in analyzing combinations of policies that leave some important economic variable unchanged. For example, a tax increase may lead to a reduction in output. We may want to distinguish the effects of a tax program on the level of output (and the effects that this may have, say, on its distribution) from the direct effects of the tax itself; thus, we may look at combinations of policies that leave the level of national output unaffected.

---

[8] Much of the focus on balanced budget incidence relates to macroeconomic consequences. Lower taxes or increased expenditures lead to higher levels of aggregate demand, unless offset by tighter monetary policy. Today, most analyses of tax and expenditure incidence assume that the monetary authorities will take offsetting actions to maintain the economy at full employment. These offsetting actions have, of course, distribution and other general equilibrium effects. Thus, a full analysis of the incidence of any set of tax or expenditure policies needs to take into account the consequences of the offsetting actions of the monetary authorities.

TAX INCIDENCE OF SPECIFIC TAX PROVISIONS

Just as we can analyze the incidence of a payroll tax or any other particular tax, we can analyze the incidence of a particular provision of the tax code—and the incidence is often not what it seems, or is supposed to be.

Consider, for instance, the provision that allows states and municipalities to issue tax-exempt bonds. The provision is supposed to allow them to benefit from lower borrowing costs. If the top marginal tax rate is approximately 40 percent, then an individual in the top rate would be indifferent between buying a taxable bond of equal risk yielding 10 percent and a tax exempt bond yielding 6 percent. Given the large amounts borrowed, this provision could be of enormous benefit to states and localities.

In practice, however, things do not work out like this. The yield on municipal bonds is lower as a result of the tax preferred treatment, but only slightly so. As an example, assume that the interest rate fell to 8 percent. That would mean that rich individuals would get a much higher *after-tax* income by investing in municipal bonds. Half the benefits of the tax provision go to helping states and municipalities, but half of the benefits go to enriching those at the top.

Similarly, many taxes have an effect on the level of capital accumulation. The fall in the capital stock, in turn, may lower wages. Again, one may want to distinguish the direct from the indirect effects of a tax resulting from its impact on capital accumulation. This is particularly the case if one believes that other instruments can be used to offset these indirect effects. If an inheritance tax reduces capital accumulation, for example, it may be possible to undo the effects by providing an investment tax credit. We may examine a set of policies whose effect is to leave capital accumulation unaffected; incidence analysis of this sort is called **balanced growth incidence analysis**.

INCIDENCE OF TAXES IN THE UNITED STATES

In this chapter, we have explained why the actual burden of taxes does not necessarily fall on those on whom the tax is imposed. Officially, the United States, like most advanced countries, has a **progressive** tax system, one

in which the rich are supposed to pay a higher proportion of their income in taxes than the poor. The income tax imposes a 35 percent tax rate on the rich, whereas poor families receive as much as a 45 percent subsidy (through the earned income tax credit). However, there is a consensus that, overall, the U.S. tax system is far less progressive than the official tax code might suggest. (A tax system is said to be **regressive** if the poor pay a higher percentage of their income in taxes than the rich do.)[9]

There are three reasons for this view. First, the income tax itself is less progressive than appears, because it has certain specific design features that allow certain types of income or categories of individuals to escape taxation, at least partially. For instance, capital gains are taxed at lower rates than other forms of income, and there are a variety of special provisions, discussed later, that lower the **effective tax rate** even more. Individuals can put away savings into tax-exempt accounts, and richer individuals tend to avail themselves of this opportunity more than poor individuals. There may be, as we shall see, good reasons for these and other provisions of the tax code, but their net effect is to reduce the progressivity of the tax system.

Second, the income tax itself is only one of several taxes; many of the other taxes, such as state and local sales taxes[10] and the payroll tax, are less progressive, or even regressive.[11]

Third, the incidence of many taxes differs from those on whom the tax is legislated; workers often bear the effect of taxes that are "intended" for others. As noted previously, there is a consensus among economists that workers, not employers, bear the full burden of the employer share of the Social Security tax. There is also a consensus that much of the corporation income tax is shifted, although there is disagreement about both the extent and to whom it is shifted. As world capital markets have become more integrated, it becomes more likely that the tax is not borne by capital. Whether it is shifted forward to consumers or back to workers is less apparent, but in either case, its impact is less progressive than it would be if it were borne by the owners of corporations.

Precise estimates of the overall burden of the federal tax system clearly depend on assumptions concerning who bears the burden of various

---

[9] The discussion that follows considers progressivity in terms of the ratio of annual taxes to annual income. A more appropriate measure would be lifetime taxes relative to lifetime income or consumption. This distinction is important. Changes in tax policy that may look regressive in the annual measure may not be so in terms of the more fundamental measure, as we shall see in later chapters.

[10] State and local sales taxes tend to be at fixed rates, but they are levied only on the purchases of certain goods. The fraction of income spent on those goods tends to be lower for richer individuals than for the poor; in states where food is exempt, it is the middle-income individuals who pay the highest percentage of their income.

[11] The payroll tax is a fixed percentage of wage income, up to a cap. Thus, higher-wage individuals pay a tax on only a portion of their wage income, and because wealthier individuals, on average, derive a smaller fraction of their income from wages, payroll taxes are an even smaller percentage of the total income for richer individuals. Interpreting whether the Social Security system is regressive as a result is far more controversial, for we need to take into account not only the contributions but also the benefits. Historically, as we saw in Chapter 16, richer individuals have gotten back far more in excess of what they contributed than did poorer individuals, but today, there is a close correspondence between contributions and payments, except for the poor, who receive back more than they contribute.

taxes, such as Social Security payroll taxes and the corporation income tax. Figure 18.12 looks at the effective federal individual income tax rates, whereas Figure 18.13 looks at the effective tax rates, including all federal taxes, assuming that workers bear the full burden of payroll taxes (including those supposedly paid by employers) but that the corporation tax is attributed to households according to their share of capital income. What is remarkable is that although the overall tax rates are clearly higher, the pattern is strikingly similar, with differences in tax rates from quintile to quintile being roughly comparable.

There was a high degree of progressivity at the bottom in 2007, with the poorest 20 percent of the population paying approximately 4 percent of their income in taxes, less than half that of the next quintile's 11 percent. On the other hand, at the very top, progressivity was limited, with the top 1 percent paying only a few percentage points more in taxes than those in the top 10 percent. The data probably overstate the overall degree of progressivity of the U.S. tax system because, as already noted, state and local taxes tend to be less progressive than federal taxes. Also, because only realized capital gains are included in income, the unrealized capital gains—which, until the financial crisis of 2008, were huge—are not included. To include them would result in a lower effective tax rate among the richest Americans. The 2001 and 2003 tax cuts under George W. Bush, by lowering marginal tax rates and providing still further special treatment of capital gains, reduced the progressivity of the tax system, undoing some of the increase in progressivity that was introduced in 1993, when marginal tax rates on upper-income individuals were raised from 28 percent to close to 40 percent. Although they were due to expire at the end of 2010, the Bush tax cuts were extended through 2012.[12]

The issue of incidence has played a major role in recent tax reforms. With each proposal, tables have been drawn up showing how effective tax rates would change. Because politicians find it hard to defend tax cuts for the very rich, a great deal of effort goes into trying to find alternative ways of characterizing the impact of a tax change. For instance, advocates of cutting capital gains tax rates—the main beneficiaries of which are the very rich—argue that such cuts will induce these individuals to sell more of their assets; and because capital gains are taxed only when the asset is sold, at least initially, tax collections from the rich will go up, even if, in the long run, tax collections go down. Thus, advocates of capital gains tax cuts for upper-income individuals focused not on the cut in tax rates, but

---

[12] The increased flows of international capital suggest that more of the burden of the corporate income tax may be borne by consumers and less by capital than is reflected in the figure. In that case, the overall degree of progressivity is less than depicted.

Because benefits are roughly commensurate with Social Security contributions, the *net* tax imposed by the Social Security system is associated only with its redistributions. Lower-income individuals receive somewhat more than they contribute, and higher-income individuals receive somewhat less. Figure 18.13 looks at only payroll taxes, not benefits.
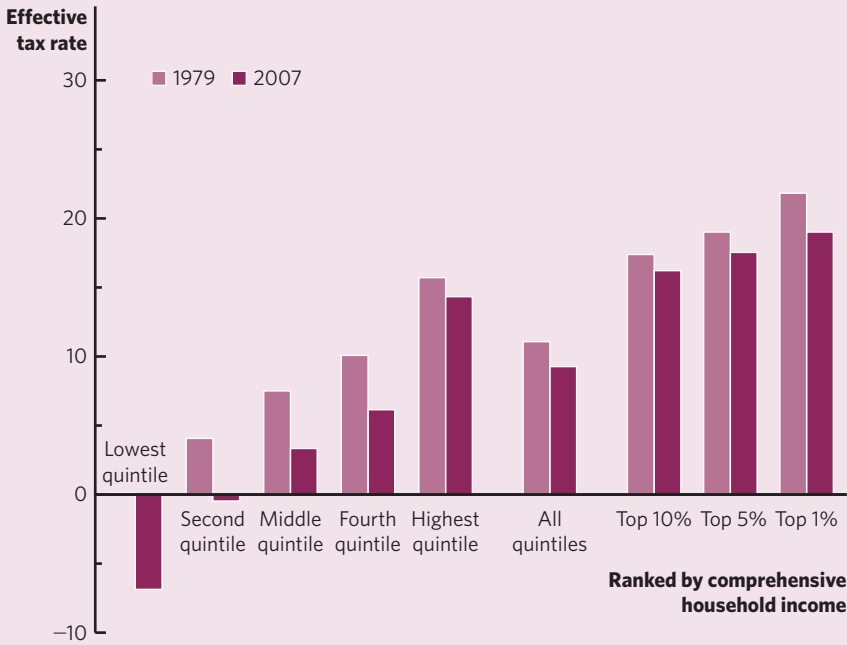
PROGRESSIVITY OF FEDERAL INDIVIDUAL INCOME TAX
Effective tax rates on the individual income tax (the ratio of tax payments to incomes) were far lower than the legislated rates, as a result of a variety of special provisions. Still, the tax schedule exhibited considerable progressivity.

SOURCES: Urban Institute and Brookings Institution Tax Policy Center, and the Congressional Budget Office.
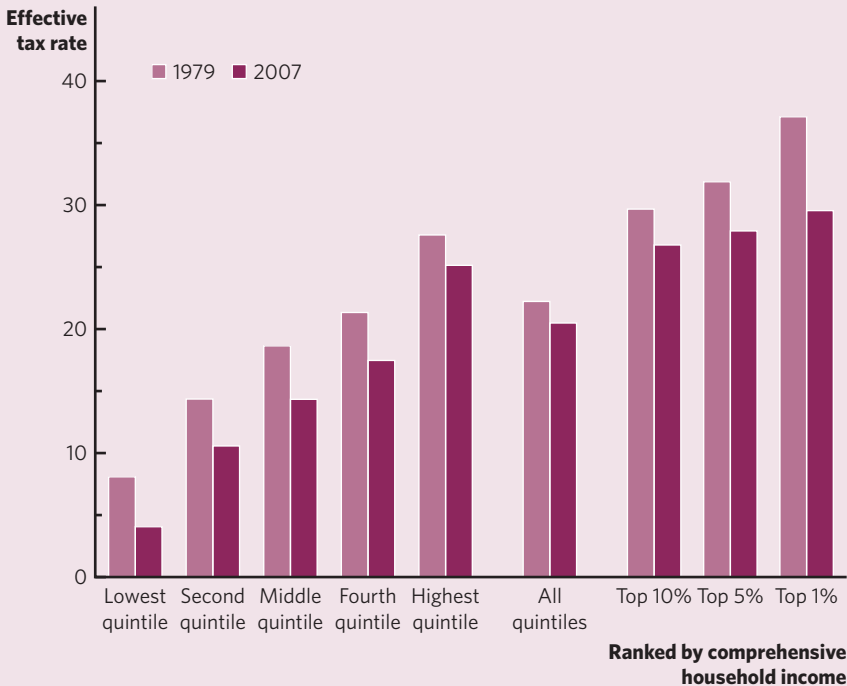
PROGRESSIVITY OF FEDERAL TAX STRUCTURE
When all federal taxes are included, effective tax rates are higher, and remain progressive.

SOURCES: Urban Institute and Brookings Institution Tax Policy Center, and the Congressional Budget Office.

on the increases in tax payments in the initial years. More generally, controversies over incidence—for instance, over who really pays taxes such as the corporation income tax—play a key role in debates over whether particular reforms increase or decrease the progressivity of the tax system. At issue are matters of both theory and empirical analysis, and the impacts often depend on detailed provisions of the tax code. The following chapters of this book will elucidate many of the key issues in these debates.

REVIEW AND PRACTICE

SUMMARY

1. It makes no difference whether a tax is imposed on the suppliers of a factor or commodity or on the consumers. Instead, who bears the burden of the tax depends on the demand and supply elasticities, and on whether the market is competitive or noncompetitive. Taxes induce changes in relative prices, and this market response determines who bears the tax.

2. In a competitive market, if the supply is completely inelastic or if demand is completely elastic, the tax is borne by producers. If the supply is completely elastic or demand is completely inelastic, the tax is entirely borne by consumers.

3. A tax on a monopolist may be shifted more than 100 percent—that is, the price paid by consumers may rise by more than the tax.

4. The general equilibrium incidence of a tax, taking into account repercussions in all industries, may differ from the partial equilibrium incidence. The incidence of a tax may be different in the long run than in the short run.

5. It is almost never possible for the government to change one policy at a time. Differential tax incidence focuses on how substituting one tax for another will affect the distribution of the tax burden.

6. A tax on output (a uniform sales tax), a proportional income tax, and a uniform value-added tax are all equivalent. A uniform tax on wages and a uniform tax on consumption are equivalent.

7. Empirical studies of who bears the burden of the set of taxes imposed in the United States show that the degree of progressivity of the tax structure depends on assumptions concerning the incidence of taxes on corporations and on payrolls. The current United States tax structure has some progressivity, though less than appears "on paper."

KEY CONCEPTS

Ad valorem tax

Balanced budget tax incidence analysis

Balanced growth incidence analysis

Behavioral economics

Differential tax incidence analysis

Effective tax rate

Elasticity of demand

Elasticity of supply

Equivalent taxes

General equilibrium analysis

Managerial capitalism

Marginal cost

Marginal revenue

Oligopoly

Partial equilibrium analysis

Progressive

Regressive

Shareholder capitalism

Shifted backward

Shifted forward

Specific tax

Tax burden

Tax incidence

QUESTIONS AND PROBLEMS

1. Consider a mineral that is in fixed supply, $Q^S = 4$. The demand for the mineral is given by $Q^D = 10 - 2p$, where p is the price per pound and Q^D is the quantity demanded. The government imposes a tax of $2 per pound on the consumer.

   a. What is the price paid by the consumer before the tax is imposed, and in the post-tax equilibrium?

   b. What is the price received by producers?

   c. How much revenue is raised?

2. Consider a small town in which workers are highly mobile (i.e., they can be induced to leave the town if opportunities elsewhere improve slightly). What do you think the incidence of a tax on wages in that town would be, compared with the incidence in a town in which workers are immobile?

3. It is frequently asserted that taxes on cigarettes and beer are regressive, because poor individuals spend a larger fraction of their income on such items than do better-off individuals. How would your estimate of the degree of regressivity be affected if you thought these commodities were produced by:

   a. Competitive industries with inelastic supply schedules?

   b. Monopoly with a linear demand schedule?

   c. Monopoly facing a constant elasticity demand schedule?

4. It is often asserted that gasoline taxes used to finance highway construction and maintenance are fair because they make users of roads pay for them. Who do you think bears the burden of such taxes?

5. If you believed that a proportional consumption tax was the best tax, what are various ways in which you could levy it? Might there be differences in administrative costs associated with levying such a tax in different ways?

6. In what ways may the actual incidence of a government expenditure program differ from the legislated intent? Why might the effects be different in the short run than in the long run? Illustrate with examples drawn from Part Four of the book, or with a discussion of the effects of government farm programs. Similarly, discuss how the short-run and long-run effects of a regulatory program, such as rent control, may differ.

APPENDIX: COMPARISON OF THE EFFECTS OF AN AD VALOREM AND SPECIFIC COMMODITY TAX ON A MONOPOLIST

Suppose the government imposes a tax on the output of a monopolist. We asserted in the text that an ad valorem tax (a tax based on a fixed percentage of the value of sales) would reduce output less than a specific tax (a fixed tax on each unit sold) for any given revenue raised by the government.

The reason is that the ad valorem tax reduces marginal revenues by less than the tax, whereas the specific tax reduces marginal revenues by exactly the amount of the tax. Because a monopolist sets marginal revenue equal to marginal cost, if marginal revenue is reduced by less, output is reduced by less.

We can see this diagrammatically in Figure 18.14. Figure 18.14A illustrates the effect of a specific commodity tax. Earlier, we represented the effects of such a tax by an increase in the marginal cost. Alternatively, we can represent the effects of this tax as a *decrease* in the price received by the firm at any given quantity sold, that is, as a downward shift in the demand schedule. Both the demand and marginal revenue schedules shift down by the magnitude of the tax, t.

With an ad valorem tax, if an individual pays a price p for a commodity, the amount received by the producer is $p(1 - \hat{t})$, where $\hat{t}$ represents the ad valorem tax rate. Thus, the tax paid is a function of the market price. If the price were zero, there would be no tax paid, as we saw in the text. The effect of the tax is to rotate the demand curve as in Figure 18.14B, rather than to shift it down uniformly as in Figure 18.14A. The ad valorem tax at rate $\hat{t}$ reduces revenue by a fixed percentage—to $(1 - \hat{t})pQ$—and therefore lowers marginal revenue by the same percentage—to $(1 - \hat{t})\mathrm{MR}_{bt}$; that is, to $1 - \hat{t}$ times the before-tax level. The marginal revenue schedule, too, is rotated around the point at which it intersects the horizontal axis.

The important point is that the marginal revenue is reduced by $\hat{t} \times \mathrm{MR}$, and because marginal revenue is less than the price, it is reduced by less than $\hat{t} \times p$, the tax revenue per unit of the product sold. By contrast, with the specific tax, marginal revenue is decreased by precisely the amount of the specific tax. Thus, for any given level of equilibrium output—any given reduction in marginal revenue—the ad valorem tax raises more revenue, as shown in the figure; or, equivalently, for any given tax revenue per unit ($t = \hat{t} \times p$), output will be higher with an ad valorem tax, so price will be lower and total government revenue will be higher.
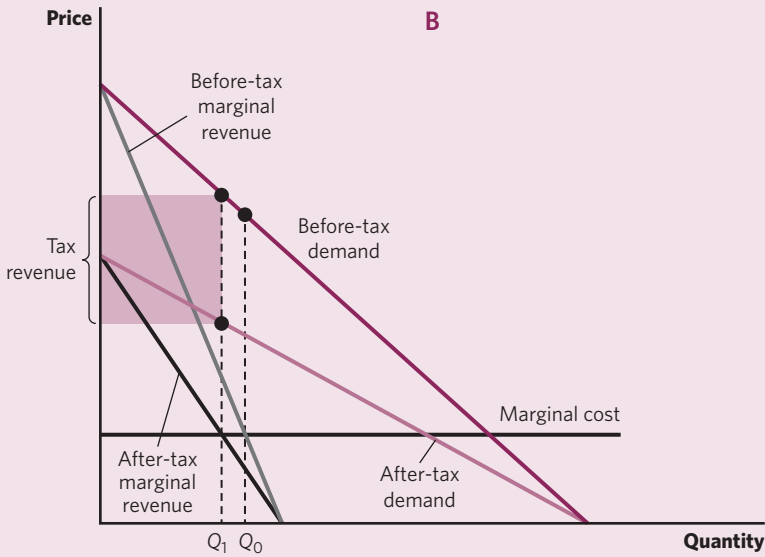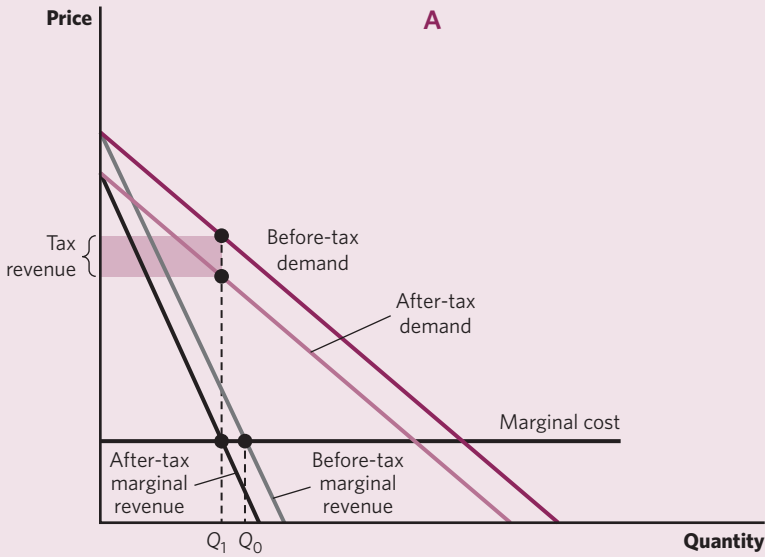
COMPARING THE EFFECTS OF A SPECIFIC TAX AND AN AD VALOREM TAX ON A MONOPOLIST

(A) The effects of a specific commodity tax on a monopolist can be viewed either as a shift upward in the marginal cost schedule (as in the earlier diagrams) or, as here, a shift downward in the demand and marginal revenue schedules. (B) Analysis of the effects of an ad valorem tax on a monopolist. For any given level of output, Q_1, tax revenue is higher with an ad valorem tax than with a specific tax.

19

TAXATION AND ECONOMIC EFFICIENCY

All taxes affect economic behavior. They transfer resources from individuals to the government. As a result, individuals must alter their behavior in some way. If they do not adjust the amount of work they do, they must reduce their consumption. They may work more, enjoying less leisure; by working more, they need reduce their consumption less.

No matter how individuals adjust, an increase in taxes must make them worse off.[1] However, some taxes reduce individuals' welfare less, for each dollar of revenue raised, than do other taxes. Tax policy is concerned with designing tax structures that minimize welfare loss for any given amount of revenue raised—while still attaining the other objectives of tax policy discussed in Chapter 17. This chapter analyzes the determinants of welfare loss; Chapter 20 then uses the results to describe the basic principles of optimal taxation.

This chapter is divided into six sections. The first analyzes the effects of a tax on a consumption good, such as beer. After describing the effects

---

[1] This ignores, of course, the benefits that may accrue from the increased government expenditures that result from the increased taxes. In a sense, this chapter looks at the "costs" of government programs, which are associated with the taxes to finance them, whereas earlier chapters in the book looked at the benefits. An overall assessment requires balancing the two. Throughout this chapter, we also ignore general equilibrium effects: before-tax wages and prices will be assumed to be unaffected by the imposition of a tax.

qualitatively, the second section shows how the distortions can be quantified. The third section analyzes inefficiencies associated with taxes on producers. The fourth and fifth sections show how the same principles may be applied to taxes on the return to savings and wages. The final section discusses various attempts to quantify the effects of taxation on labor supply.

EFFECT OF TAXES BORNE BY CONSUMERS

We begin the analysis with the simplest case: that of a tax borne fully by consumers. For example, let's assume that Crusoe's income is fixed, and he can choose between purchasing two commodities: soda and beer. His budget constraint is the line *SB* in Figure 19.1. This gives the various combinations of soda and beer that he can purchase. If he spent all his income on soda, he could purchase the amount *S*; if he spent all his income on beer, he could purchase the amount *B*.

Suppose that the government imposes a tax on beer. What will be the effect? (Throughout this section, we will assume that the consumer price rises by the full amount of the tax; that is, consumers bear the full burden of the tax. This will happen if the supply curves for beer and soda are infinitely elastic, as we showed in Chapter 18.) The tax on beer shifts the budget constraint in to *SB'*. Crusoe can still, if he wishes, spend all his income on soda, in which case he obtains *S* units of soda. Beer, however, is now more expensive, so he can purchase less of it with his income.
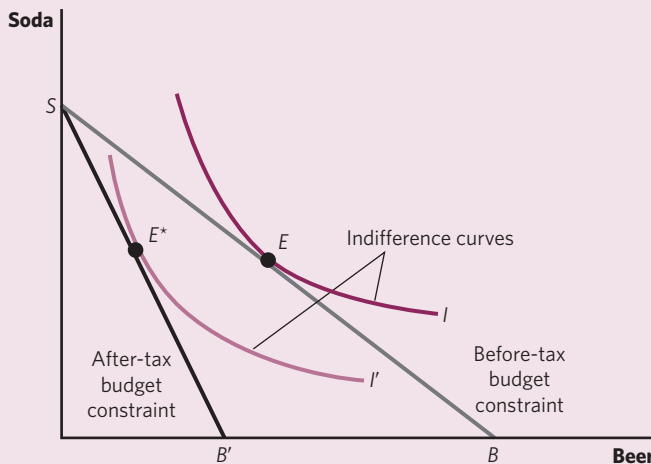
Initially, Crusoe allocated his income by choosing point *E* on this budget constraint. This is the point of tangency between the budget constraint and the indifference curve. After the imposition of the tax, there is a new equilibrium, at point E*. At *E**, Crusoe consumes less beer than at *E*.

SUBSTITUTION AND INCOME EFFECTS

The tax decreases an individual's consumption of beer, for two reasons. First, the tax—like any tax or loss of income—makes the individual worse off, by leaving him or her with less money to spend. Normally, when an individual is worse off, he or she consumes less of all goods. The amount by which the individual's consumption of the taxed good is reduced because he or she is worse off is called the **income effect** of the tax. Second, the tax makes beer more expensive than other goods. When a good becomes relatively more expensive, individuals find substitutes for it. The extent to which consumption of the taxed good is reduced because of the increased *relative price* is the **substitution effect**.

Figure 19.2 shows how to decompose the movement from *E* to *E**— the reduction in beer consumption—into income and substitution effects. We first ask, how would consumption of beer have been reduced if we had taken away income from the individual—to put him or her on the new, lower indifference curve—but, at the same time, had not changed relative prices? This change is reflected in the budget constraint $\hat{S}\hat{B}$, which is parallel to the original budget constraint (implying the same prices) but tangent to the indifference curve I', at $\hat{E}$. The corresponding reduction in beer consumption is the income effect.

The movement from $\hat{E}$ to *E**, and the corresponding reduction in beer consumption, is the substitution effect. It represents the reduction in consumption *due solely to changes in relative prices*.

Income and substitution effects work in the same direction in the case of a beer tax: beer consumption drops continually as we move from *E* to $\hat{E}$ to *E**.

DETERMINING THE SIZE OF THE SUBSTITUTION EFFECT The magnitude of the substitution effect depends on how easy it is to substitute other goods for the taxed good. This is reflected in the shape of the indifference curves. If they are relatively flat, then substitution is easy, and the substitution effect is large.[2] Figure 19.2B illustrates the extreme case in which indifference curves are L-shaped and there is no substitution effect.

---

[2] More precisely, it depends on the *elasticity of substitution,* which is defined as the percentage change in relative quantities consumed from a percentage change in relative prices. The L-shaped indifference curves in Figure 19.2B have a zero elasticity of substitution. The other extreme case is a straight-line indifference curve, in which case the elasticity of substitution is said to be infinite.
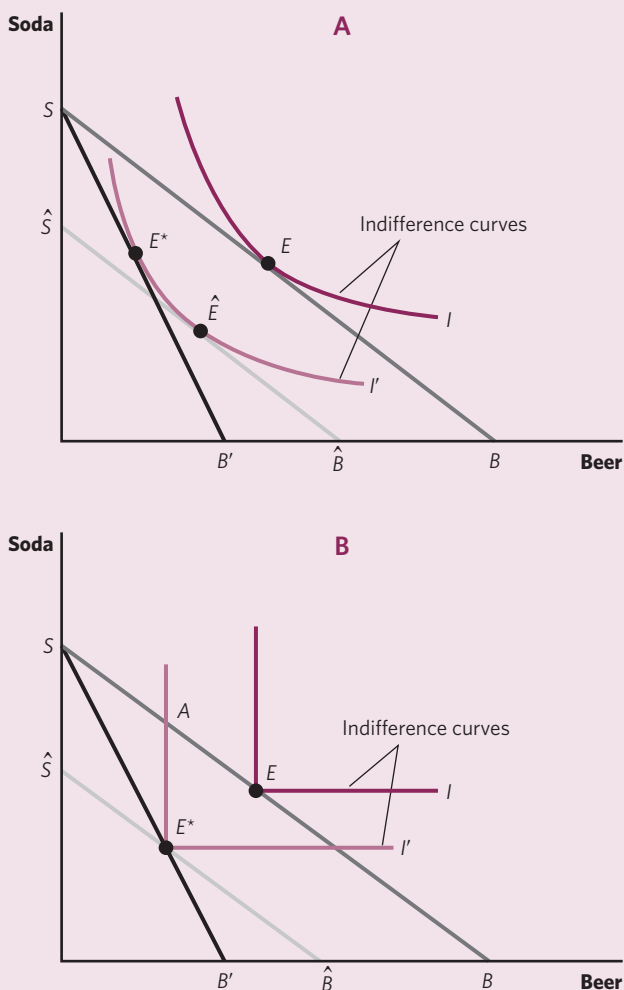
INCOME AND SUBSTITUTION EFFECTS OF A TAX ON BEER CONSUMPTION

(A) Decomposes the movement from E to E^* into income and substitution effects. The movement from E to $\hat{E}$ is the income effect, and the movement from $\hat{E}$ to E^* is the substitution effect. (B) Represents the case in which there is no substitution effect; indifference curves are L-shaped.

QUANTIFYING THE DISTORTIONS

Any tax must have effects on consumption. After all, the purpose of a tax is to transfer purchasing power from the individual to the government. Individuals must reduce their consumption of something. An efficient tax minimizes the welfare loss per unit revenue raised. Chapter 17 introduced the concept of a lump-sum tax, a tax that the individual must pay regardless of what he or she does. Such a tax simply moves the budget

FIGURE 19.3

LUMP-SUM TAX

The vertical distance between the two budget constraints measures the magnitude of the lump-sum tax.
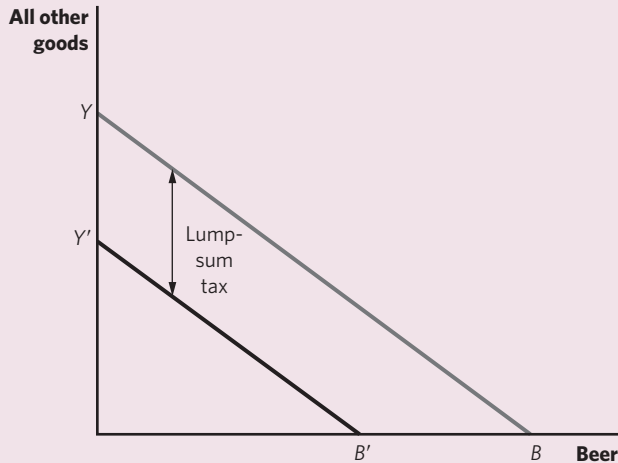
constraint in a parallel way, as illustrated in Figure 19.3. In the figure, we have put expenditures on beer on the horizontal axis and expenditures on all other goods on the vertical axis. Thus, point Y, at which the individual consumes no beer, measures his or her income before tax; point Y' measures his or her income after tax; and the vertical distance YY', measures the lump-sum tax. The budget constraint is

Expenditures on beer + expenditures on all other goods
= income − lump-sum taxes,

where expenditures on beer = $p_B B$, the price of beer times the quantity of beer purchased.

We compare the effect of any tax—such as a tax on beer—with the effect of a lump-sum tax by asking: For the same revenue, how much worse off are individuals with the tax on beer than they would have been with the lump-sum tax? The extra loss in welfare is called the *deadweight loss*. Equivalently, we can ask: For the same effect on individual welfare, how much *extra* revenue would a lump-sum tax have raised? How much *less* revenue does the beer tax raise? The difference in revenue is how we measure the deadweight loss of the tax.

MEASURING DEADWEIGHT LOSS USING INDIFFERENCE CURVES

Figure 19.4 contrasts the effect of a tax on beer with a lump-sum tax. The beer tax rotates the individual's budget constraint down, from YB to YB'. The income raised by the tax is the vertical distance between the
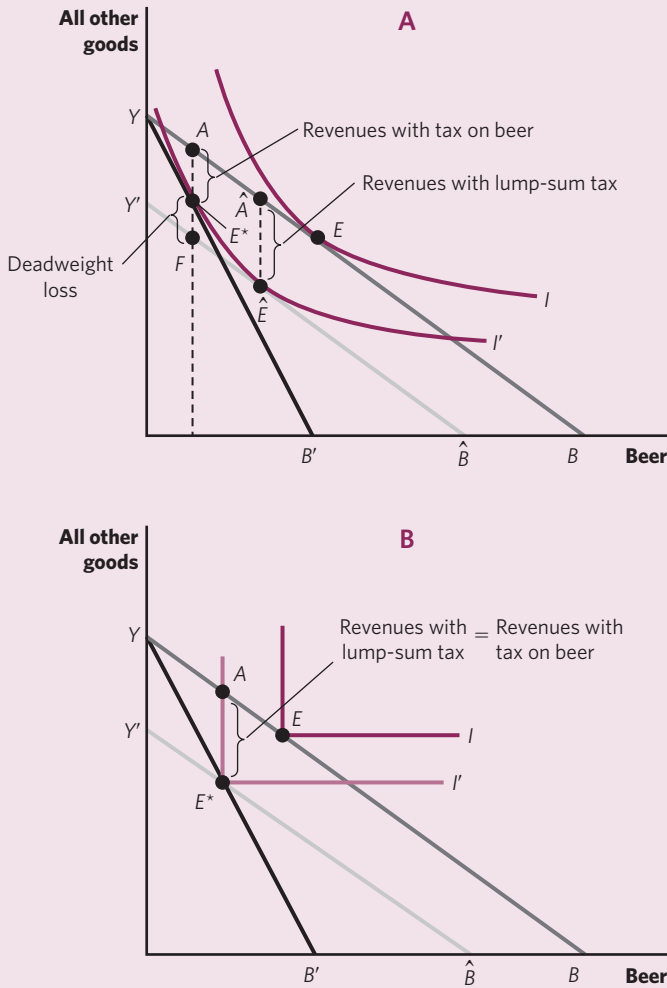
MEASURING THE DEADWEIGHT LOSS USING INDIFFERENCE CURVES
Individuals choose the amount of beer to consume by the tangency between their indifference curve and the budget constraint. The beer tax rotates the budget constraint. The lump-sum tax moves the budget constraint down parallel.
(A) The extra revenue raised by the lump-sum tax is E^*F.
(B) When there is no substitution effect, the beer tax has no deadweight loss; a lump-sum tax and a tax on beer raise the same revenue.

before-tax budget constraint and the after-tax budget constraint. Clearly, when no beer is consumed (point Y), no revenue is raised. The more beer that is consumed, the greater the tax revenue. The revenue raised is AE.*

The lump-sum tax with the same effect on utility moves the budget constraint from YB to $Y'\hat{B}$ and the equilibrium is now $\hat{E}$. The revenue raised is again the vertical difference between the new and the old budget constraints—this represents the amount of income that had to be taken away to leave the individual on the same indifference curve. Because the new and old budget constraints are parallel, the vertical distance $\hat{A}\hat{E}$ is exactly equal to AF. (The vertical distance between parallel lines is the same at any location.) Thus, *the lump-sum tax with the same effect on*

utility raises an additional revenue in the amount of E*F. E*F is the measure of the deadweight loss associated with the tax.

The *magnitude of the deadweight loss depends on the substitution effect*. This is illustrated in Figure 19.4B, which is identical to Figure 19.4A, except now the indifference curves are L-shaped, so there is no substitution effect, and it is apparent that there is no deadweight loss.

MEASURING DEADWEIGHT LOSS USING COMPENSATED DEMAND CURVES

Another way of measuring deadweight loss makes use of the concepts of consumer surplus and compensated demand curves introduced in Chapter 7. Assume we have imposed a tax of 30 cents per bottle of beer, and, with the tax, Crusoe consumes ten bottles a week. We ask him how much he would be willing to give to the government if the tax were eliminated. In other words, what lump-sum tax would leave him at the same utility level reached when he was subject to the 30-cent tax on beer? Clearly, Crusoe would be willing to pay at least 30 cents × 10 per week. Any extra revenue that such a tax would generate is the deadweight loss associated with the use of a distortionary tax system.

We now show how to calculate the deadweight loss using a consumer's *compensated demand curve*. The compensated demand curve gives Crusoe's demand for beer, assuming that as the price is lowered, income is being taken away from him in such a way as to leave him on the same indifference curve. We use the compensated demand curve because we wish to know how much more revenue we could have achieved with a nondistortionary tax, still leaving Crusoe just as well off as with the distortionary tax.

Assume that initially the price of a bottle of beer is $1.50, including the 30-cent tax, and Crusoe consumes ten bottles a week. We then ask him how much extra he would be willing to pay to consume eleven bottles a week. He is willing to pay only $1.40. The total amount that he would be willing to pay us as a lump-sum tax if we lowered the tax from 30 cents to 20 cents (and lowered the price of beer from $1.50 to $1.40) is 10 cents × the 10 bottles he previously purchased, or $1.00 (the area *FGCD* in Figure 19.5A).

We now ask Crusoe to assume that he is in a situation in which we levied a $1.00 lump-sum tax and charged $1.40 each for eleven bottles of beer. How much *extra* would he be willing to pay for one extra bottle? Assume that he said $1.30. We can now calculate the total lump-sum tax that he would be willing to pay if the price were reduced from $1.50 to $1.30. He would be willing to pay 20 cents a bottle for the first ten bottles (the area *JKCD*), and 10 cents for the next (the area *GKLH*), for a total of $2.10.
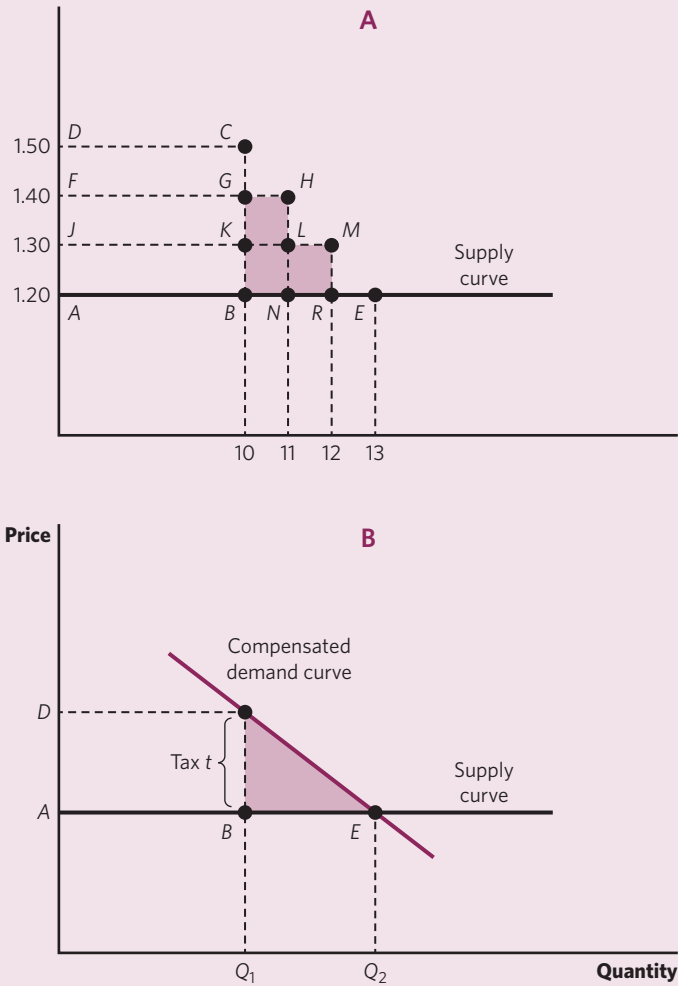
Finally, we ask him to assume that he is in a situation in which we levied a $2.10 lump-sum tax and charged $1.30 each for twelve bottles. How much extra would he be willing to pay for one extra bottle? Assume that he said $1.20. We could now calculate the total lump-sum tax that he would be willing to pay for the elimination of the 30-cent tax. He would be willing to pay 30 cents on the first ten bottles (the area *ABCD*), 20 cents on the next bottle (the area *BNHG*), and 10 cents on the twelfth bottle (the area *NRML*), for a total of $3.30. The tax revenue from the tax was $3.00 (the area *ABCD*). The deadweight loss is 30 cents (the shaded area).

FIGURE 19.6

EFFECT OF AN INCREASE IN TAX RATE ON THE DEADWEIGHT LOSS

A doubling of the tax rate more than doubles the deadweight loss. (The area *B'C'E* is four times the area *BCE*.)
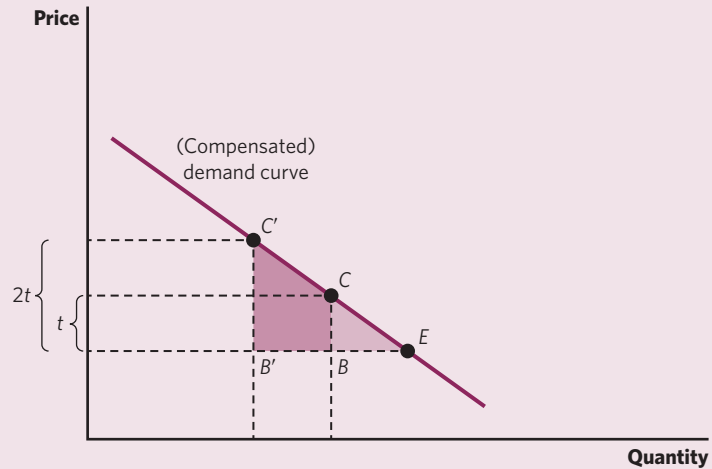
More generally, the amount that an individual would be willing to pay to have the price reduced by 1 cent is just 1 cent times the quantity consumed. As we lower the price, the quantity consumed increases. In Figure 19.5B, the total the individual would be willing to pay to have the price reduced from *D* to *A* is the area *AECD*, which takes account of the change in the after-tax quantity consumed as the price is reduced. Of that, however, *ABCD* is the tax revenue (the tax *AD*—which equals *BC*—times the quantity consumed, *AB*). Hence, the deadweight loss—the difference between the two—is just the triangle *BCE*. Figure 19.6 shows that as we double the tax rate, we more than double the deadweight loss.

Figure 19.7 shows that, for a given tax rate, the deadweight loss is greater the flatter—or, more precisely, the more elastic—the demand curve. (Remember that the elasticity of the demand curve gives the percentage change in demand as a result of a 1 percent change in price.)

We now make these insights more precise.

CALCULATING THE DEADWEIGHT LOSS

Return to Figure 19.5, in which we used compensated demand curves to measure deadweight loss. The height of the triangle, *BC*, is equal to the tax, *t*. *BE* is the change in quantity as a result of the tax. Recall that the elasticity of demand gives the percentage change in quantity as a result of a 1 percent change in price, that is,
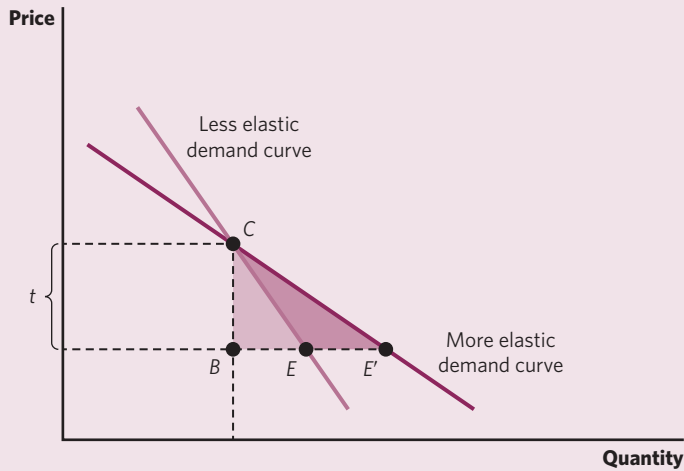
$$\eta = \frac{\Delta Q/Q}{\Delta p/p}$$

where the symbol ΔQ represents the change in quantity and the symbol Δp represents the change in price. (The symbol Δ is the [capital] Greek letter delta and is conventionally used to represent a change. The symbol η is the Greek letter eta and is conventionally used to represent the elasticity of demand.) Rearranging, we can write the change in quantity as

$$\Delta Q = \frac{\Delta p}{p}\, Q\eta.$$

This equation has the natural interpretation that the change in quantity will be larger, the larger the change in price and the larger the elasticity of demand. However, the change in price is just the per unit tax, t. Thus, substituting, we obtain

$$BE = \frac{t}{p}\, Q\eta.$$

Now the area of the triangle *BCE* is just

$$\frac{t \cdot BE}{2} = \frac{1}{2}\frac{t^2}{p}\, Q\eta$$

$$= \frac{1}{2}\left(\frac{t}{p}\right)\left(\frac{t}{p}\right)pQ\eta$$

$$= \frac{1}{2}\, \hat{t}^2\, pQ\eta$$

where $\hat{t} \equiv t/p$ is the tax *rate*, the ratio of the tax to the price.

<div style="border:2px solid; padding:1em;">

DEADWEIGHT LOSS OF A TAX

The deadweight loss of a tax increases with the magnitude of the substitution effect (or the elasticity of the compensated demand curve) and with the *square* of the tax rate.

</div>

DETERMINANTS OF DEADWEIGHT LOSS

The preceding formula identifies two of the primary determinants of deadweight loss. Deadweight loss increases with the *square* of the tax rate. High tax rates are far more distortionary than low tax rates.

Deadweight loss increases with the *elasticity of the compensated demand curve*.[3] The latter is precisely the substitution effect identified earlier as the critical determinant of deadweight loss. When indifference curves are very flat, the elasticity of the compensated demand curve is large—that is, a small percentage change in price leads to a large change in consumption. (Remember, the compensated demand curve simply describes a movement along an indifference curve, as, by definition, individuals are being compensated to keep them on the same indifference curve.) Many of the goods on which excise taxes are imposed have relatively low elasticities of demand, so the deadweight loss is relatively small. For instance, the 10 percent airline ticket tax is estimated to have a deadweight loss equal to 2.5 percent of the revenue raised (on the basis of an estimated 0.5 price elasticity of demand), an 8 percent beer tax generates a deadweight loss equal to 1.2 percent of the revenue raised (on the basis of an estimated price elasticity of 0.3), and a 15 percent cigarette tax is estimated to lead to a deadweight loss equal to 3 percent of the revenue raised (on the basis of a price elasticity of demand of 0.4).

EFFECT OF TAXES BORNE BY PRODUCERS

Up to now, this chapter has focused on the distortionary effects of a tax on a consumption good. We assumed that supply curves were horizontal, so the entire burden of the tax was on consumers.

At least in the short run, however, most supply curves are upward sloping. This means that part of the burden of any tax on a consumption good will fall on producers. Will this cause an excess burden on producers, above and beyond the direct burden of the tax revenue? The answer

---

[3] Recall from Chapter 7 that the compensated demand curve is closely related to the ordinary demand curve. When price rises, say, as a result of the tax, individuals are worse off. If the individual previously purchased 100 bottles a beer a year, a 10-cent price increase makes him or her worse off; if we gave the individual $10, he or she would be fully compensated. The effect of a compensated price increase is just the ordinary direct effect, plus the effect of giving an individual an extra $10. If the individual spends only 0.1 percent of his or her income on beer, then the extra income induces an additional beer expenditure of 10 cents: there is little difference between the impact of a compensated and an uncompensated change.

is yes, except in the special case in which the supply curve is vertical; that is, the elasticity of supply is zero.

Recall how a supply schedule (curve) is constructed. At each price, firms produce up to the point at which price equals marginal cost. If the supply schedule is upward sloping, the marginal cost rises as production rises. The area between the supply curve and price measures the **producer surplus**, which is just the difference between revenues and total variable costs. Changes in this area thus measure changes in profits.

Consider the example illustrated in Figure 19.8A. What happens to profits as price increases from 1 to 4 and output increases from 1 to 4?
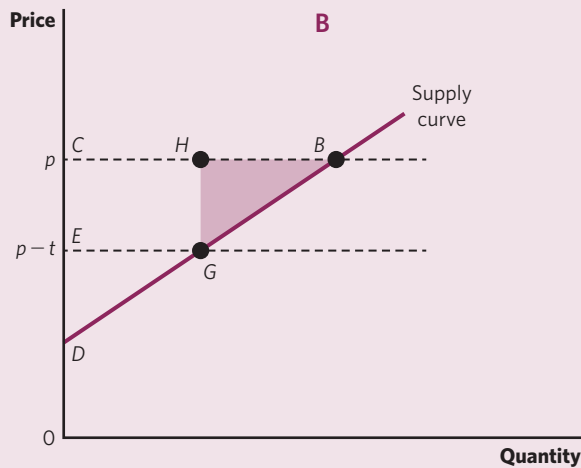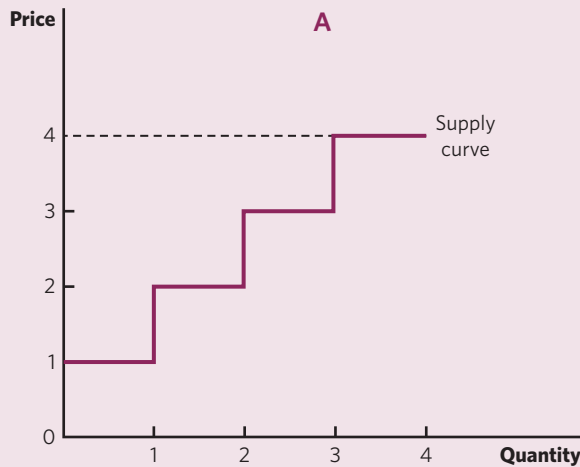
DEADWEIGHT LOSS OF A TAX ON PRODUCTION
BGH measures the deadweight loss of a tax on production.

The first unit of output costs $1; the next, $2; the third, $3; and the fourth, $4. If we pay the firm $4 for each unit, so it produces four units, the firm gets $3 more than marginal costs for producing the first unit, $2 more than marginal costs for producing the second unit, and $1 more than marginal costs for producing the third unit. The total profits are $3 + $2 + $1 = $6. Imposing a tax that lowers the price received by the producer to $3 lowers profit to $2 + $1 = $3. If the tax is $1 per unit, however, tax collection will be $2, so the deadweight loss is $1.

This can be seen more generally in Figure 19.8B. Assume initially that the producer is receiving the price p. Then a tax is imposed that lowers the amount the producer receives to $p - t$. In the initial situation, the total profits are given by the area DBC.[4] Now, the producer's profits are reduced to DGE. The change in the profits area is $EGBC$. Of this change, though, part accrues to the government as tax revenue—the rectangle $EGHC$. The tax on producers has resulted in producers' profits being reduced by more than government revenue has increased. The difference between the two is the deadweight loss associated with the tax—it is simply the shaded area BGH. To put it another way, the government could have imposed a lump-sum tax on the firm, which left price at p and which left the firm at the same level of profits as it had with a price of $p - t$. That lump-sum tax would have generated higher revenues, by the amount BGH, than the tax on the output of the firm.

It is clear that the steeper—the more inelastic—the supply schedule, the smaller the deadweight loss. In particular, we can, as before, show that for small taxes the deadweight loss increases with the square of the tax rate and with the supply elasticity.

A similar analysis applies to taxes on goods that are used in production. For instance, assume we had a tax on some input, such as steel, into an industry (automobiles). We can ask what lump-sum tax we could impose on the industry that would have the same effect on profits as the tax on steel.[5] The difference in revenues raised by the lump-sum tax and the tax on steel is the deadweight loss from the tax. The magnitude of the deadweight loss will depend on the possibilities of substitution. If the firm cannot substitute any other input for steel (even partially), the tax on steel is no different from a tax on output. There is no distortionary effect on the input mix and, hence, no deadweight loss associated with a change in the input mix.

---

[4] More accurately, the shaded area measures the difference between revenues and total variable costs. To calculate profits, we need to subtract fixed costs. (Fixed costs are costs that are incurred as long as the firm operates; they do not depend on the scale of production.)
[5] This is not, of course, the only deadweight loss arising from the input tax. Because it increases the marginal cost of production, the input tax will result in an increase in the price consumers pay, and there will be a deadweight loss to consumers.

EFFECTS OF TAXES BORNE PARTLY BY CONSUMERS, PARTLY BY PRODUCERS

It is straightforward to combine our analysis of producer deadweight loss with consumer deadweight loss. Figure 19.9 illustrates the case of a tax that is borne partly by producers (the price they receive falls from p to p_s) and partly by consumers (the price they pay rises from p to p_c). The change in market demand can be decomposed into two parts, just as before. The movement from Q to $\hat{Q}$ is the income effect of the tax; the movement from $\hat{Q}$ to Q^* is the substitution effect, as consumers substitute away from the taxed good along the compensated demand curve. That is, in the new equilibrium, at the price p_c, consumers are clearly worse off than they were at the original equilibrium price, p. If we ask how much they would have consumed, at the original (non–tax-distorted) price p, but at the new lower level of welfare, the answer is $\hat{Q}$, the point along the compensated demand schedule through A at the price p. The deadweight loss is associated with the movement along the compensated demand schedule, with the reduction of consumption from $\hat{Q}$ to Q^*, and is given by the triangle ABD.

What matters for producers, however, is the total change in quantity, from Q to Q^*, so their deadweight loss is the triangle BCE. The total deadweight loss is the sum of these two triangles, and depends, as before, on the elasticities of demand and supply.
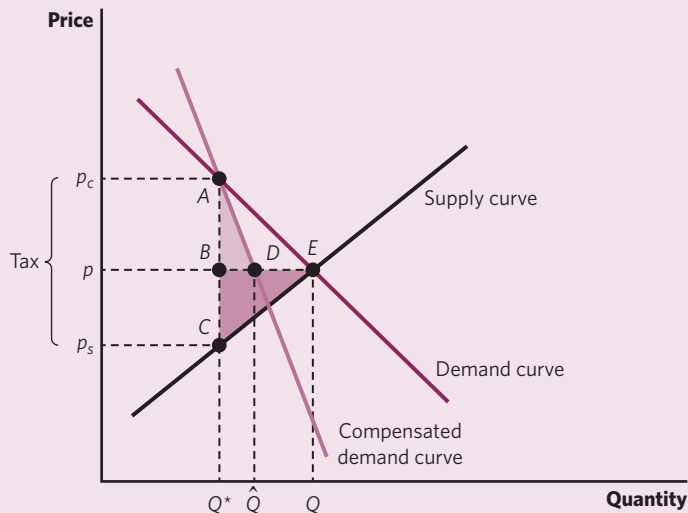
FIGURE 19.9

DEADWEIGHT LOSS FROM A TAX THAT IS BORNE PARTLY BY CONSUMERS AND PARTLY BY PRODUCERS

The consumers' deadweight loss is the triangle *ABD*; the producers' is the triangle *BCE*. If the compensated and uncompensated demand schedules coincide, as they will if the demand curve is not sensitive to small changes in income, then the total deadweight loss is the large triangle *ACE*.

TAXATION OF SAVINGS

The individual's allocation of income between consumption this period and consumption in the future is very much like his or her decision about allocating income between two different commodities.

By giving up one dollar of consumption today, this individual can obtain $(1 + r)$ of extra consumption dollars next period, where r is the interest rate. That is, if the individual saves the dollar and deposits it in a bank, he or she gets back at the end of the period the dollar plus the interest it has earned. Thus, $1/(1 + r)$ is the price of consumption tomorrow, relative to consumption today.

If the individual neither borrowed nor saved money, he or she would consume whatever the wages were in the two periods. We denote the wages in the initial period by w_0 and wages in the next by w_1. Suppose that w_0 and w_1 correspond to point W in Figure 19.10. By borrowing, the individual can consume more today, but at the expense of consuming less next period. By saving, the individual can consume more next period, but at the expense of consuming less this period.

The individual thus faces a budget constraint. He or she can either have $\overline{C}$ units of consumption today, or $(1 + r)\overline{C}$ units of consumption tomorrow, or any point on the straight line joining the two points, as depicted in Figure 19.10. The individual has a set of indifference curves between present consumption and future consumption, just as the individual has between beer and soda; each indifference curve gives the combinations of current and future consumption that leave him or her at the same level of utility. The individual is willing to consume less today in return for more future consumption. As the individual's present consumption gets smaller and smaller, he or she becomes less willing to give up more; and as the individual's future consumption gets larger and larger, the extra benefit he or she gets from each additional unit of future consumption gets smaller and smaller. Thus, the amount of increased consumption next period—required to compensate the individual for a reduction by one unit in current consumption—becomes larger and larger. That is why the indifference curve has the shape depicted. The individual chooses the point, denoted by E, on his or her budget constraint that is tangent to the indifference curve.

Figure 19.10A illustrates a situation in which the individual wishes to consume less than his or her wage income the first period, so he or she saves the rest; whereas in Figure 19.10B, the individual wishes to consume more than his or her wage income the first period, so he or she borrows the difference.

Consider now the effect of a tax, at the rate t, on interest income. (We assume that if interest income is negative—that is, the individual is a borrower—there is a negative tax; in other words, the borrower receives money from the government.) For a saver, someone whose first-period
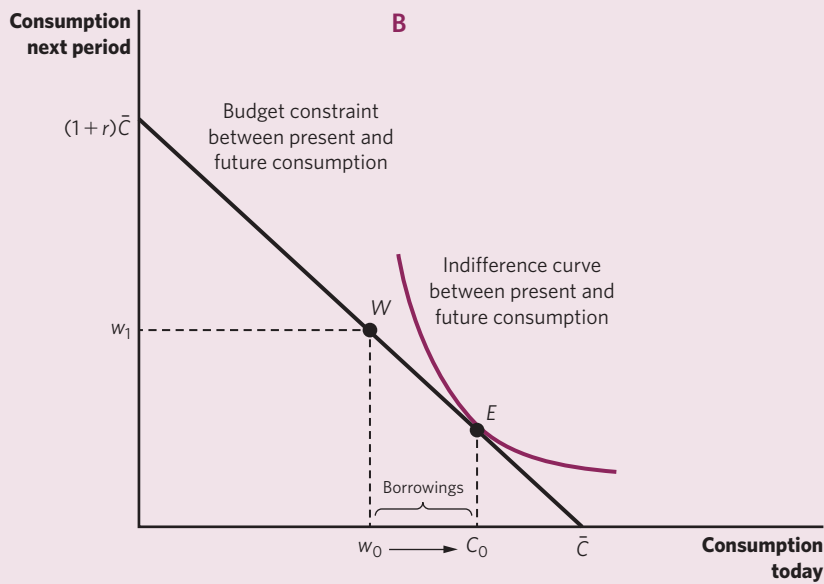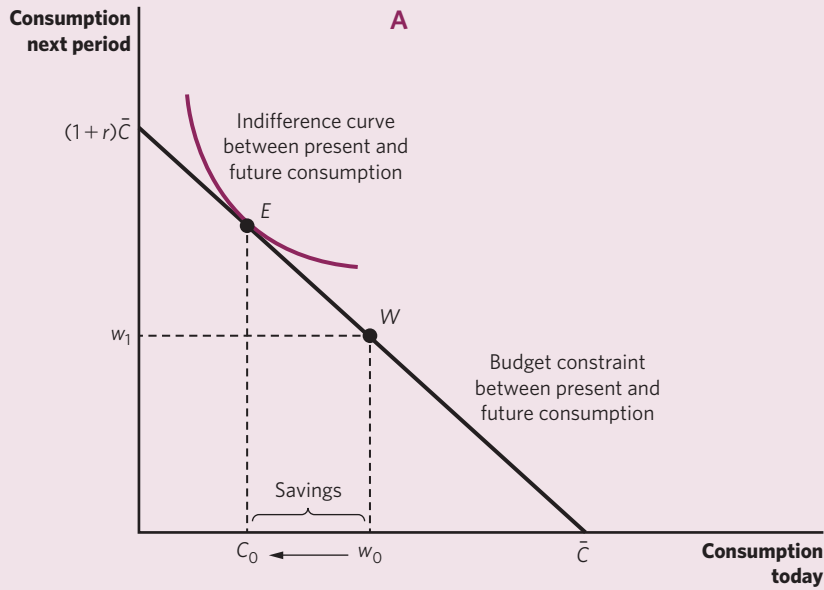
CONSUMPTION, SAVINGS, AND BORROWINGS

The individual allocates his or her income between consumption this period and next. (A) Individual saves. (B) Individual borrows.

A

Consumption next period

$(1+r)\bar{C}$

Indifference curve between present and future consumption

E

w_1

W

Budget constraint between present and future consumption

Savings

$C_0 \longleftarrow w_0$

$\bar{C}$

Consumption today

B

Consumption next period

Budget constraint between present and future consumption

$(1+r)\bar{C}$

Indifference curve between present and future consumption

w_1

W

E

Borrowings

$w_0 \longrightarrow C_0$

$\bar{C}$

Consumption today

consumption is less than his or her (after-tax) first-period wage income, the tax has both an income effect and a substitution effect. Because the individual is worse off, he or she normally will reduce consumption in both periods. Thus, the income effect leads to a lower current consumption. (Remember that savings is just the difference between the first period after-tax wage income and first-period consumption.) But because the individual receives a lower return from postponing consumption, the substitution effect discourages future consumption and encourages current consumption; it leads individuals to reduce their savings. The net effect on current consumption—and hence on savings—is ambiguous. If the substitution effect is large enough, savings are reduced.

If the substitution effect and the income effect were to cancel each other, leaving savings unchanged, would this imply that the tax is non-distortionary? No, because the tax is distortionary as long as it causes the individual to substitute between current and future consumption along his or her indifference curve.

Figure 19.11 depicts the case in which the income effect of the interest tax (the movement from E to $\hat{E}$) is just offset by the substitution effect (the movement from $\hat{E}$ to E^*). Therefore, savings, $w_0 - C_0$, are the same before and after the tax. Nonetheless, there is a substantial distortion in second-period consumption.

FIGURE 19.11

EFFECT OF INTEREST INCOME TAX

The income effect of the interest tax is just offset by the substitution effect in the first period, but there is still a deadweight loss of E^*F.
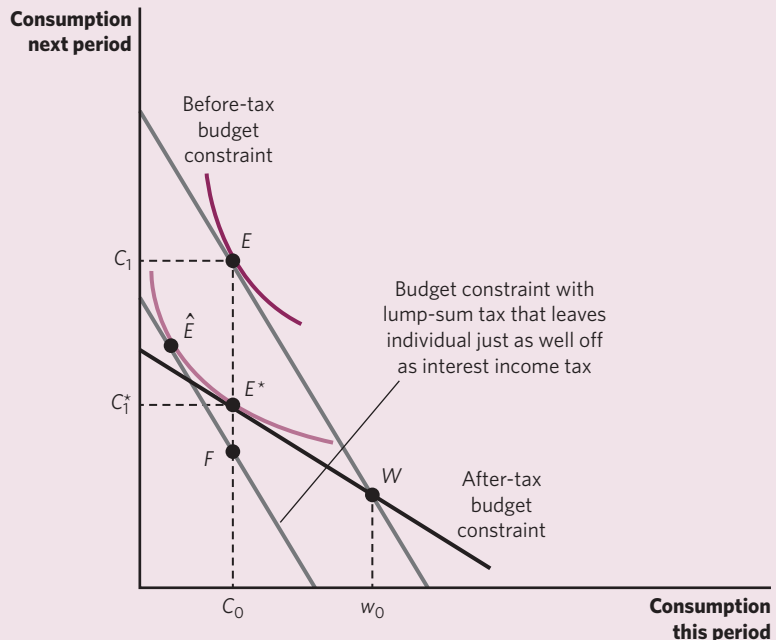
We could contrast the effect of the interest income tax with a lump-sum tax, a tax that shifted the budget constraint down in a parallel manner. Again, it is straightforward to show that such a tax will, for any given effect on the individual's utility, raise more revenue (the deadweight loss is measured by $E*F$), or that for any given level of revenue, individuals will be better off with the lump-sum tax than with the interest income tax. The magnitude of the distortion depends on the magnitude of the substitution effect, which in turn, depends on how substitutable current and future consumptions are.

QUANTIFYING THE EFFECTS OF AN INTEREST INCOME TAX

Most empirical estimates suggest that the substitution effect slightly outweighs the income effect, so an interest income tax has a slight negative effect on savings. Whereas from one perspective this is good news—the tax system may not be reducing savings by much—from another perspective it is bad news: government is unlikely to encourage savings by much through tax incentives.

The fact that the net effect is small does not, of course, mean that the distortionary effect is small. That depends on the magnitude of the substitution effect. However, as most Americans save a relatively small fraction of their incremental income, the income effect is relatively small, implying that the substitution effect is also relatively small (because the two cancel out).

As we noted, distortionary effects increase with the square of the tax. Because much of saving is done by the rich, who face relatively high taxes, the deadweight loss may still be significant, even if the elasticity of substitution is small.

The precise magnitude of the deadweight loss remains a subject of controversy, with some economists arguing that the deadweight loss is actually quite large.

TAXATION OF LABOR INCOME

Exactly the same kind of analysis can be applied to labor supply decisions. There are, of course, many dimensions to labor supply—number of hours worked, effort exerted on the job, years of education (age of entry into the labor force), and age of retirement. The same principles apply to each. Here, we illustrate the analysis for the decision concerning the number of hours worked. As with the savings decision, we can model the labor supply decision in terms of the choice between two commodities.

Here, the two commodities are leisure and all consumption of goods and services. Figure 19.12 shows the individual's budget constraint, with hours of leisure on the horizontal axis and consumption (income)[6] on the vertical axis. The wage tax, like the beer tax, rotates the budget constraint. If the individual does not work at all, he or she faces no tax—the individual still has sixteen hours of leisure a day (ignoring the eight hours of sleep). At a 50 percent wage tax, the individual's consumption is reduced by half, at any given level of work (leisure). Again, there is an income effect and a substitution effect. The substitution effect, as before, makes the individual work less (enjoy more leisure), but the income effect makes him or her work more: when the individual is poorer, he or she "consumes" less of all "goods," including leisure. The income and substitution effects work in opposite directions. Figure 19.12 illustrates a case in which the two effects are essentially offsetting; there is no effect on hours worked.

The fact that the labor supply curve is relatively inelastic—that income and substitution effects are offsetting—does not mean that the income tax is not distortionary; it is, as long as there is a substitution effect. Indeed, Figure 19.13 shows a case in which the income effect outweighs the substitution effect, so the labor supply curve is backward bending (at lower wages, individuals actually supply more labor). A tax in that case actually increases the labor supply. Nonetheless, because there is a substitution effect, the tax is distortionary—that is, there is a deadweight loss associated with the tax.

---

[6] For purposes of this section we assume there is no savings, so consumption and income are identical.

EFFECTS OF PROGRESSIVE TAXATION

So far in this chapter, we have focused on *proportional taxes.* The beer tax was a per unit tax, so that as consumption of beer increased, tax payments increased proportionally. The effect was similar for the interest income tax. Some wage taxes, such as the Social Security payroll tax, are proportional (up to some maximum), but other taxes, such as the individual income tax, are not.

EARNED INCOME TAX CREDIT   Figure 19.14 shows the effect of taxes on the budget constraint facing a low-wage individual who receives a *wage subsidy* (under the earned income tax credit) up to some level, and faces a tax beyond a certain (higher) level.[7] In the interval *LA*, the new budget constraint is actually steeper than the before-tax budget constraint *LL*; in the interval *AB*, the new and old budget constraints are parallel; and in the interval *BC*, the after-tax budget constraint is much flatter. For an individual who works little (chooses a point in the interval *LA*), the income and substitution effects work in opposite directions, with the income effect leading to less work (as the individual is better off) and the substitution effect to more work (as the return to working has increased). For an individual who works a moderate amount and chooses a point in

---

[7] The analysis simplifies the full complexity of the tax law by ignoring, for instance, both state and Social Security taxes.

FIGURE 19.14

BUDGET CONSTRAINT FOR LOW-INCOME INDIVIDUALS: THE EFFECT OF THE EARNED INCOME TAX CREDIT

Low-wage individuals who work little face positive incentives to work from the earned income tax credit. The tax credit affects both hours worked and labor force participation.
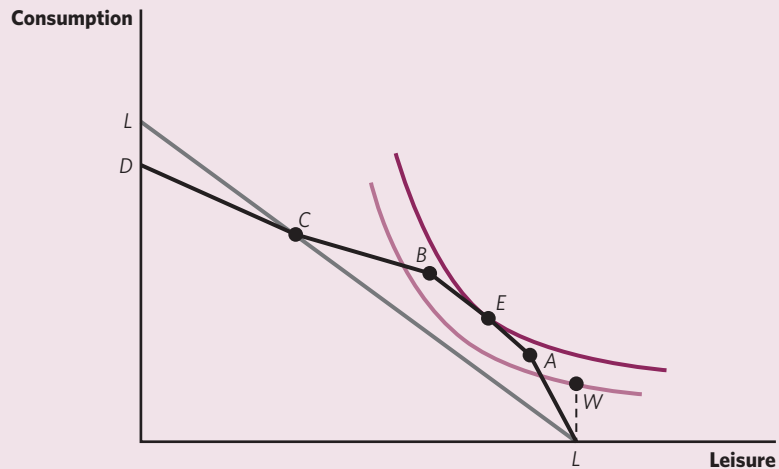
the interval AB, there is only an income effect: the individual unambiguously works less than before the subsidy. For the hardworking individual, in the interval BC, the income and substitution effects are reinforcing: the individual is better off, and the return to working is lowered, so the reduction in work effort is even greater. Finally, in the interval CD, the individual's EITC benefit is completely exhausted and he or she now faces an income tax. Thus, the individual is worse off than he or she was in the absence of any tax/subsidy program. Now the income and substitution effects work in opposite directions. (Point C, at which EITC is fully phased out, can occur to the left or the right of the before-tax budget constraint.)

PARTICIPATION VERSUS HOURS DECISIONS   Even though for many individuals (those in the interval BC), marginal incentives to work are reduced by the EITC, *incentives to participate in the labor force are increased*. Assume, as an alternative to working, individuals can receive a fixed amount of welfare payments, represented in the figure by point W (at which they enjoy full leisure). We have drawn the indifference curve through W.[8] It is clear that, with the earned income tax credit, the individual who would have been content simply to receive welfare now prefers to work.

Most economists believe that the participation decision is far more important than the hours decision. First, attachment to the labor force brings with it education, skills, and a sense of belonging to society, which contribute to social stability. Second, for many jobs, discretion over the number of

---

[8] In practice, under welfare, individuals do not lose all benefits if they work, but rather face a high marginal tax rate. There is a flat budget constraint through W: we assume it is sufficiently flat that the individual chooses W.
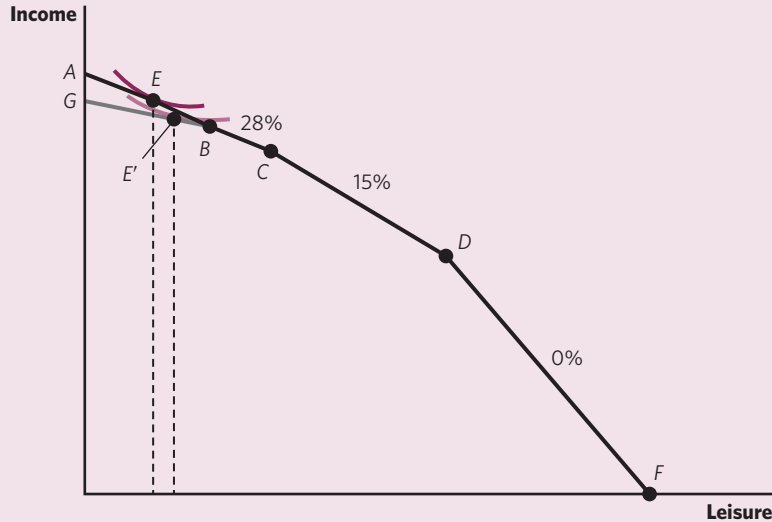
IMPACT OF 1993 TAX CHANGE

In 1993, the tax rate on upper-income individuals was increased. The curve *ABCDF* depicts schematically the original budget constraint, with three marginal tax rates at 0, 15, and 28 percent. After the change, there is a fourth segment, *BG*, at a 36 percent tax rate. Originally, the high-income individual chooses point *E*. Afterward, he or she chooses *E'*. The income effect is small, as the individual has to pay the higher tax only on the excess of income over a very high threshold. The substitution effect is large relative to the income effect, so the individual works less (enjoys more leisure).

hours is far more limited than the discussion we have presented suggests.[9] There is either a workweek of five eight-hour-days, or, say, a workweek of five seven-hour-days. To be sure, over time, the number of hours worked does adjust in response to economic forces. The average number of hours worked declined markedly during the first half of the twentieth century, whereas more recently, as wages (particularly at the lower end of the income distribution) have failed to grow or even declined, hours worked have increased; the individual, in deciding to work, is not usually in a position to bargain about whether he or she should work thirty-five hours or thirty-seven hours.

HIGH-INCOME INDIVIDUALS   Figure 19.15 analyzes the effect of taxation on high-income individuals in the 36 percent tax bracket. Someone in the 36 percent tax bracket faces a budget constraint with four segments, representing the 0, 15, 28, and 36 percent brackets. Assume initially that the top bracket was 28 percent. An increase in the top bracket is represented by a downward rotation of the budget constraint at point *B*. Because of the kink in the budget constraint, many individuals—with different preferences between leisure and work—may be clustered at the level of income associated with the kink.

---

[9] This observation is consistent with a study of high-income physicians in Britain, which found that whereas self-employed physicians exhibited considerable sensitivity to the marginal tax rate, those who were employees had no discernible sensitivity. See M. H. Showalter and N. K. Thurston, "Taxes and Labor Supply of High-Income Physicians," *Journal of Public Economics* 66 (1997): 73–97.

THE 1993, 2001, AND 2003 TAX REFORMS

In 1993, as the government faced ever-increasing deficits, some tax increases appeared imperative. The Clinton administration proposed that those who had benefited most from the economic expansion and the tax cuts of the 1980s should bear the brunt of these tax increases. Only the top 1.2 percent of taxpayers experienced rate increases. For example, married couples with incomes in excess of $140,000 had their income tax rates increased, from a marginal tax rate of approximately 28 to 36 or 39.6 percent.*

Critics, such as Professor Martin Feldstein of Harvard University, who was chairman of the Council of Economic Advisers under the Reagan administration, predicted that the tax increase would raise less revenue than both Congress and the administration had estimated, because of large responses in labor supply. As was illustrated in Figure 19.15, many of those facing tax increases were clustered near point B, the income level at which taxes were increased. For these individuals, as we have noted, there was little income effect from the tax but a large substitution effect; therefore, Feldstein argued, there would be a large labor supply response. To support his theoretical arguments, he cited evidence that after the 1986 tax reform, which had reduced marginal rates for upper-income individuals, there had been a marked increase in income tax collections from upper-income individuals.

As is so often the case, though, reading the evidence is not easy. The 1986 tax reform also closed a number of loopholes that had allowed upper-income individuals to avoid taxation, thereby broadening the tax base; and in the expansion of the economy that had begun in 1983, as the economy recovered from its worst recession since World War II, there was a strong trend of increased inequality, with earnings at the top growing far more rapidly than earnings in the middle. There was little direct evidence that upper-income individuals had either worked longer hours or worked harder as a result of the decrease in marginal tax rates.

As it turned out, the tax revenues raised on upper-income individuals in the years following 1993 were far higher than had been anticipated, and, indeed, these increased revenues were largely responsible for the elimination of the deficit in the late 1990s. Furthermore, there was little evidence that the upper-income individuals had reduced their labor supply in the way Feldstein had predicted.

In 2001 and 2003, President George W. Bush lowered marginal tax rates and reduced taxes on investment income from dividends and capital gains, as well as narrowed the tax base by phasing out limits on itemized deductions and personal exemptions for high-income taxpayers. The combined effect of these measures was to decrease both tax revenue relative to GDP and tax progressivity: federal receipts as a share of GDP, which had risen from 17.5 percent in 1993 to 20.6 percent in 2000, fell to 18.5 percent in 2007 (before the Great Recession helped to lower the number further to 15.1 percent by 2009); extension of the Bush tax cuts to 2011 is estimated to result in an average tax reduction of $45 for taxpayers in the bottom 20 percent, but an average tax savings of $63,000 for those with incomes in the top 1 percent.

---

*This ignores the Medicare tax, which was also raised.

For those near the kink, the income effect is small relative to the substitution effect; hence, their labor supply would have been expected to decrease, as depicted. (If the tax rate on all income had increased by the same percentage points, there would have been a large income effect; for someone with an income just above the level at which the higher rate set in, though, the income effect was, in fact, negligible.)

SECONDARY LABOR FORCE PARTICIPANTS

Historically, in two-earner families with primary and secondary earners, the labor supply response of the secondary earner (typically, the wife) has been markedly different from that of the primary earner. There are obvious reasons for this, particularly during the years in which there are small children at home. The *net* income—after subtracting out the costs of child care, which would not have to be paid if she did not work, costs of commuting, and so forth—may be far lower than the gross income. Thus, if in 2011 a woman earned $25,000 a year, and child care expenses amounted to $10,000, her net income was just $15,000. The tax rate she faced was determined by *household* income, including that of her husband. If her husband earned more than $69,000, she faced a 25 percent (or higher) income tax on all her income. (This ignores the payroll tax and state taxes.) A 25 percent tax on her total income, however, translates into a 42 percent tax on her net income. Her incentives to participate in the labor force are thus greatly reduced. Secondary workers have, accordingly, shown much more sensitivity in their labor force participation to changes in tax rates.

MEASURING THE EFFECTS OF TAXES ON LABOR SUPPLIED

The fact that, theoretically, the effect of a tax on wages is indeterminate makes it all the more important to attempt to determine empirically what its effects in fact have been. Research in this area has been extensive and has yielded important (but controversial) results. Two main methods have been employed to study these questions: statistical analyses using market data, and experiments.[10]

---

[10] Earlier, several studies approached the problem using *qualitative* approaches, simply asking individuals whether taxes led them to work more or less. The different responses reflected the presence of income effects (leading to more work) and substitution effects (leading to less work). See D. M. Holland, "The Effect of Taxation on Effort: Some Results for Business Executives," *National Tax Association Proceedings of the Sixty-Second Annual Conference* (1969); and G. Break, "Income Taxes and Incentives to Work: An Empirical Study," *American Economic Review* 47 (1957): 529–549.

STATISTICAL TECHNIQUES USING MARKET DATA

The first method entails using statistical techniques to analyze how individuals in the past have responded to changes in their after-tax wages. In general, we do not have data on how particular individuals responded to changes in wages. Rather, we have data on how many hours individuals who earn different wages work. Those who earn higher wages seem to work more hours. From this we can calculate the "average" effect of wages on hours worked.

Up to now, we have simply described a *correlation*, an observed relationship between two economic variables. We now wish to use this to make an *inference*, a prediction or a statement about the effect of lowering take-home wages resulting from, say, the imposition of a tax. To make such an inference, we must make an assumption; for example, that the reason individuals who receive higher wages are observed to work more is that they choose to work more because of the higher wage. In other words, that an individual who receives a higher wage is essentially like one who receives a lower wage; the only important difference is the difference in pay, and it is this difference that leads to a difference in the number of hours worked. There are, of course, other important differences, such as age, occupation, or gender, and more sophisticated statistical analyses attempt to take as many of these as possible into account; they attempt to see, among individuals of the same age, occupation, or gender (or who have other characteristics in common), whether those who receive higher wages work more.

The vast literature on labor supply suggests that estimated labor supply elasticities may depend on the precise statistical methods used as well as on the data employed.[11] There appears to be widespread agreement (though not unanimity) on the following:

- The labor supply of married men is fairly unresponsive to changes in the wage rate.[12]
- The compensated labor supply elasticity also appears to be small, although there appears to be more disagreement over this finding. (The uncompensated labor supply elasticity can be small either because of

---

[11] For instance, Jim Heckman of the University of Chicago has argued that reporting errors may have obscured a larger decline in hours worked by men, and this may also account for much of the observed decline in real wages. This measurement error may also bias labor supply elasticities toward zero. See J. J. Heckman, "What Has Been Learned about Labor Supply in the Past Twenty Years?" *American Economic Association Papers and Proceedings* (1993): 116–121.

[12] For instance, one study estimated that the tax system (as of 1983) had resulted in only a 2.6 percent reduction in hours worked by married men. See R. K. Triest, "The Effect of Income Taxation on Labor Supply in the United States," *Journal of Human Resources* (Summer 1990): 491–516. An earlier study using similar data and model had suggested a somewhat larger effect. See J. Hausman, "Labor Supply," in *How Taxes Affect Economic Behavior,* ed. H. J. Aaron and J. Pechman (Washington, DC: Brookings Institution, 1981), pp. 27–72.

small offsetting substitution and income effects or because of large off-setting substitution and income effects.)[13]

- The labor supply elasticity of married women is larger, but more problematic to estimate.[14]
- The labor supply elasticity of female heads of household is somewhere between that of married men and married women.
- The participation decision is more sensitive to the wage rate than are marginal hours of work.[15]
- Labor supply parameters estimated from market data tend to be larger than those estimated from experimental data (discussed in the next section).

Large changes in tax rates, of the kind that occurred in 1981, 1986, 1993, 2001, and 2003, provide natural tests of the effect of taxation on labor supply.

Each of these changes was complex; different individuals faced different changes in tax rates, and a variety of loopholes were closed, or, in some cases, opened up. For example, although the evidence is that the 1986 tax law led to an increase in taxable income faster than would have been predicted based on the assumption of fixed incomes,[16] there is controversy about the reason. Did the changes following the 1986 tax reform, for instance, reflect mostly the closing of loopholes, or did they largely reflect underlying trends? Increasing inequality was placing a larger fraction of the nation's income in the hands of those facing higher tax rates.

Barry Bosworth and Gary Burtless of the Brookings Institution argued that the observed changes had little to do with changes in taxes, but mostly reflected underlying economic trends.[17] Although it appears that after the tax cuts, labor supply was higher than it would have been had past trends continued (by 1989, for men, work effort was 6 percent above what it would have been had the 1967–1980 trend continued, and for women it

---

[13] Jerry Hausman of MIT, using a technique meant to capture the effects of the nonlinearities in the budget constraints described earlier, obtains much larger estimates than do others. See his "Labor Supply" (note 12). His techniques have been criticized by Tom MaCurdy of Stanford, who claims that his technique "forces higher estimates of substitution effects or lower estimates of income effects than are obtained from other procedures. This ... raises serious questions about the reliability of evidence cited by much of the literature to support tax reforms aimed at lowering marginal tax rates." See T. MaCurdy, "Work Disincentive Effects of Taxes: A Reexamination of Some Evidence," *American Economic Association Papers and Proceedings* (May 1992): 243–249.

[14] See Thomas A. Mroz's widely cited study, "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica* (July 1987): 765–799. After noting that the estimated elasticity is sensitive to the specification of the model, Mroz concludes that when the correct specification is used, "factors such as wage rates, taxes and nonlabor incomes have a small impact on the labor supply behavior of working married women" (p. 795). His analysis focused only on working women, ignoring the participation decision.

[15] This could be because of either the constraints imposed by employer or the fixed costs of working. Robert Triest estimates that, despite the small responsiveness of hours worked, participation is so responsive that the U.S. tax system (state and federal) reduced total hours worked by wives by as much as 30 percent, depending on the specification employed. See "The Effect of Income Taxation on Labor Supply in the United States" (note 12).

[16] See M. Feldstein, "The Effect of Marginal Tax Rates on Table Income: A Panel Study of the 1986 Tax Reform Act," *Journal of Political Economy* 103, no. 3 (1995): 551–572.

[17] B. Bosworth and G. Burtless, "Effects of Tax Reform on Labor Supply, Investment, and Saving," *Journal of Economic Perspectives* (Winter 1992): 3–25.

was 5.4 percent higher), the interpretation of what went on is not so clear. If tax reform was the impetus for the growth in labor supply, we would expect higher-income workers (who experienced a greater change in marginal tax rates) to have increased their labor supply by proportionately more than lower-income workers (who were less affected by tax reform). However, it was the lower-income individuals whose labor supply seems to have increased the most.[18] The fact that income tax revenues among the very rich continued to increase so robustly after 1993, in spite of the large increase in marginal tax rates, is consistent with the hypothesis that it was long-term trends and the closing of loopholes, not incentive effects on labor supply, that accounted for the increased tax revenues after the 1986 reform.

EXPERIMENTS

The second approach to obtaining a quantitative estimate of the magnitude of the labor supply responses to tax changes is an experimental one. We are interested in the question: What would happen to the labor supply if we raised or lowered tax rates (or changed the tax structure in some other way)? One approach is to say, "Let's change the tax structure and see what happens." This could be an expensive approach: the change might have a very negative effect on labor supply, but before the effects were recognized and the tax structure changed again, considerable damage (welfare loss) could occur.

However, we can learn something by changing the tax structure for just a small portion of the population. Just as opinion polls can give fairly accurate estimates of how voters will vote in an election simply by asking a small sample of the population (often fewer than 1000 individuals), so, too, the response of a small sample may give a fairly reliable indication of how other, similar individuals would respond facing the alternative tax structure. Opinion polls are careful to obtain a representative sample of views. They make sure that views of young and old, rich and poor, married and unmarried, skilled and unskilled workers, and so on are all represented; in forming their estimate of how the population as a whole will vote, they weight the relative importance of the various groups in the population. (When they are attempting to predict the outcome of elections, they assign weights corresponding to the known likelihood that members of different groups vote.)

Between 1968 and 1982, a series of such experiments attempted to ascertain, in particular, the effects of changes in the tax structure and welfare system on the labor supplied by poorer individuals. Different individuals were confronted with different levels of guaranteed income and

---

[18] Men in the lowest quintile increased hours worked by 31 percent, whereas those in the highest quintile increased hours worked by 3.2 percent; the corresponding numbers for women were 16.7 and 11.8 percent (Source: Bosworth and Burtless, "Effects of Tax Reform," note 17).

tax structures, making it possible, in principle, not only to estimate the overall effect of tax changes but to separate out the income effects from the substitution effects. The results were consistent with the view that the overall effect of taxes on labor supply is relatively small. The report on the first such experiment, conducted in New Jersey, described it as presenting "a picture of generally small absolute labor supply differentials between" those who were confronted with the alternative tax/welfare structures and those who faced the existing tax/welfare structure. "Only among wives, whose mean labor supply is quite small to begin with, are the differentials large in relative terms."[19] (In subsequent years, labor participation of wives has increased enormously, so the aggregate effect of such adverse incentives is now far more significant.) The experiments yielded some further results concerning the possible effects of changes in the welfare/tax system. Providing more income to the poor resulted in their searching longer for a job when they became unemployed.

Whereas the early experiments focused on the effect of alternative tax-subsidy schemes on labor supply, and related variables like job search, later studies attempted to ascertain whether there were other effects as well. For instance, an experiment in Gary, Indiana, found a higher birth weight of babies—an indication of the health of the child—in families whose income had been increased. A large-scale experiment sponsored by the U.S. Department of Health, Education, and Welfare and conducted in Seattle and Denver, found that providing women with a guaranteed income, as the negative income tax does, might contribute to the breakup of families. However, the most generous negative income tax programs in the Seattle–Denver experiment had the *least* effect on family dissolution rates. It has been argued that income guarantees have two opposing effects on dissolution rates: on the one hand, they stabilize marriages by improving the family's ability to buy essential goods and services; on the other hand, they destabilize marriages by improving the economic viability of alternatives to marriage. Under this theory, the experimental results suggest that for low guaranteed income levels, the second effect (the "independence effect") dominates the first.

The experiments represented an important advance in the tools that are available to social scientists. At the same time, some important limitations to the experimental approach need to be borne in mind when evaluating the results.

First, there is a well-known phenomenon called the **Hawthorne effect**, which plagues all experimental work with individuals: when an individual is included in an experiment and knows his or her behavior is being examined, the behavior is often altered.

---

[19] U.S. Department of Health, Education, and Welfare, *Summary Report: New Jersey Graduated Work Incentive Experiments* (Washington, DC: Government Printing Office, 1973).

Second, there are problems associated with ensuring that the sample is representative. Because participation in the experiment is voluntary, there may be systematic biases associated with the kinds of individuals who refuse to participate.

Third, the response of individuals to short-run changes may differ from their responses to long-run changes. On the one hand, a temporary change in the tax structure that leads them to be better off has a smaller effect on lifetime income than a permanent change in the tax structure; hence, the income effect may be understated. On the other hand, because the experiment discussed earlier often involved individuals facing a higher or lower marginal tax rate during the course of the experiment, the after-tax wage was temporarily reduced or increased; a temporary reduction in the wage may have different effects than a permanent reduction. In the absence of costs of adjustment, there is a presumption that individuals will reduce their work (increase their leisure) more than they would with a permanent wage reduction. Thus, an individual who was planning to take some time off from work (say, a woman thinking of having children in the near future) might have taken advantage of the temporary availability of a large subsidy combined with a high marginal tax rate. If this is true, the experiments overstate the effects relative to what they would be with a permanent change. On the other hand, costs of adjustment may be very high; an individual might be reluctant to quit his or her current job, knowing that he or she will want it back in three years' time when the experiment is over, believing that it will be difficult to get it back then. If these effects are important, the experiment may have understated not only the income effects, but the substitution effects as well. Some of the more recent experiments have attempted to ascertain the magnitude of the biases in the estimates resulting from the fact that the change in tax structure/welfare payments was only temporary, by guaranteeing to the individual the same tax structure/welfare structure over a more extended period (up to twenty years).

A final important qualification on interpreting how accurately the experiments describe the extent to which labor supply is affected by changes in tax laws or welfare programs relates to the role of institutions in determining the length of the workweek. We commented earlier that, in the short run, institutional practices play an important role in restricting individuals' choices over the number of hours worked. But in the long run, these institutional practices themselves change, partly in response to changes in the economic environment. Thus, many of the individuals in the experiment may have had only limited discretion over the number of hours they worked; however, if everyone in society were confronted with the new tax/welfare payments structure, pressures might develop to alter these institutional practices to bring them more into conformity with individuals' preferences.

The high cost and ambiguous results of such experiments have meant that there have been few experiments of the scale and scope of the earlier studies. On the other hand, more care is placed in the design of pilot programs, so stronger inferences can be made concerning what works and what does not work. There have been, for instance, a large number of studies of training programs and of programs designed to move people from welfare to work. Still, primary reliance must be placed on "natural experiments," the experiments that occur as a result of, for example, different states' trying different programs. For instance, before the 1996 welfare reform, several states had experimented with time-limited welfare programs and welfare programs with work requirements. Such experiments suggested that the welfare reform would result in significantly reduced welfare dependency, a prediction borne out in the months after passage of the legislation—with the reduction in welfare roles far greater than could be explained by the declining unemployment rates.[20]

---

[20] On negative income tax experiments, see P. K. Robins, "A Comparison of the Labor Supply Findings from the Four Negative Income Tax Experiments," *Journal of Human Resources* 20, no. 4 (Fall 1985): 567–582; D. Greenberg and H. Halsey, "Systematic Misreporting and Effects of Income Maintenance Experiments on Work Effort: Evidence from the Seattle–Denver Experiments," *Journal of Labor Economics* 1, no. 4 (October 1983): 380–407; and R. G. Spiegelman and K. E. Yaeger, "The Seattle and Denver Income Maintenance Experiments: Overview," *Journal of Human Resources* 15, no. 4 (Fall 1980): 463–479. On natural experiments of tax changes, see N. Eissa, "Labor Supply and the Economic Recovery Tax Act of 1981" (pp. 5–32); and J. J. Heckman, "Comment on Labour Supply and the Economic Recovery Tax Act of 1981" (pp. 32–38) in *Empirical Foundations of Household Taxation*, ed. M. Feldstein and J. Porteba (Chicago and London: University of Chicago Press, 1996).

REVIEW AND PRACTICE

SUMMARY

1. The imposition of a tax that is not a lump-sum tax introduces inefficiencies. The magnitude of the inefficiencies is measured by the deadweight loss, the difference in revenues that could be obtained from a lump-sum tax as compared to a distortionary tax, with the same effect on the level of welfare of consumers.

2. The effect of any tax can be decomposed into an income effect and a substitution effect. There is an income effect associated with a lump-sum tax, but no substitution effect. The greater the substitution effect, the greater the deadweight loss.

3. There is also a deadweight loss associated with the reduction in the price received by producers as a result of the imposition of a tax. The reduction in their profits exceeds the tax revenues they effectively pay to the government.

4. For a tax on a commodity, both the income effect and the substitution effect usually lead to a reduction in the level of consumption of that commodity. For an interest income tax, as viewed by a saver, the income effect typically leads to an

increase in savings, and the substitution effect leads to a decrease in savings; the net effect is ambiguous. Even if the net effect is to leave savings unchanged, however, there is still a distortion associated with the interest income tax. For workers, the income and substitution effects of an increase in wages have opposite effects; thus, higher wages may lead to either an increase or a decrease in labor supply.

5. Empirical evidence suggests that for men, the substitution and income effects of wage taxes virtually cancel, so the total effect of the tax on the male labor supply is probably not large. For women, there may be a marked effect on labor force participation. On the other hand, even though the total effect may be small for males, the substitution effect, and hence the deadweight loss associated with the tax, may be significant.

KEY CONCEPTS

Hawthorne effect
Income effect
Producer surplus
Substitution effect

QUESTIONS AND PROBLEMS

1. If savings do not respond to changes in the interest rate, does it mean that there is no deadweight loss associated with the taxation of interest?

2. What is the deadweight loss from the mineral tax in Chapter 18, problem 1? What is the relationship between deadweight loss and supply curves? Relate this to the discussion of lump-sum taxes.

3. Taxes and government expenditure programs affect a variety of other aspects of household behavior. Some economists, for instance, argue that they affect birth rates. What provisions of the tax system might affect the decision to have a child? What government expenditure programs?

4. Instead of representing the individual's decisions as a choice between consumption and leisure, they could be represented in terms of a choice between consumption and work. Draw the indifference curves, and identify the income and substitution effects resulting from a change in the tax rate on labor.

5. Compare the effects of a wage tax and a lump-sum tax raising the same revenue. In particular, show that the individual's utility is higher with the lump-sum tax than with the income tax.

6. Compare the effects of a proportional income tax and a progressive flat-rate income tax; that is, one in which there is a lump-sum grant from the government of, say, $3000, and then a constant marginal tax rate on all income. In particular, show that if the two taxes raise the same revenue, and all individuals have the same income, utility will be higher with the proportional tax.

7. Prior to 1981, the government imposed only a 67 percent (instead of a 100 percent) marginal tax rate on income earned by a mother receiving AFDC. Draw the budget constraint before 1981 and after 1981. Draw the indifference curve of someone who prefers to remain out of the labor force under both regimes. Draw the indifference curve of someone who worked before 1981 but chose not to work after 1981. Show how, for this person, lowering the tax rate will increase utility, reduce costs to the welfare system, and increase labor supply. Finally, draw the indifference curve of someone who worked both before and after 1981. Show how, for this person, the lower tax rate affects AFDC costs and affects labor supply. What can you say about government policy if there are some individuals of the first type, some of the second type, and some of the third type?

8. What would be the effect of a switch to taxing individuals on the basis of their own income (rather than family income) on labor force participation of wives?

9. Describe the income and substitution effects of an increase in the interest rate for a borrower. What does this imply for the effect of eliminating tax deductibility of interest payments?

APPENDIX: MEASURING THE WELFARE COST OF USER FEES

We can measure the cost of user fees, say, for the use of a bridge, using the techniques that are introduced in Chapter 7.

The loss in welfare is given by the shaded triangle in Figure 19.16. This is referred to as the *deadweight loss*. To see this, we recall that the points on the demand curve measure the individual's marginal willingness to pay for an extra trip at different quantities. Assume a price, p, was charged for the use of the bridge. The number of trips taken would then be Q_e. The welfare loss from not taking the trip is the difference between what the individual is willing to pay (the marginal benefit) and the marginal cost. The willingness to pay at Q_e is p, and the cost of providing an extra trip is zero; thus, the welfare loss is just p. At slightly higher levels of usage, the loss is still the marginal willingness to pay, but this is now smaller. To find the total welfare loss, we simply add up the welfare loss associated with each of the trips *not* taken as a result of charging the toll. At a zero price, Q_m trips are taken; at a price of p, Q_e trips are taken. Hence, the toll results in $(Q_m - Q_e)$ trips not being taken. The loss in welfare from the first trip not taken is p, the loss in welfare from the last trip not taken is zero. (The willingness to pay for one additional trip at Q_m is zero.) The *average* welfare loss from each trip not taken is thus $p/2$, and the *total* welfare loss is $p(Q_m - Q_e)/2$, the area of the shaded triangle in Figure 19.16.[21]

---

[21] As noted in Chapter 7, this is only an approximation for the deadweight loss. The correct calculation entails using the *compensated* demand schedule, not the ordinary demand schedule. However, if the fraction of income spent on traveling across the bridge is small, the two demand curves differ by very little.
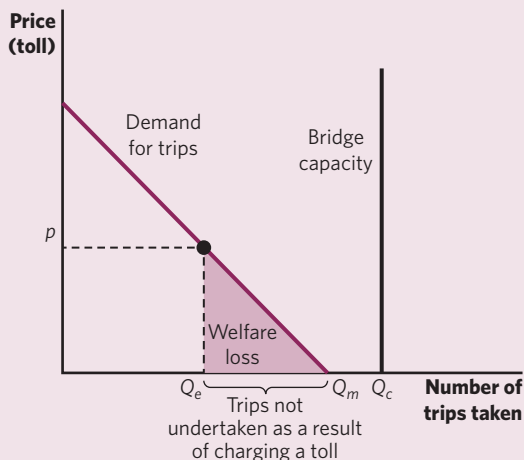
BRIDGES: HOW A USER FEE CAN RESULT IN WELFARE LOSS

As a result of a toll, p, some trips across the bridge are not taken, even though they would be beneficial to society as a whole. The total welfare loss created by the toll is represented by the shaded region.

20

OPTIMAL TAXATION

In the previous chapter we observed that there may be a significant welfare loss (the *deadweight loss*) associated with any tax other than a lump-sum tax. Two questions immediately arise: Why, if this is the case, do we not just impose a lump-sum tax? And if we are to impose distortionary taxes, is there some way that they can be designed to minimize the deadweight loss? These questions have been at the center of theoretical research in taxation. The research has produced some remarkably simple and insightful answers, answers that may help us to design better tax systems in the future.

The chapter is divided into four sections. The first section disposes of two fallacies that have long confused discussions of tax design. Next, the basic principles of optimal taxation are described, and then applied to analyze the design of income tax structures. The final two sections analyze commodity taxation. The third section focuses on the effectiveness of taxing consumers' purchases of different commodities at different rates in achieving redistributive goals, and the fourth on the role of taxation of producers.

TWO FALLACIES OF OPTIMAL TAXATION

Before turning to the details of the analysis, we need to dispose of two fallacies that have misled discussions of tax design—one suggesting an overly simplified approach, the other that the world is so complex that nothing can be said.

THE FALLACY OF COUNTING DISTORTIONS

The first fallacy says we should simply have a tax on wage income. Additional taxes—taxes on commodities such as cigarettes or alcohol, or taxes on savings—just add to the number of distortions and thus to economic inefficiency. One distortion is better than several distortions.

A tax on wage income would be optimal if there were no distortions associated with that tax, for then that tax would be equivalent to a lump-sum tax. However, we showed in the previous chapter that an income tax distorts individuals' decisions to work, and it is not necessarily the case that one large distortion is better than several smaller distortions. Chapter 19 showed that the deadweight loss from a tax was proportional to the square of the tax rate. This suggests that it may be better to have a number of small taxes than a single large tax.

MISINTERPRETATIONS OF THE THEORY OF THE SECOND BEST

In earlier chapters we characterized Pareto efficient resource allocations. All the required conditions are seldom satisfied. The **theory of the second best** is concerned with the design of government policy in situations in which the economy is characterized by some important distortions that cannot be removed.[1] This is in contrast to "first-best" economies, in which all the conditions for Pareto efficiency can be satisfied. Second-best considerations say that it may not be desirable to remove distortions in sectors in which they can be removed. The theory of the second best is often interpreted fallaciously as saying that as long as there

---

[1] Early formulations of the theory of the second best include those of J. Meade, *Trade and Welfare: Mathematical Supplement* (Oxford: Oxford University Press, 1955), and R. G. Lipsey and K. Lancaster, "The General Theory of Second Best," *Review of Economic Studies* 24 (1956–1957): 11–32.

FOCUS QUESTIONS

1. What are the trade-offs involved in designing a progressive income tax system?

2. What should be the role of the taxation of commodities (such as luxuries) and savings in achieving greater equity in taxation?

3. If the government imposes taxes on different commodities, how should the tax rates be set to minimize the total deadweight loss?

are some distortions, economic theory has nothing to say. This is incorrect, as we shall shortly show. Economic theory can tell us under what circumstances two small distortions are preferable to one large one, when it is better to have inefficiencies in both consumption and production, and when it is better not to have inefficiencies in production. Second-best theory tells us that we cannot blindly apply the lessons of first-best economics. Finding out what we should do when some distortions exist is often a difficult task, but it is not impossible.

OPTIMAL AND PARETO EFFICIENT TAXATION

Chapter 3 introduced the concept of Pareto efficiency. Recall that a resource allocation was Pareto efficient if no one could be made better off without someone else being made worse off. Similarly, in judging tax structures we again use the concept of Pareto efficiency: a **Pareto efficient tax structure** is one such that there exists no alternative tax structure that can make some individuals better off without making other individuals worse off.[2] If such an alternative tax system exists, then the current tax system is clearly inefficient.

There are many Pareto efficient tax structures, just as there are many Pareto efficient resource allocations without taxes. In each, no one can be made better off without someone else being made worse off. They differ in distribution. In the two-person economy of Crusoe and Friday, Crusoe is better off in some Pareto efficient allocations, whereas Friday is better off in others.

In Chapter 7, we learned how one can choose among Pareto efficient resource allocations using a social welfare function. So, too, in choosing among Pareto efficient tax structures: the **optimal tax system** is the set of taxes that maximizes social welfare. Clearly, different social welfare functions will generate different optimal tax structures. At a practical level, for instance, a social welfare function that reflects a greater concern for equality (such as a Rawlsian social welfare function) may imply that the optimal tax structure is more progressive, with the rich bearing a larger fraction of the burden for paying for public goods. One of the

---

[2] For a more detailed description of Pareto efficient tax structures, see J. E. Stiglitz, "Self-Selection and Pareto-Efficient Taxation," *Journal of Public Economics* 17 (1982): 213–240; and J. E. Stiglitz, "Pareto Efficient and Optimal Taxation and the New New Welfare Economics," in *Handbook of Public Economics*, ed. A. J. Auerbach and M. Feldstein (Amsterdam and New York: North Holland; distributed in Canada and United States by Elsevier Science Publishers, 1987), pp. 991–1042.

objectives of optimal tax theory is to determine whether there are some general properties of all Pareto efficient tax systems—that is, properties that hold regardless of the social welfare function.

LUMP-SUM TAXES

If all individuals were identical and were treated for tax purposes identically, a lump-sum tax would be the only efficient tax: any other tax would introduce distortions, so the government could raise the same amount of revenue *and* make each individual better off. Furthermore, if everyone were identical, there would be no reason to redistribute income. Both equity and efficiency would thus require that any revenue that the government needed be raised by imposing a uniform lump-sum tax on all individuals.

   In the real world, things are more complicated. Individuals differ, governments wish to redistribute income, and, in any case, there is a strong belief that individuals who can pay taxes more easily should pay more taxes than those who cannot pay as easily. Even if the government wishes to make different people pay different taxes, it does not follow that it would have to impose distortionary taxes, such as income or excise taxes.

WHY IMPOSE DISTORTIONARY TAXES?

Indeed, it can be argued that if the government had perfect information about the characteristics of each individual in our society, it would not impose distortionary taxes. If the government could ascertain who had greater abilities, and who therefore was in a better position to pay taxes, it would simply impose higher lump-sum taxes on those individuals.

   How can abilities be measured? Consider a family. Parents often believe that they have good information concerning the abilities of their children. A parent who has two sons, one of whom has a great deal of ability but chooses to become a beachcomber, and the other of whom has limited ability that he uses to the fullest, is more likely to provide financial assistance to the latter than to the former; the assistance is not made on the basis of income—the beachcomber may, in fact, have a lower income than his hardworking but low-ability brother.

> ## PARETO EFFICIENT AND OPTIMAL TAXATION
>
> - A Pareto efficient tax structure is one such that there exists no other tax structure that can make some individuals better off without making others worse off.
> - The optimal tax structure, given a particular social welfare function, is the Pareto efficient tax structure that maximizes that social welfare function.

ESTIMATING THE OPTIMAL TAX RATE

Determining the optimal tax rate depends on estimating the response to higher tax rates and making judgments about the value of increased equality—the value of an extra dollar to someone at the top versus someone at the bottom.

Those who argue that the "optimal" tax rate should be low typically argue that supply elasticities are large. The extreme version of this—that lowering top tax rates from the 70 percent that prevailed before President Reagan would actually raise revenue—was disproved by what happened subsequent to the tax decrease.

Nobel Prize-winning economist Peter Diamond and Emmanuel Saez of Berkeley, one of the country's leading authorities on inequality at the top, have argued that the "optimal" top marginal tax rate is considerably higher than it is today—it should be over 50 percent, and could be as high as 70 percent (taking into account other taxes that individuals face).

SOURCE: P. Diamond and E. Saez, "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives* 25 (Fall 2011): 165–190.

The government, however, is not in the position of the parent who can observe the ability and drive of his or her children. The government can base its tax only on *observable variables*, such as income and expenditure (and even these, as we shall see, are not easily observable). The choice facing the government is to have either a *uniform lump-sum tax*—one that individuals pay regardless of what they do or what their abilities are—or a tax that depends on easily measured variables, such as expenditures or income; such a tax is inevitably distortionary. An income tax does not always succeed in taxing those whom we might think ought to be taxed: it treats equally the individual who has low ability but works extremely hard and the individual who is of high ability and takes it easy, provided the two have the same income. Still, most people believe that those who have a higher income ought to pay a higher share of government costs because those with a higher income are, *on average*, more able or have had better-than-average luck. Moreover, society may reasonably value the loss of income by the rich (implying, say, one less yacht) less than it values the loss of income to lower-income individuals.

The use of distortionary taxes is thus an inevitable consequence of our desire to redistribute income in a world in which the government can observe the characteristics of individuals only imperfectly. Still, some tax systems are less distortionary than others.

RENT SEEKING, INEQUALITY, AND OPTIMAL TAXATION

With the top 1 percent of the country receiving more than 20 percent of the nation's income—a level not seen since before the Great Depression eighty years ago—there has been renewed attention on inequality and what gives rise to it.

A growing number of economists believe that much of the inequality at the top is at least partially a result of rent seeking—successful efforts by those at the top to seize a larger share of the nation's economic pie. If that is the case, their incomes are not commensurate with their contributions in making the size of the nation's economic pie larger. Indeed, what they do may have the opposite effect. Among those at the top are those who have gained their income and wealth from the exercise of monopoly power, taking advantage of deficiencies in corporate governance rules that allow them to take a disproportionately large share of the corporation's income, and those in the financial sector, some of whose income is derived from predatory lending practices, abusive credit card practices, and market manipulation. Low tax rates at the top increase the return to such rent-seeking activities. If that is the

case, higher taxes at the top might actually increase growth and efficiency, at the same time that they increase equality.

Making matters worse are tax provisions like favorable treatment to capital gains, including those originating from speculaton that increase inequality (they are of benefit overwhelmingly to those at the very top) and distort the economy, shifting resources away from more productive activities into speculation.

There is some evidence that is consistent with the hypothesis attributing much of the income at the top to rent seeking: countries that have raised taxes at the top have not seen a decrease in their growth rates.

It would, of course, be still better to target the rent seeking activities directly, but that may not be easy or possible. What should be clear is that, given the level that inequality has reached in the United States and the sources of that inequality, simple models postulating that there is a trade-off between inequality and efficiency may be missing the mark: there is much that can be done through tax as well as expenditure policies that could improve both.

SOURCE: T. Piketty, E. Saez, and S. Stantcheva, "Taxing the 1%: Why the Top Tax Rate Could Be Over 80%," *Vox* 8 (December 2011), http://www.voxeu.org/article/taxing-1-why-top-tax-rate-could-be-over-80.

DESIGNING AN INCOME TAX SYSTEM

Pareto efficient tax structures minimize distortions. For instance, one might ask: Is it better to redistribute income just through a progressive income tax or to supplement a progressive income tax with a tax on luxuries consumed by the rich? Before addressing that question, however, we ask a simpler one: Assuming there are no savings, so the only source of income is wages, and the only tax is an income tax, how progressive

should the tax system be? That is, how much larger a portion of their income should rich people pay?

As always, economists focus on trade-offs. Here, the more progressive the tax, the larger the deadweight loss, the inefficiencies from the tax, but the less the degree of inequality. We can view much of the political debate concerning how progressive the tax structure should be as one involving differences in values, in how much deadweight loss one is willing to accept for a given decrease in inequality.

There may be disagreements not only about values, but also about the empirical question of what the trade-offs are. Those who advocate more progressive taxes tend also to argue that the cost, in terms of the deadweight loss, of reducing inequality is relatively small. In Chapter 19 we showed that the magnitude of the deadweight loss from a tax was related to the substitution effect. If leisure and consumption goods are very substitutable, then the compensated labor supply schedule will be very elastic, and there will be a large deadweight loss from a tax on consumption or labor income. If consumption this period and consumption next period are very substitutable, then the savings schedule will be very elastic, and the deadweight loss associated with an interest income tax will be large. Those who believe that the deadweight losses are small are often referred to as *elasticity optimists*; they believe, for instance, that the (compensated) labor supply and savings elasticities are low, so the distortions associated with high tax rates are low, whereas those who believe that the distortions are large are often referred to as *elasticity pessimists*, because they believe that the labor supply and savings elasticities are large.

WHY DOES MORE PROGRESSIVITY IMPLY MORE DEADWEIGHT LOSS?

The preceding section argued that as we use our tax system to attain greater equality, the deadweight loss increases. Figure 20.1 illustrate this general proposition by contrasting two tax schedules. Figure 20.1A is a proportional income tax, in which the tax liability is the same percentage of income for all individuals, no matter how large or small their income. Figure 20.1B is a simple progressive income tax that imposes a tax at a flat rate on the difference between the individual's income and some critical level of income, $\hat{Y}$. Individuals whose income falls below the critical level receive a grant from the government equal to the tax rate times the shortfall between their income and the critical level. Notice from Figure 20.1B that the marginal tax rate, the extra tax an individual pays or receives on an extra dollar of income, is constant for
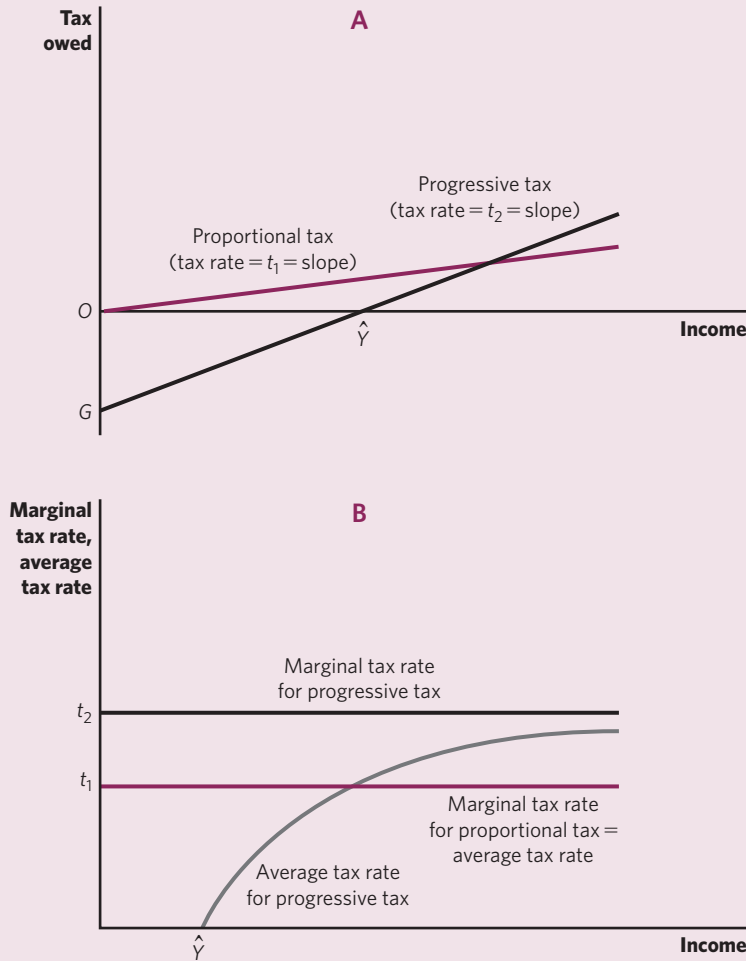
both tax systems. Therefore, both are called **flat-rate taxes**. With the progressive tax, however, the *average* tax rate—the ratio of the total tax payments to the individual's income—increases with income. This is why we call the tax progressive.[3]

Because, as we have depicted it, the progressive flat tax provides for a payment to individuals whose income falls short of the critical level, we

---

[3] Usage is not standardized. Some prefer to reserve the term *progressive* for tax structures in which the *marginal* tax rate increases. Nothing important hinges on these semantic points. Notice that a flat-rate tax combined with a lump-sum tax is regressive, in the sense that the average tax rate decreases with income. For a more general discussion of the definition of progressive and regressive tax structures, see A. B. Atkinson and J. E. Stiglitz, *Lectures on Public Economics* (New York: McGraw-Hill, 1980), Chapter 2.

sometimes refer to that portion of the tax schedule below $\hat{Y}$ as a **negative income tax**.[4]

The progressive flat tax can be thought of as a combination of a uniform lump-sum grant to all individuals and a proportional income tax. Thus, in Figure 20.1A, a proportional tax at the rate t_2, combined with a grant of *OG*, is identical to an income tax on incomes in excess of $\hat{Y}$ ($\hat{Y}$ is the exemption level) at a rate of t_2, provided that those with incomes less than $\hat{Y}$ receive a rebate equal to t_2 times the difference between $\hat{Y}$ and their income. If the government is both to finance its public goods and other public expenditures *and* pay everyone a uniform lump-sum grant, the revenue raised must be higher than if it just financed the public goods, so the marginal tax rate must be higher than with just a proportional tax.

In the last chapter, we learned that the deadweight loss increases with the *marginal* tax rate: the magnitude of the deadweight loss is related to the substitution effect, and the magnitude of the substitution effect is related to the marginal tax rate. More progressive taxes have higher marginal tax rates, and thus greater deadweight loss.

Moreover, the more progressive the tax, the greater the *likelihood* of a smaller labor supply and national output necessitating on that account a still higher tax rate. All lower-income individuals are better off, so both substitution and income effects lead to a smaller labor supply. For higher-income individuals, income and substitution effects are off-setting. Unless they have very backward-bending labor supply schedules, overall labor supply will be reduced.

A DIAGRAMMATIC ANALYSIS OF THE DEADWEIGHT LOSS OF PROGRESSIVE TAXATION

The fact that more progressive tax results in greater deadweight loss can be seen by looking at any individual, and comparing the revenues the government can obtain with two taxes that leave the individual just as well off. The more progressive tax has a higher marginal tax rate.

Figure 20.2 shows a budget constraint with a proportional tax and a budget constraint with a progressive tax, one that gives the individual a fixed income, even if he or she does not work. The marginal rate with the progressive tax is higher, but is set so that the individual is on the same indifference curve. We compare the total revenue. It is reflected in the distance between the before- and after-tax budget constraints. As income is

---

[4] In some tax systems, those with income above $\hat{Y}$ are taxed on the difference between their income and this exemption level, but those below the critical level neither pay taxes nor receive a rebate.
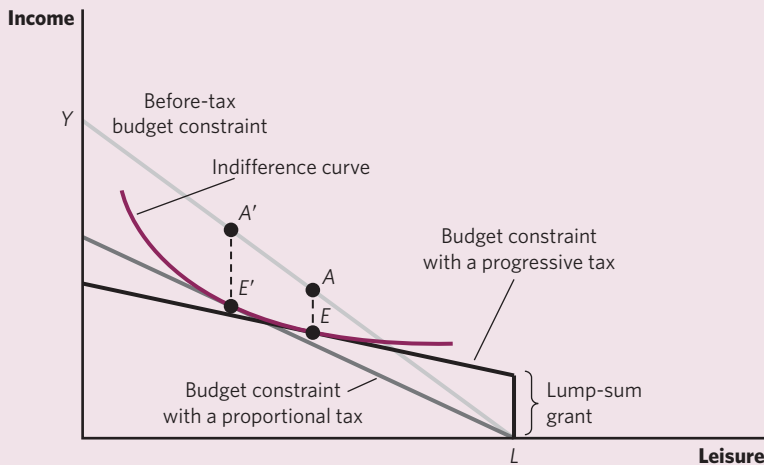
measured along the vertical axis (and hours along the horizontal axis), the tax revenue in dollar terms is measured as the vertical distance between the two budget constraints: for the proportional tax, by the distance $E'A'$; for the progressive tax, by the distance EA. It is apparent that $E'A'$ is much larger than EA—for any given effect on utility, the progressive tax yields lower revenue,[5] so it is less efficient than the proportional tax.

CHOOSING AMONG FLAT-RATE TAX SCHEDULES

The analysis has clarified the trade-offs faced as we increase the degree of progressivity: poorer individuals gain, and richer individuals lose. Like a leaky bucket, the dollar gains of the poor are less than the "dollar equivalent" losses of the rich, because of the deadweight losses associated with the tax. However, the *social value* of the gains of the poor may well exceed the social value of the losses of the rich. Whether this is so depends, of course, on the social welfare function.

In Chapter 7, we introduced the concept of a Rawlsian social welfare function, under which society is concerned about the welfare of the

---

[5] The vertical distance between the indifference curve and the before-tax budget constraint is maximized at the point where the slope of the indifference curve is the same as the slope of the before-tax budget constraint. That is why a lump-sum tax, which does not alter the slope, maximizes revenue for any given impact on utility. Because at E' and E, the slope of the indifference curve is flatter than the slope of the budget constraint, the vertical distance is larger the farther "up" the indifference curve we move.

THE 1993 TAX INCREASE ON UPPER-INCOME INDIVIDUALS: A PARETO INEFFICIENT TAX?

I n 1993, Congress raised taxes on upper-income individuals. The tax increase clearly made them worse off. Critics claimed that it was a Pareto inefficient tax change (although the popular press did not use that vocabulary). The claim was made that these individuals would reduce their work effort—the substitution effect was larger than any income effect—so that tax revenues would be reduced. Thus, funds available to redistribute to the poor would actually be lowered. This did not happen. Instead, tax revenues for the rich increased faster than for others, and far faster than national income.

To be sure, upper-income individuals are worse off than they would have been with lower taxes, but the taxes they paid increased. (Of course, it is possible that without the tax rate increase, incomes of the rich would have increased even more, enough so that tax revenues would have increased. However, this would have required an implausibly large growth in their income, not commensurate with historical experience.) Thus, whether the tax change was desirable depends on the social welfare function, but it does not appear to have been Pareto inefficient.

worst-off individuals. The worst-off individuals are those at the bottom of the income distribution, and their welfare is typically related directly to the size of the lump-sum grant. For a Rawlsian, the optimal tax structure is simply that which maximizes the lump-sum grant; that is, which maximizes the revenue that can be extracted from taxpayers. Such a tax rate may be quite high—one estimate put the number at 80 percent,[6] although others have estimated lower rates. At 80 percent, the deadweight loss incurred by higher-income individuals can be quite high, depending on the elasticities. Other social welfare functions, which put more weight on middle- and upper-income individuals, accordingly suggest a lower optimal tax rate. One estimate put the optimal tax with a utilitarian social welfare function—where all individuals are weighed equally—at 19 percent.[7]

---

[6] N. H. Stern, "On the Specification of Models of Optimum Income Taxation," *Journal of Public Economics* 6 (1976): 123–162. Stern assumed that expenditures on public goods amount to 20 percent of national income. The results of the calculations are very sensitive to all the assumptions made, and, in particular, to assumptions concerning the compensated and uncompensated elasticities of labor supply. As we noted in Chapter 19, there is considerable controversy concerning their magnitude. Those who believe that the uncompensated elasticity is quite high believe that revenues that can be obtained from taxing the rich, to pay the lump-sum grant to the poor, peak out at much lower rates of taxation. Those who believe that the compensated elasticity is quite high believe that the deadweight loss from the progressive taxation is very high.

[7] Stern, "On the Specification of Models of Optimum Income Taxation."

GENERAL EQUILIBRIUM EFFECTS

So far, we have assumed that the income tax has no effect on before-tax incomes, that there is, in other words, no *shifting* of the income tax. Some economists, however, believe that there may be considerable shifting. In particular, it has been argued that the income tax system has increased the degree of before-tax inequality.

Some believe, first, that the wages and fees of managers and professionals adjust to the taxes, leaving their after-tax income relatively unchanged. Moreover, if, as a result of the income tax, skilled workers supply less labor and investment is discouraged, unskilled laborers' productivity (and hence, their wage) will decline. At the present time, unfortunately, we do not know the quantitative significance of these effects. If they are important, it suggests that the benefits of progressivity are less than they seem when these effects are ignored.[8]

RAISING BENEFITS FOR THE POOR   The analysis also makes clear why it is so difficult to provide increased benefits for the poor. It is not "just" that financing those benefits requires raising taxes. There is a real problem in designing the "phaseout"—the rules stipulating how benefits get reduced as income increases. A rapid phaseout implies a high marginal tax rate (as benefits are reduced greatly for each extra dollar earned) over the phaseout income range, thus weakening work incentives. A slow phaseout reduces the magnitude of the disincentive effect, but—if the poorest are to receive the same benefit—raises the benefit levels of others, including lower-middle-income individuals, thus necessitating further tax increases. The objectives of *targeting* and *good incentives* are inevitably in conflict.

Consider the earned income tax credit (EITC), meant to supplement the wage income of poor families with dependent children. The idea behind the EITC was simple: reward the poor for working, thus encouraging them to work more and acquire more skills. In 1993, the EITC was greatly expanded and indexed, so in 1997, the maximum benefit for a family with two or more eligible children was $3656, phased out over

> ### BASIC TRADE-OFF IN TAX DESIGN
>
> More progressive tax systems entail greater dead-weight loss; more "equalitarian" social welfare functions (placing more weight on equality) will choose more progressive tax systems.

---

[8] The importance of these general equilibrium effects in the design of optimal taxes was noted by Martin Feldstein using a simulation model in "On the Optimal Progressivity of the Income Tax," *Journal of Public Economics* 2 (1973): 357–376. His results were corroborated and extended in subsequent theoretical work by N. Stern, "Optimum Taxation with Errors in Administration," *Journal of Public Economics* 17 (1982): 181–211; F. Allen, "Optimal Linear Income Taxation with General Equilibrium Effects on Wages," *Journal of Public Economics* 17 (1982): 135–143; and J. E. Stiglitz, "Self-Selection and Pareto-Efficient Taxation" (note 2). See also L. J. Kotlikoff and L. H. Summers, "Tax Incidence," Chapter 16 in *Handbook of Public Economics*, vol.2, pp. 1043–1092 (note 2).

FLAT-RATE TAXES ARRIVE ON THE POLITICAL SCENE

The presumed simplicity of the flat-rate tax system has long attracted academic economists. In the early 1980s, Robert Hall and Alvin Rabuschka of Stanford University wrote a widely read book advocating the flat tax. In 1996 and 2000, Steve Forbes ran presidential primary campaigns centered around the flat-rate tax. He proposed a high exemption level and a low rate. Like the supply-siders of the 1980s, he believed that the supply response to the lower tax rate would be a huge increase in national income. Most economists, however, thought that the supply response would be far smaller, leaving a huge deficit, estimated in the hundreds of billions per year. Raising the flat rate to eliminate the deficit made the proposal sound less attractive, but even then, it would have represented a huge change in who bears the burden of taxation, with the rich facing markedly lowered burdens and the middle class facing higher burdens. As people examined the idea more closely, their enthusiasm for it languished, and so did Forbes's campaigns. In addition to distributional concerns, it also became clear that a flat tax did not necessarily mean a low tax rate, a large tax base, or a simple tax system. Still, the idea of a flat-rate tax is likely to be an active one on the political scene for years to come, and has been adopted for the state income tax in seven states. It has also been adopted in different forms and with varying degrees of success by several members of the former Soviet bloc, including Russia.

the range of \$11,950 to \$29,290, implying a marginal tax rate of 21 percent from the EITC. When the Clinton administration took office, it had hoped to expand the EITC so all those working full-time would be lifted out of poverty. However, the overriding desire to reduce the deficit led to a lower maximum benefit than was required to achieve this goal. Fiscal constraints forced a shorter phaseout range, thus leading to greater marginal disincentives. In 2011, the implied marginal tax rate remained 21 percent for a two-child family in which the heads of household were married and filing jointly, although the maximum benefit had risen to \$5112 and the phaseout had expanded to a range of \$21,800 to \$46,000.

The discussion so far has focused on the optimal flat-rate tax. In fact, the United States has had, for a long time, a highly nonlinear schedule, with marginal rates varying from zero to almost 40 percent. Nonlinear tax structures increase complexity and, for reasons that are explained more fully in Chapter 24, increase incentives and opportunities for tax avoidance. At the same time, they can reduce the total deadweight loss associated with attaining any set of revenue and distributive goals.

Earlier, we saw that the deadweight loss is related to the marginal tax rate and the elasticity of the (compensated) labor supply. The basic principles of efficient progressive income taxation are derived from that insight:

1. Impose high average tax rates with low marginal tax rates.

2. Make as few people as possible face high marginal tax rates.

3. Impose high marginal tax rates on those for whom the tax is least distorting.

Figure 20.3 compares two tax structures: a progressive flat-rate tax and a tax structure with high marginal tax rates at low incomes and very low marginal tax rates at very high incomes. *OB* is the lump-sum grant given to someone who does not work and has no other source of income. This grant gets phased out as income rises. The high marginal tax rates (high phase-out rate) over the interval *BC* mean that at incomes beyond *C*, average rates can be higher while marginal rates are lower. For these middle- and upper-income individuals, this means the government is collecting more taxes with less distortion. The price is greater distortions for those with income in the range *BC*. The total deadweight loss will be low, however, if there are relatively few people in this interval, or if those in it have a relatively low labor supply elasticity. Even if they reduce their labor supply significantly, however, the total economic loss may be relatively low if they produced relatively little.
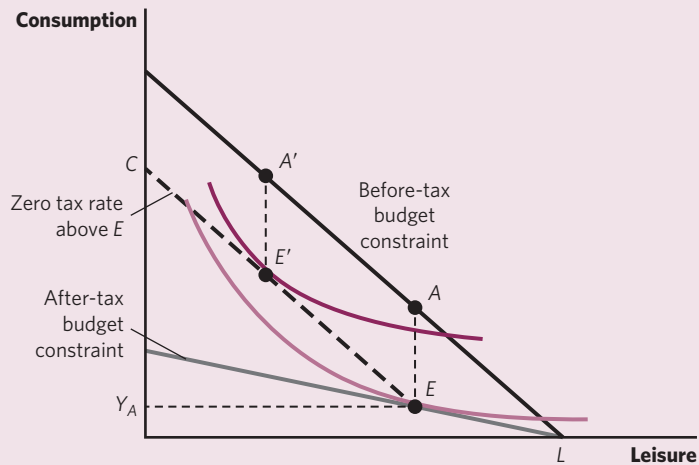
LOWERING TAX RATES FOR THE RICH   Figure 20.4 shows why lowering the marginal tax rate for the highest income groups may be desirable. The figure depicts the budget constraint facing the highest income group, with individuals in the group choosing point E. The revenue raised is the amount EA. If we now lower the marginal tax rate to zero for those who have income above Y_A, the government still collects the same revenue, but with a lower marginal tax rate, these individuals work harder and are still better off. (The new budget constraint is the dotted line $EE'C$, and they choose point E' on a higher indifference curve.) Thus, the tax reform is a Pareto improvement: the rich are better off, and no one is worse off. Now, if instead of imposing a zero marginal tax rate beyond Y_A we had imposed a low marginal rate, the higher-income individuals would still be better off, but there would be additional tax revenues collected, which could be used to reduce taxes on the middle class and/or increase subsidies for the poor. *All individuals could be made better off.*

Such reasoning provided part of the rationale for the reduction in the tax rates at the upper end of the income distribution enacted in 1986, and again in 2001 and 2003. The hope is that lowering the marginal tax rate will lead to more labor supply, which will lead to more output and more tax revenue. As we have noted elsewhere, the evidence that there are significant supply side effects at the top is scant. While lowering capital gains taxes temporarily may induce some individuals to sell their assets (they worry if they wait, the tax rate may go up again), the gain in revenues now comes at the expense of lower tax revenues in the future. Much of the income at the very top is associated with what economists call rents, such

as monopoly profits. A recent study[9] shows that increasing tax rates at the top seems to have no adverse effect on economic performance; that is, it does not have the adverse "supply" side effects predicted by the standard model. (See case study, "Rent Seeking, Inequality, and Optimal Taxation.")

DIFFERENTIAL TAXATION

The government imposes a huge array of taxes on various commodities, from airline tickets to tires to gasoline to perfume. Taxes that are imposed at different rates on different commodities are called **differential taxes**. Some of these taxes, such as the airline ticket tax, are designed as *benefit taxes*—that is, to make those who benefit from airline travel pay for the costs of the air traffic controller system and airports. Others, such as the taxes on gasoline, tobacco, and alcohol, are partially designed as *corrective taxes* to ameliorate some of the negative externalities they generate, such as traffic congestion and air pollution from automobiles. Finally, some, such as the tax on perfume, are *luxury taxes*, intended to increase the redistributive nature of our tax system.

In this section we address two key questions. First, if the government cannot impose an income tax to redistribute income—as is the case in many less-developed countries—what rates should it impose on different commodities? Second, if the government *can* impose an income tax to redistribute income, should it *also* impose taxes on different commodities at different rates? The two questions turn out to have markedly different answers.

RAMSEY TAXES

We begin with an even simpler question posed by the great Cambridge economist Frank Ramsey. Ramsey was not concerned with redistribution, only with efficiency, but he assumed that the government could not impose a lump-sum tax,[10] and hence, it had to raise revenues through commodity taxation. The question he asked was: What is the least distortionary pattern of taxes? For instance, should every commodity be taxed at the same rate, in which case the tax is just a tax on income? (Recall the discussion of equivalent taxes in Chapter 18.) That was the answer suggested by those who simply wanted to count distortions, since such a tax would have only

---

[9] T. Piketty, E. Saez, and S. Stantcheva, "Taxing the 1%: Why the Top Tax Rate Could Be Over 80%," *Vox* 8 (December 2011), http://www.voxeu.org/article/taxing-1-why-top-tax-rate-could-be-over-80.

[10] F. Ramsey, "A Contribution to the Theory of Taxation," *Economic Journal* 37 (1927): 47–61. The question had been posed to him by his teacher, A. C. Pigou. See A. C. Pigou, *A Study in Public Finance,* 3rd ed. (London: Macmillan, 1947).

one distortion. Ramsey showed not only that that was wrong, but also that there was a simple formula for the optimal tax rate.

The commodity taxes that minimize the deadweight loss are called **Ramsey taxes**. Under certain simplifying conditions, Ramsey taxes are proportional to the sum of the reciprocals of the elasticities of demand and supply:

$$\frac{t}{p} = k \left(\frac{1}{\eta_u^d} + \frac{1}{\eta^s} \right),$$

where k is a proportionality factor that depends on the total amount of revenue the government is attempting to raise, t is the per unit tax, p is the (after-tax) price, η_u^d is the compensated elasticity of demand, and η^s is the elasticity of supply. If the elasticity of supply is infinite (a horizontal supply schedule), the tax should simply be inversely proportional to the compensated elasticity of demand. Ramsey's result should not come as a surprise. In Chapter 19, we showed that the deadweight loss from a tax increased with the compensated elasticity of demand and with the elasticity of supply. (Recall also Figure 18.8 in Chapter 18.)[11]

Figure 20.5 shows the solution to the optimal commodity tax problem. Figure 20.5A depicts the deadweight loss as a function of the tax rate imposed on commodity i. Figure 20.5B shows the revenue raised as a function of the tax rate imposed on commodity i. From these two diagrams we can calculate, at each tax rate, the ratio of the increase in deadweight loss to the increase in tax revenues from raising the tax a little bit—that is, the marginal deadweight loss from raising an extra dollar of revenue from a tax on commodity i. Notice that we have drawn the curve not only so that excess burden increases as the revenue raised increases, but also so that the extra deadweight from raising an extra dollar of revenue increases with the tax rate (and thus with the revenue raised). This follows from the fact that the deadweight loss increases with the square of the tax rate.

A similar curve can be derived for commodity j, as shown in Figure 20.5C. The tax rates should be set so that the increase in deadweight loss per extra dollar raised is the same for each commodity. If the increase in excess burden per extra dollar raised were greater for one commodity than for another, by adjusting tax rates so that one less dollar was raised on the first commodity and one more dollar was raised on the second commodity, total deadweight loss would be reduced.

---

[11] If there is a tax rate t on corporate profits, the Ramsey formula is modified to

$$\frac{t}{p} = k \left(\frac{1}{\eta_u^d} + \frac{1-t}{\eta^s} \right).$$

Hence, if corporate profits are taxed at 100 percent, the tax rate is simply inversely proportional to the elasticity of (compensated) demand.

Deadweight loss

A

Deadweight loss increases with tax rate

Tax on commodity i (t_i)

Tax revenue

B

Revenue function

t_i

Marginal deadweight loss / Marginal tax revenue

C

Extra deadweight loss from raising an extra dollar of revenue

Commodity i

Commodity j

$t_i^{\star}$ $t_j^{\star}$ t
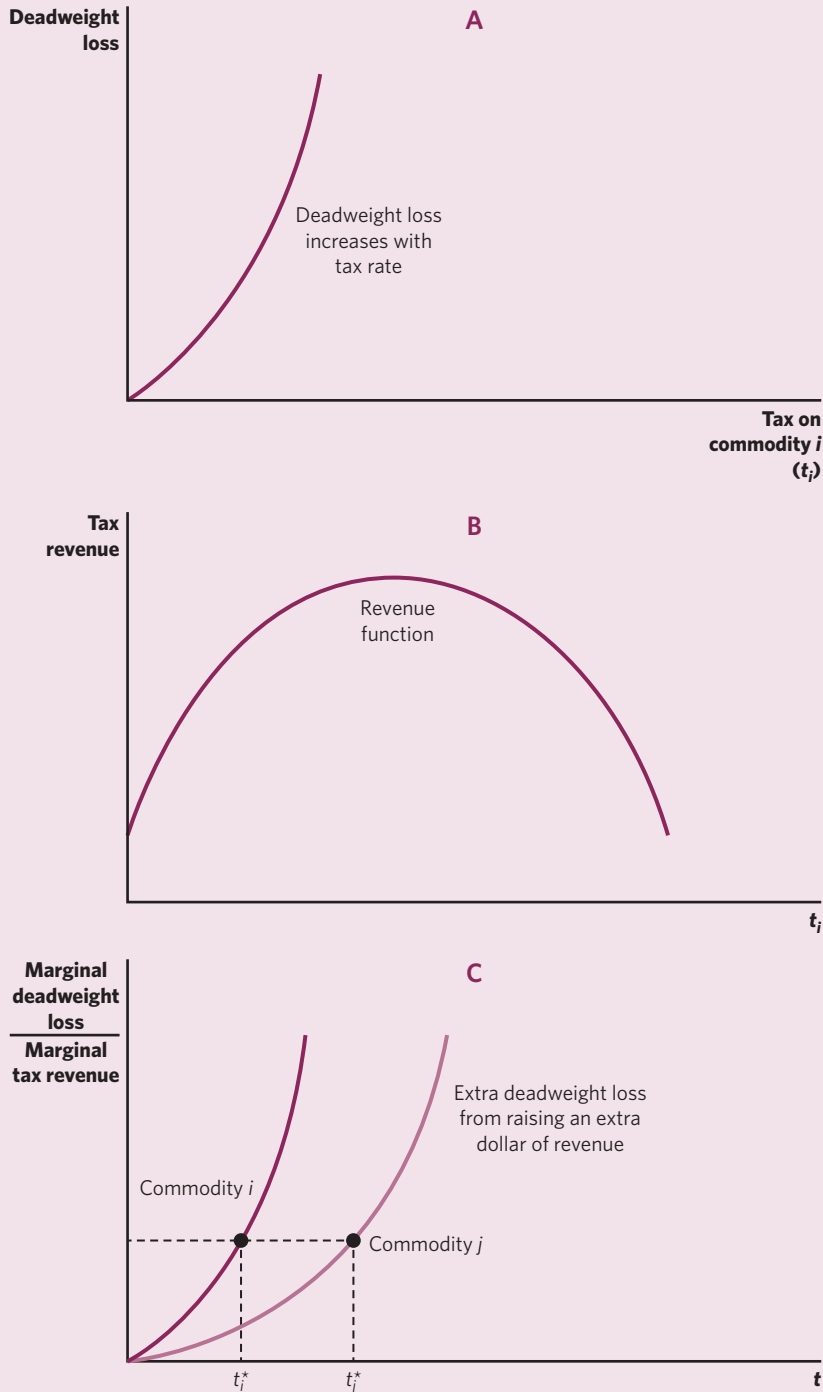
FIGURE 20.5

OPTIMAL COMMODITY TAXATION
The marginal excess burden (deadweight loss) per marginal dollar raised must be the same for all commodities.

In Figure 20.5C, the marginal deadweight loss per marginal dollar of revenue raised is higher for commodity i than for commodity j at any given tax rate. To equate the marginal deadweight loss per marginal dollar of revenue raised, we must impose a lower tax rate on i than on j. Ramsey's basic insight was to observe that *commodities with low elasticity of demand (or low elasticity of supply) have a lower marginal deadweight loss per marginal dollar of revenue raised, and thus should face higher marginal tax rates*.

OPTIMAL COMMODITY TAXATION WITH INTERDEPENDENT DEMANDS[12]   The result we have just given requires that the compensated demand curves of each commodity are independent; that is, the demand for one commodity does not depend on the price of another. Another interpretation of Ramsey's result holds when supply curves are infinitely elastic, whether or not demand curves are interdependent: *The optimal tax structure is such that the percentage reduction in the compensated demand for each commodity is the same.*[13]

ALTERNATIVE INTERPRETATION: OPTIMAL COMMODITY TAX STRUCTURE WITH INTERDEPENDENT DEMANDS   An income tax is distortionary because it induces individuals to make "incorrect" decisions concerning the amount of labor they wish to supply. Commodity taxation may help correct that distortion. If we tax commodities that are complements for leisure and subsidize commodities that are complements for work, we encourage individuals to work, and thus reduce the distortion caused by a uniform commodity tax (which is equivalent to just a wage tax). For instance, by taxing ski equipment and subsidizing commuter costs, we induce individuals to work more and consume less leisure.[14]

REDISTRIBUTION AND RAMSEY TAXES: COMMODITY TAXATION IN LDCs   There is one very disturbing feature of Ramsey's analysis. The major reason that governments use distortionary rather than uniform lump-sum taxes is that they have certain redistributive goals that they cannot achieve otherwise. However, the early discussions of optimal taxation assumed that all individuals were identical, in which case the natural assumption would be that the government would employ uniform lump-sum taxation.

---

[12] This subsection and the remaining subsections of this part of the chapter deal with more advanced topics and can be omitted.

[13] Note that if $\eta^s = \infty$, $t/p = k/\eta_u^d$; with horizontal supply curves, the percentage tax is inversely proportional to the compensated demand elasticity. The percentage change in output is equal to the percentage increase in price multiplied by the percent change in demand from a percent change in price $= (k/\eta_u^d) \times \eta_u^d = k$; that is, it is the same for all commodities.

[14] This interpretation was noted in W. J. Corlett and D. C. Hague, "Complementarity and the Excess Burden of Taxation," *Review of Economic Studies* 21 (1953): 21–30.

This was particularly vexing because the results described earlier suggest that high tax rates should be imposed on commodities with low price elasticities like food. These commodities often have low income elasticities, so if a high tax is imposed on them, the poor will bear a larger burden than the rich. However, the original reason for employing commodity taxation was to shift more of the burden onto the rich than they would face, say, with a uniform lump-sum tax. Ramsey's analysis thus seemed to provide little guidance for any serious policy analysis and was, accordingly, largely dismissed.

> ## RAMSEY TAXES
>
> In the absence of any income or profit taxes, and with all individuals identical, raising revenues so as to minimize deadweight loss requires imposing taxes in inverse relationship to the elasticity of demand and supply.

Subsequent research has extended Ramsey's original analysis to include redistributive goals.[15] Not surprisingly, whether one wishes to tax income-elastic and price-elastic commodities, such as perfume, at a higher or lower rate than income-inelastic and price-inelastic commodities like food depends, in part, on the strength of one's concern for income redistribution.

Less-developed countries typically place little reliance on income taxes, as they have difficulty monitoring income. Indeed, in many cases, they cannot even impose a tax on all commodities, but only on commodities that are imported or exported (as they have some control over what passes over their borders), and on commodities manufactured in the urban sector. Most LDCs have sufficient concern for redistribution that they tax luxuries at higher rates than basic necessities.

DIFFERENTIAL COMMODITY TAXES IN ADVANCED COUNTRIES WITH PROGRESSIVE INCOME TAXES

All advanced industrialized countries, however, do have progressive income taxes. For them, the issue is markedly different from that posed by Ramsey. They ask: *If* there is an optimally designed income tax, does the marginal benefit of the extra redistribution that, say, a tax on luxuries provides exceed the marginal cost in terms of the excess deadweight loss? The naïve answer to this question was rejected in the introduction

---

[15] See, in particular, P. Diamond and J. Mirrlees, "Optimal Taxation and Public Production, I: Production Efficiency and II: Tax Rules," *American Economic Review* 61 (1971): 8–27, 261–278; P. Diamond, "A Many-Person Ramsey Tax Rule," *Journal of Public Economics* 4 (1975): 335–342; A. B. Atkinson and J. E. Stiglitz, "The Structure of Indirect Taxation and Economic Efficiency," *Journal of Public Economics* 1 (1972): 97–119; and A. B. Atkinson and J. E. Stiglitz, "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics* 6 (1976): 55–75; reprinted in A. B. Atkinson, ed., *Modern Public Finance*, vol.2, in *International Library of Critical Writings in Economics* (Aldershot, UK, and Brookfield, VT: Elgar, 1991), pp. 82–102.

to this chapter. Earlier discussions had suggested that introducing more distortions was bad, and therefore, differential commodity taxation was undesirable, but this fallacy was dismissed: one simply cannot count the number of distortions. Remarkably enough, though, the conclusion of these earlier discussions was correct: if an income tax is well designed, adding differential commodity taxation is likely to increase the ability to redistribute income little, if at all. The objective of taxation is to redistribute income, or to impose the burden of taxation on those most able to afford it, and it turns out that the best way to do this, after all, is to focus taxation on what we are really interested in, namely income.[16]

INTEREST INCOME TAXATION AND COMMODITY TAXATION

In our earlier discussion, we showed how a tax on interest income discourages future consumption. It changes the slope of the budget constraint in the same way that a tax on future consumption only would.

Thus, an income tax that taxes interest can be viewed as a differential commodity tax in which future consumption is taxed more heavily than current consumption. The question whether it is desirable to tax interest income is then equivalent to the question whether it is desirable to tax future consumption at higher rates than current consumption.

Just as little may be gained by adding differential commodity taxation with a well-designed income tax, so little is gained from taxing consumption at different dates at different rates. This means, in effect, that interest income should be exempt from taxation. An income tax that exempts interest income is, of course, equivalent to a wage tax, and we showed in Chapter 18, that a wage tax was equivalent to a consumption tax (in the absence of bequests). This suggests that it may be optimal to have a consumption tax. We discuss this further in Chapter 25.

This conclusion, however, illustrates the care that one has to take in ensuring that the assumptions that go into any analysis are appropriate. Implicit in this analysis are assumptions such as (a) we tax all wage income with an optimal progressive income tax; and (b) we do not care about inequality directly, including whether it is perpetuated through inheritance, leading to the establishment of a self-perpetuating plutocracy. For instance, especially in the financial sector, managers are able to get themselves paid in forms that appear as capital income (capital gains). They are, for instance, rewarded by giving them a share of the firms they restructure. If there were no taxes on capital income, they would avoid paying taxes entirely.

---

[16] Indeed, under standard assumptions, Pareto efficient taxation requires that there be no differential taxation of commodities. See Atkinson and Stiglitz, "The Design of Tax Structure" (note 15).

When an individual earns a higher than normal return on his invest-ments, it can be viewed as a result of his own "efforts"—that is, a return to his labor—even though it appears as a return to capital. The model just described ignores these very important differences.

So, too, large amounts of money are transferred from one generation to the next *without paying any inheritance tax*. Those receiving these inheritances are better able to pay taxes than similar individuals without such an inheritance, and it would be inefficient not to tax them: the tax on their capital income would enable a lowering of tax rates paid by others, and thus the distortions of the tax system. There is little evidence that a tax on capital income of one's heirs (at a rate corresponding to taxes paid on other forms of income) would have a significant effect on the efforts of those providing the bequest.

TAXES ON PRODUCERS

So far this chapter has focused on taxes on households: on their wage and interest income and their consumption. Many people believe that it is only fair that firms pay taxes too. Such reasoning is misguided: firms never bear the incidence of a tax, as we have seen, but individuals do, as shareholders, workers, or consumers. Figuring out the incidence of taxes on corporations is a complicated matter.

However, we can ask a more general question: Does Pareto efficient taxation imply that taxes should be imposed on production processes? The taxes described thus far interfere with one of the three conditions for Pareto efficiency discussed in Chapter 3, product mix efficiency: the marginal rate of transformation differs from the marginal rate of substitution.[17,18] Do we want to maintain production efficiency even if we cannot maintain product mix efficiency?

Many of our taxes also affect the production efficiency of the economy, which is to say that they result in the economy's not being on its produc-tion possibilities schedule. Production efficiency requires that the mar-ginal rate of technical substitution between any two inputs be the same in all firms, and that the marginal rate of transformation between any two outputs (or between an input and an output) be the same in all firms.

---

[17] With a tax on wage income, the marginal rate of transformation (wage) exceeds the marginal rate of substitution (after-tax wage); with differential commodity taxes, relative producer prices, which equal the marginal rate of transformation, differ from relative consumer prices.

[18] We can also ask if it is possible to charge individuals with different incomes different taxes on consumption, and whether it is desirable to do so. In other words, is it desirable to maintain *exchange efficiency*? Under the conditions in which no differential commodity taxation is desirable, of course, there is exchange efficiency in the consumption of all goods; but when differential taxation is desirable, it is also, in general, desirable to have relative tax rates dependent on income.

Productive efficiency is attained when all firms face the same prices for inputs and outputs. Thus, any tax on an input that is not uniform across all firms, or any tax on an output that is not uniform across all firms, results in the economy's not being productively efficient. For instance, the corporation income tax is widely viewed as a tax on capital inputs used in incorporated firms, because it raises the after-tax cost of capital in corporations above that in unincorporated businesses. In addition, whereas gasoline that is used for most business purposes is taxed, gasoline used for farming is not. These are only the most obvious examples.

Many production activities are performed in both the market and non-market sectors. Only activities performed in the market sector are taxed. Thus, an individual driving to work is performing the same service that a taxicab driver who drives the individual to work performs. However, there is a tax on the latter and not on the former. A person who bakes a loaf of bread at home is performing a service similar to that of a baker, but is not taxed in the same way that the baker is taxed. There is thus a distortion between the marketed and nonmarketed sectors, and the economy is not productively efficient. The same holds true for taxing formal but not informal markets in developing countries, and is central to the debate about disincentives for microenterprises and small businesses to formalize.

Any tax on *intermediate goods*—goods used to produce other goods—is distortionary. To see this most clearly, consider a firm that produces and uses computers in its own production plants; the cost of the computer is simply the cost of the factors of production, including the return to capital employed in the production. In a competitive economy, this firm would be forced to sell the computers at its costs of production, so the cost of any other firm using a computer would be the same as the cost of the manufacturing firm in using it. Now, however, when a sales tax is imposed, the cost to the firm manufacturing the computer and using it is less than the cost to another firm using the computer in its production processes. There is thus an important distortion, and the economy is no longer productively efficient.

Should the government impose such distortionary taxes if it wishes to minimize the deadweight loss of the tax system? One naïve answer to this question is to say, of course not, the government should not introduce any additional distortions that it does not need to. This kind of argument is similar to the arguments we discussed earlier concerning differential commodity taxes. It makes no sense simply to count the number of distortions. However, it turns out that under some circumstances, the conclusion of the naïve argument is correct.

If the government is able to tax away all profits in the private sector, and if there are no other restrictions on the ability of the government to impose taxes (other than the ability to impose lump-sum taxes), it is

possible to show that productive efficiency is desirable. Hence, the government should impose no distortionary taxes on businesses. *Whatever the government could do with a distortionary tax on producers, it could do better with a direct tax on consumers that maintained the economy on the production possibilities schedule*.[19]

This analysis has some very strong implications. It suggests, in particular, the undesirability of import duties and of taxes on corporations that differ from taxes on unincorporated businesses.

There are many instances, however, when governments face difficulties in imposing taxes. For instance, governments cannot distinguish between final consumer use of a commodity and the use of the commodity by a business; thus, if a government is to impose a tax on consumers it must also impose a tax on business use. Whenever the government is not able to identify and tax away *all* pure profits in the private sector, and whenever there are other restrictions on the ability of the government to impose taxes, it may be desirable to impose distortionary taxes on producers.[20]

The basic insight, though—suggesting that one look unfavorably on taxes that interfere with productive efficiency—is still a valuable one. Taxes on imports, for example, introduce an important inefficiency in the economy; at least in more developed countries, governments can impose a tax on the consumption of these goods rather than on just imports; and, *in general*, such consumption-based taxes are preferable.[21]

THE DEPENDENCE OF OPTIMAL TAX STRUCTURE ON THE SET OF AVAILABLE TAXES

Throughout this chapter, we have noted the dependence of the optimal tax results on the assumptions made concerning the set of available taxes. This was particularly true for commodity taxation. Whether there should be differential commodity taxation, and, if so, how the difference in rates should be chosen, depends on whether there is an income tax and if there is, on its structure. Ramsey showed that *in the absence of any income tax* (and assuming no redistributional objectives), different commodities should be taxed at

---

[19] This result was originally established in the important paper by Diamond and Mirrlees, "Optimal Taxation and Public Production, I: Production Efficiency" (note 13). See also A. J. Auerbach, "The Theory of Excess Burden and Optimal Taxation," Chapter 2 in *Handbook of Public Economics,* vol. 1, pp. 100–101.

[20] This result was established in J. E. Stiglitz and P. Dasgupta, "Differential Taxation, Public Goods and Economic Efficiency," *Review of Economic Studies* 39 (1971): 151–174.

[21] For a more extended discussion of the relationship between trade taxes and commodity taxes, see P. Dasgupta and J. E. Stiglitz, "Benefit–Cost Analysis and Trade Policies," *Journal of Political Economy* 82 (January–February 1974): 1–33.

different rates depending only on the elasticities of demand and supply. When there is an optimally chosen income tax, it may not be desirable to impose differential commodity taxes. When it is desirable to impose differential commodity taxes, they do not depend simply on the elasticities of demand.[22]

It should be emphasized, however, that the set of taxes that is feasible should itself be a subject for analysis: it depends, in particular, on what variables are easily observable and verifiable. In developing countries in which there are many barter transactions (trade not for cash) and in which the level of record keeping is low and there are few businesses with formal payroll operations, it is difficult to enforce an income tax, and commodity taxes must be relied on to redistribute income and to ensure that the burden of taxation is equitably borne. In the United States and other high-income countries, however, the case for the use of redistributive commodity taxation is weak.

---

[22] The central question is whether the additional redistribution that might be obtained from differential commodity taxation is worth the extra deadweight loss.

When there is a flat-rate income tax, with the tax rate chosen optimally, the optimal tax rate on a commodity is simply inversely proportional to the elasticity of demand, and proportional to a parameter that measures the extent to which the good is consumed relatively more by the rich (so a tax on that good is progressive). In some simple cases, that distributional parameter itself is proportional to the price elasticity of demand; goods with low elasticities of demand like food have low deadweight losses, but a tax on them is regressive. The two effects (efficiency or deadweight loss and distribution) are offsetting, and there should be either no differential taxation on different commodities, or it should depend on parameters other than the elasticity of demand.

In the more general case in which an optimal income tax can be imposed that is not necessarily flat—that is, marginal rates can vary with income—a critical determinant of the commodity tax structure is how the marginal rate of substitution between two commodities depends on leisure; in the case in which marginal rates of substitution among commodities do not depend at all on leisure, there should be no differential commodity taxation.

REVIEW AND PRACTICE

SUMMARY

1. Pareto efficient tax structures are such that there is no alternative that can make any individual better off without making some other individual(s) worse off. The nature of the Pareto efficient tax structure, in turn, depends on the information available to the government.

2. There are important trade-offs between distributional goals and efficiency in the design of tax structures. The optimal tax structure balances the gains from additional redistribution with the costs in terms of loss in efficiency.

3. The deadweight loss associated with the magnitude of the substitution effect suggests that it is desirable to have low marginal tax rates in the parts of the income distribution in which there are a large number of individuals—which is to say, in the middle income ranges. On the other hand, high marginal rates in such ranges enable the government to collect the same or greater revenue with a lower marginal tax rate from upper-income individuals. This reduces the deadweight loss per dollar of revenue raised from upper-income individuals.

4. Ramsey taxes minimize the deadweight loss associated with raising a given revenue through commodity taxes alone. In the simple case of independent demand and supply curves, the higher a good's supply and compensated demand elasticities, the lower the tax rate on a good.

5. Whether different commodities should be taxed at different rates depends on the taxes that are available to the government. If the government has imposed an optimal income tax, there may be little, if any, gain from the imposition of differential commodity taxes.

6. If there are no pure profits in the private sector (the economy is perfectly competitive, or the government can impose a 100 percent profits tax) and if there are no other restrictions on the ability of the government to impose taxes, then the government should not impose any taxes that interfere with the productive efficiency of the economy. When these stringent assumptions are removed, it may be desirable to introduce taxes that interfere with productive efficiency.

KEY CONCEPTS

Differential taxes

Flat-rate taxes

Negative income tax

Optimal tax system

Pareto efficient tax structure

Ramsey taxes

Theory of the second best

QUESTIONS AND PROBLEMS

1. "If there are groups in the population who differ in their labor supply elasticity, they should be taxed at different rates." Justify this in terms of the theory of optimal taxation, and discuss its implications for the taxation of working spouses.

2. Earlier, we noted that consumption at different dates could be interpreted just like consumption of different commodities at the same date. What do the results on optimal taxation imply about the desirability of taxing interest income? (Hint: Recall that the price of consumption tomorrow relative to the price of consumption today is just $1/(1 + r)$, where r is the rate of interest.)

3. Explain why it might be desirable to have a regressive tax structure, even if the social welfare function is utilitarian, when general equilibrium effects of taxes are taken into account. Would it ever be desirable to impose a negative marginal tax rate on very–high-income individuals?

4. If you believed that those who were more productive in earning income also had a higher marginal utility of income (they were more efficient in consumption), what would that imply for the design of tax structures? Discuss the reasonableness of alternative assumptions.

5. Under what circumstances will an increase in the progressivity of the tax schedule increase the degree of before-tax inequality?

6. To what extent do you think that differences in views concerning how progressive our tax structure should be reflect differences in values, and to what extent do they reflect differences in judgments concerning the economic consequences of progressivity (deadweight loss, shifting)?

7. One argument sometimes made in favor of the use of commodity taxation rather than income taxation is that people do not accurately perceive the amount they pay in commodity taxes. They will object less to a 20 percent income tax supplemented by a 10 percent sales tax than to a 30 percent income tax. Do you think this is true? If it is, what do you think it implies about the design of tax policy?

8. Explain why the EITC may actually lower total work effort of the poor even if it increases labor force participation. (Hint: Focus separately on those below and above the maximum benefit level.)

APPENDIX A: DERIVING RAMSEY TAXES ON COMMODITIES

The formula for Ramsey taxation, given horizontal supply schedules, may be derived using calculus and certain standard results from microeconomic theory. We represent the individual's utility by the *indirect utility function*, giving the individual's level of utility as a function of consumer prices $(p_1, p_2, p_3, \ldots)$ and of income (I): $V = V(p_1, p_2, p_3, \ldots, I)$. A standard result[23] is that the change in utility from a change in price is just equal to the (negative of the) quantity consumed times the marginal utility of income $\partial V/\partial I$:

$$\frac{\partial V}{\partial p_i} = -Q_i \frac{\partial V}{\partial I}.$$

Let us now increase the per unit tax on, say, the first commodity (t_1) and reduce the per unit tax on the second commodity (t_2) in such a way as to leave utility unchanged. Because with horizontal supply curves producer prices are fixed, the change in the consumer price is just equal to the change in the tax: $dp_1 = dt_1 > 0$, $dp_2 = dt_2 < 0$. Clearly, to keep utility unchanged, the required change in the tax on the second commodity must satisfy $dV = (\partial V/\partial p_i)\, dt_1 + (\partial V/\partial p_2)\, dt_2 = 0$. We can substitute in the values of $(\partial V/\partial p_1)$ to obtain

$$\frac{dt_2}{dt_1} = -\frac{Q_1}{Q_2}.$$

Thus, if the quantity consumed of the first commodity is large (so the loss in welfare from the tax increase is large), the reduction in taxes on the second commodity must be large.

If the demand for each commodity depends only on its own price, then the change in revenue induced by an increase in the tax on the first commodity is just

$$\frac{\partial(t_1 Q_1)}{\partial t_1} = Q_1 + \frac{t_1 dQ}{dp_1} = Q_1 \left(1 + \frac{t_1 dQ_1 p_1}{p_1 dp_1 Q_1}\right) = Q_1 \left(1 - \frac{t_1}{p_1} \eta_u^1\right),$$

where η_u^1 is the compensated demand elasticity for good 1. The term $t_1(dQ_1/dp_1)$ represents the *loss* in revenue resulting from reduced sales in

---

[23] This result is known as Roy's identity. For a proof, see H. Varian, *Microeconomic Analysis*, 3rd ed. (New York: W. W. Norton & Company, 1982), pp. 106–107; or A. Deaton and J. Muellbauer, *Economics and Consumer Behavior* (London: Cambridge University Press, 1980), pp. 37–41.

response to the changed price. The reason why it is the compensated demand elasticities that are relevant is that we are considering variations in two tax rates that, together, *leave the individual at the same level of welfare*.

Similarly, for each change in the tax on the second commodity, the change in revenue is given by

$$Q_2\left(1 - \frac{t_2}{p_2}\eta_u^2\right).$$

The total change in revenue is thus

$$\frac{dR}{dt_1} = Q_1\left(1 - \frac{t_1}{p_1}\eta_u^1\right) + \frac{dt_2}{dt_1}Q_2\left(1 - \frac{t_2}{p_2}\eta_u^2\right)$$

$$= Q_1\left[\left(1 - \frac{t_1}{p_1}\eta_u^1\right) - \left(1 - \frac{t_2}{p_2}\eta_u^2\right)\right]$$

$$= Q_1\left(\frac{t_2}{p_2}\eta_u^2 - \frac{t_1}{p_1}\eta_u^1\right).$$

With an optimal tax structure, this must be zero; that is, given that we are keeping the level of utility of the individual constant, revenues must be maximized. But this requires that

$$\frac{t_2}{p_2}\eta_u^2 - \frac{t_1}{p_1}\eta_u^1 = 0.$$

Generalizing this condition to all commodity taxes, $t_1, t_2, \ldots t_i, \ldots$, we know that $(t_i/p_i)\,\eta_u^i$ must be the same for all, that is, for all commodities. Let k be that value, so that

$$\frac{t_i}{p_i} = \frac{k}{\eta_u^i}.$$

This means that tax rates must be inversely proportional to compensated demand elasticities—this is the Ramsey rule.

APPENDIX B: DERIVATION OF RAMSEY FORMULA FOR LINEAR DEMAND SCHEDULE

Figure 20.6 illustrates a linear compensated demand schedule, $Q = a - b(p + t)$, with a fixed producer price (infinite elasticity supply schedule) and a tax t. The slope of the demand schedule is b. The deadweight loss

$$\text{DWL} = \frac{1}{2}\,bt^2,$$

so the marginal deadweight loss from increasing the tax is

$$\text{MDWL} = bt.$$

The revenue raised by the government is

$$R = tQ = at - b(pt + t^2),$$

so the marginal revenue from increasing the tax is

$$\text{MR} = a - b\,(p + 2t).$$
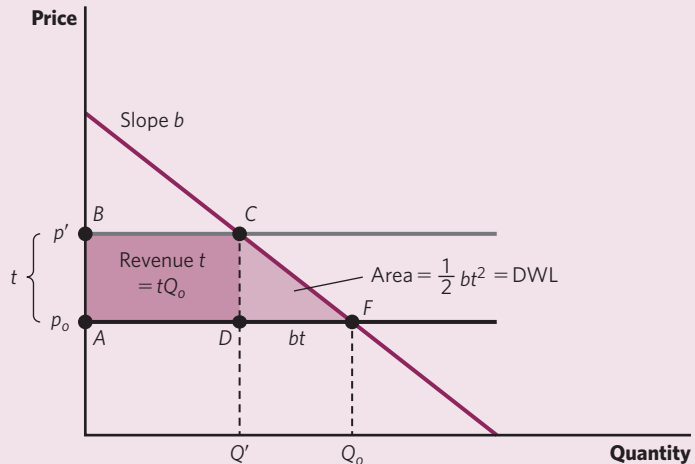
The ratio of marginal revenue to marginal deadweight loss is

$$\frac{\text{MR}}{\text{MDWL}} = \frac{a - b(p + 2t)}{bt}$$

$$= \frac{Q}{bt} - 1$$

$$= k', \text{ the same for all commodities}$$

or

$$\frac{Q}{bt} = 1 + k' \equiv \frac{1}{k}$$

or

$$t = \frac{kQ}{b}.$$

But the elasticity of demand is just

$$\eta_u^d = -\frac{\Delta Q/Q}{\Delta p/p} = \frac{bp}{Q}$$

so

$$\frac{t}{p} = \frac{kQ}{bp} = \frac{k}{\eta_u^d},$$

taxes are inversely proportional to demand elasticities.

21

TAXATION OF CAPITAL

The previous chapter analyzed several of the central problems in tax design, including the trade-offs between redistribution and efficiency, their implication for the degree of progressivity of the tax system, and the role of commodity taxation in the design of an efficient tax system. However, we made only passing reference to one of the most central problems in tax design: how to tax the returns to capital. Should income from capital—interest on bonds and savings accounts, dividends from stock, and the gains that come from selling assets at prices higher than original cost—be taxed at a higher or a lower rate than wage income?

Popular arguments have raged on both sides of this issue. "Capitalists are wealthier than working people," so, it is alleged, capital should be taxed more heavily based on capacity to pay.

On the other side: "Taxation of capital income represents double taxation—taxes have already been imposed when the principal [the original amount invested] was first earned." "Our economy depends on capital investment, and without incentives, there will be insufficient savings and investment." "Taxation of capital is distortionary; by eliminating capital taxation we eliminate one more government-imposed distortion."

The tax system has reflected many of the vagaries of popular discussions. Until 1981, the maximum tax rate on wage income was 50 percent, whereas the maximum tax rate on capital income was 70 percent. At the same time, certain forms of capital income were taxed at lower rates.

Many of these popular arguments on both sides are, if not fallacious, at least misleading. We have already explained what is wrong with arguments that just count the number of distortions. Even if capitalists are, on average, richer, it does not follow that capital income should be taxed more heavily, or even that it should be taxed at all. The issue is: What is the appropriate basis for levying taxes?

This chapter is divided into four sections. The first sets out the basic reasons for and against taxing income from capital. The middle sections discuss key efficiency and administrative issues: the impacts on savings and investment and on risk taking in both open and closed economies. The final section focuses on the complications for the taxation of capital posed by inflation, capital gains (the increase in the value of an asset over time), and depreciation (the decrease in value of an asset as it wears out or becomes obsolete).

At a practical level, corporations make the taxation of income from capital very complicated. The income of corporations, after paying wages, can largely be thought of as a return on capital. That return may be distributed to bondholders or shareholders or may be retained inside the firm. Eventually, of course, the owners of the company will receive the benefits of these retained earnings, in the form of either higher dividends or an increased value of the shares in the corporation when they come to sell them. In principle, we could disregard the existence of corporations—that is, we could pretend that the firm distributed all its profits to its shareholders, who then sent the company back a check (equal to the retained earnings) for reinvestment. While the United States does this for some corporations (closely held, say, by a single individual), it and most countries do not engage in such imputations. They impose taxes on the income of corporations and on the income of individuals, when they receive their dividends or capital gains.

In this chapter, we attempt to ignore these complications by focusing on a simple economy in which individuals own their own firm, investing their savings in capital, facing decisions on how much to save and invest, and what to invest in. The effects of taxes on capital are dependent on a number of the detailed provisions of the tax code—for instance, how the government treats assets as they wear out. We return to several of the themes in more detail in Chapters 22 and 23, in which we look more closely at the individual and corporate income tax code in the United States. The issues raised in this chapter are of broader applicability, facing all countries as they think about the design of their tax system.

SHOULD CAPITAL BE TAXED?

The debate on whether capital should be taxed has centered around three issues: equity, efficiency, and administrative complexity. Before beginning our detailed discussion, we need to review certain basic results on tax equivalence.

RELATIONSHIP AMONG CONSUMPTION TAXES, A WAGE TAX, AND EXEMPTING CAPITAL INCOME FROM TAXATION

In Chapter 18, we showed the equivalence among four tax structures: a proportional consumption tax, a proportional wage tax, an income tax with a tax exemption for income from capital, and a value-added tax, with an exemption for investment goods.[1] The equivalence is important, for it implies that the belief that capital should not be taxed is equivalent to the belief in a consumption tax or a wage tax. Although the different taxes are equivalent, one way of describing a tax may make it look far less attractive than another way.

The equivalence is also important because it provides several alternative ways in which the same tax results can be implemented. Critics of consumption taxes often suggest that consumption is hard to measure. Consumption need not be measured directly, though. A prime example is the value-added tax with an exemption for investment, which is a very important tax in most of the world.

EQUITY ISSUES

Both those who believe that capital should be taxed and those who believe it should not marshal equity arguments in their favor.

One of the most forceful arguments for basing taxes on consumption rather than on income was put forward more than one hundred years ago by Irving Fisher, one of America's most distinguished economists. He argued that it was more appropriate to tax individuals on the basis of what they take out of society (their consumption) than on what they contribute to society (measured by their income).

Beyond this broad philosophical argument is the perspective that (ignoring inheritance), *taxing consumption is equivalent to taxing lifetime income*, as we saw in Chapter 18. Thus, with a consumption tax, two

---

[1] Recall that this equivalence holds only if individuals receive no inheritances and leave no bequests.

individuals with the same lifetime income have the same total tax burden. On the other hand, with an income tax, Imprudence, who puts nothing aside for the future, has a lower present discounted value of tax burden than does her sister, Prudence, even though they have the same present discounted value of lifetime income. (Inequities are introduced if inheritances escape taxation.[2]) This argument supports excluding capital income from taxation, or equivalently, taxing consumption.

On the other hand, consumption taxes are often equated with sales taxes, and sales taxes are widely viewed as regressive. One reason for this is that sales taxes are imposed on only a portion of consumption, and the portion represents a smaller fraction of upper-income individuals' overall consumption than that of lower-income individuals. The consumption taxes under discussion here are levied on *all* consumption; indeed, they can be levied at a progressive rate, so the tax paid may go up more than proportionately with consumption. Progressive consumption taxes are thus also markedly different from value-added taxes, which are proportional.

Critics of a comprehensive consumption tax observe that because richer individuals save a larger fraction of their income than do poorer individuals, a proportional consumption tax is actually regressive, as the ratio of taxes to income for richer individuals is lower. However, this ignores the basic issue of the appropriate tax base. If the *correct* tax base is consumption, then it makes no sense to compare tax payments with income; progressivity or regressivity should be judged by measuring tax payments relative to the appropriate tax base.

ARGUMENTS FOR A CONSUMPTION-BASED TAX

Equity
- People should be taxed on what they take out of the system, not on what they contribute.
- Taxing life-cycle income (ignoring inheritances and bequests) is equivalent to taxing consumption: an income tax discriminates against those who prefer to consume later in life.
- A consumption-based tax can be made progressive.

Efficiency: A Consumption Tax Lowers Deadweight Loss
- Eliminates discrimination against consumption later in life.
- Eliminates distortions arising from hybrid tax system, with some savings taxed differently from others.

Administrative Simplicity
- Much of the tax system's complexity arises from attempts to tax capital income and to reduce avoidance of capital income taxation.

EFFICIENCY ARGUMENTS

There are three categories of efficiency arguments. One category focuses on the deadweight losses, using the kind of analysis introduced in Chapter 20.

[2] Though tax rates on very large inheritances are high, because of loopholes in the tax law, many bequests escaped taxation.

There, we showed that consumption at different dates could be treated like consumption of different commodities. The basic argument that the most efficient way of obtaining redistributive goals is through a progressive wage income tax—with no differential commodity taxation—implies that there should be no differential taxation on consumption at different dates, which, in turn, implies that capital income should not be taxed.[3]

The second category of efficiency arguments focuses on the fact that we have a *hybrid system*, a mixture of a consumption tax and an income tax. Some forms of capital income, such as owner-occupied housing and retirement income, are essentially tax exempt. Others, such as capital gains, are taxed at a preferential rate. This hybrid may be less equitable, more distortionary, and more administratively complex than either a true income tax or a true consumption tax would be.

Popular discussions have not focused on these technical arguments of economists, but rather on a perception that the tax system discourages saving, investment, and risk taking, which are vital for a market economy. These potential effects have become of particular concern in recent years, as savings rates are so much lower in the United States than in some other high-income countries. Later sections of this chapter will address the validity of these concerns.

ADMINISTRATIVE PROBLEMS

Concern with the administrative complexity of our tax laws provides one of the strongest rationales for moving to a consumption base (exempting interest income). Much of the tax code's complexity arises from attempts to reduce opportunities for the avoidance of capital taxes. On the other hand, concern that consumption itself may be difficult to measure at one time provided an argument against a consumption tax. One does not need actually to monitor an individual's purchases of goods in order to tax

---

[3] The formal analysis makes a number of assumptions, such as that the only differences in individuals' income arise from differences in abilities, that relative wages are fixed, and that individuals' marginal rates of substitution between consumption early in life and later in life do not depend on how much they work. If these assumptions are only approximately satisfied, there will still be little to gain from taxing interest income. In some cases, it will be desirable to impose an interest income subsidy, not an interest income tax. There are two circumstances under which a case for an interest income tax can be made. First, if such a tax changes the before-tax distribution in a desirable way. For instance, if decreasing the after-tax return to capital discourages savings, and if unskilled labor and capital are substitutes, the lower capital supply will increase the relative wages of the unskilled. As there is a deadweight loss in redistributing income, it is always desirable to incur some deadweight loss to change the before-tax distribution of income. Second, if individuals differ in their ability to invest, with some individuals obtaining a much higher return to their investments than others, then a wage tax alone or, equivalently, a consumption tax, will not be able to redistribute income efficiently. (The standard models assume that all individuals receive the same return on their capital.) Formally, if one attributes the extra return on capital to the individual's investing ability, then that extra return can be thought of as a return to labor. It is not possible, however, to make this distinction administratively.

consumption, though. Rather, all one needs to observe is an individual's cash flow. Because

$$\text{Income} = \text{consumption} + \text{savings},$$

if one can measure income (total receipts, including gifts) and savings, one can infer the level of consumption. (The measurement of income for an income or consumption tax is identical.) The problem of measuring an individual's savings is not particularly difficult: one simple method calculates the total value of sales of securities and all other assets during a year, less the total value of purchases during the same period.[4] The difference plus the individual's wage income is the cash flow and is equal to the individual's consumption. Thus, the practical problems of implementing a consumption tax may not be that much different from those for implementing an income tax.

EFFECTS ON SAVINGS AND INVESTMENT

In Chapter 18, we discussed the ambiguous evidence concerning the effects of the interest income tax on savings. In this chapter, we assume that there is a negative effect on aggregate savings, and ask: What are the effects on the economy, and, in particular, on the level of investment and, eventually, on the capital stock? To the extent that a reduction in the level of savings gets translated into a reduction in the capital stock, output per worker will fall, and eventually there will be a reduction in standards of living.

EFFECTS OF REDUCED SAVINGS IN A CLOSED ECONOMY

Figure 21.1 illustrates the basic concern. In this figure, we have drawn a curve showing the demand for investment as a function of the interest rate and the supply of savings.[5] The intersection shows the initial equilibrium

---

[4] This understates the problem: not all asset sales and purchases are recorded, and there is some ambiguity between expenditures on assets and those on consumption. Particular problems arise in the transition—because there is no record of *current* asset holding, individuals could consume by selling currently owned assets. An individual might claim the purchase of a painting or a farm as an asset, not as an item of consumption. Although these problems arise today, they might be exacerbated under a consumption tax. (Today, the problem is that an individual may sell a painting and fail to report the capital gain; under a consumption tax, the individual will claim that paintings really purchased for consumption purposes were purchased as investments.)

[5] This analysis makes one crucial simplification: there is only one asset (capital goods) that can be purchased with savings, when in fact there are other assets, including land and government bonds. Government tax policy affects the value of land. Tax policies that result in higher interest rates lead to lower land values; hence, the magnitude of the decrease in capital is less than it would be in an economy with no land. If the value of land decreases, then more of savings can be directed into capital accumulation.

FIGURE 21.1

EFFECT OF TAX ON
INTEREST INCOME
ON EQUILIBRIUM
INVESTMENT IN A
CLOSED ECONOMY

A tax on the return to capital
lowers the equilibrium level
of investment, and in the
long run thus lowers the
equilibrium capital stock.
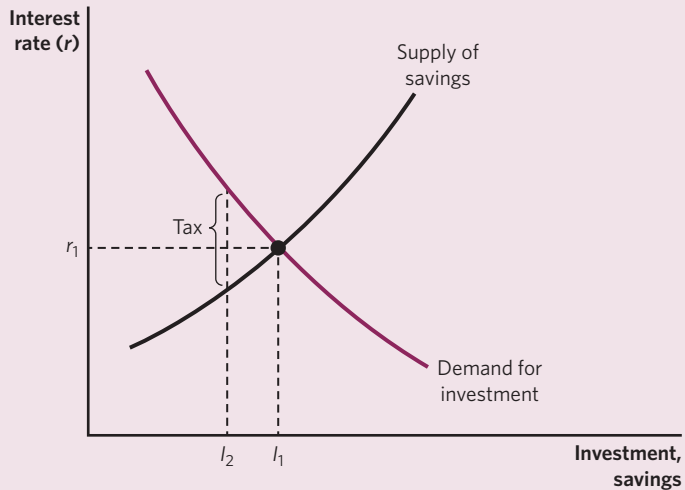
level of the interest rate, r_1, and investment, I_1. We now impose an interest income tax. The tax imposes a wedge between the return to investment and to savings, reducing the equilibrium level of investment to I_2. With lower levels of investment year after year, the economy's capital stock eventually is lowered enough to adversely affect standards of living.

THE DISTINCTION BETWEEN SAVINGS AND INVESTMENT

In a closed economy, in equilibrium, savings must equal investment. Thus, *in equilibrium*, a policy that promotes savings must promote investment, and conversely.

Some policies shift the supply of savings curve and some shift the demand for investment curve. For instance, an investment tax credit, in which the government effectively pays part of the price of capital goods, shifts the demand for investment curve up, as depicted in Figure 21.2A. By itself, this will lead to higher rates of interest and higher levels of investment.

Figure 21.2B shows how such policies can be used to offset (partially or totally) the effects of an interest income tax. In the new equilibrium, investment is the same.

Why, one might ask, impose a tax on savings and then simply offset its effects by a subsidy to investment? The answer is that even though
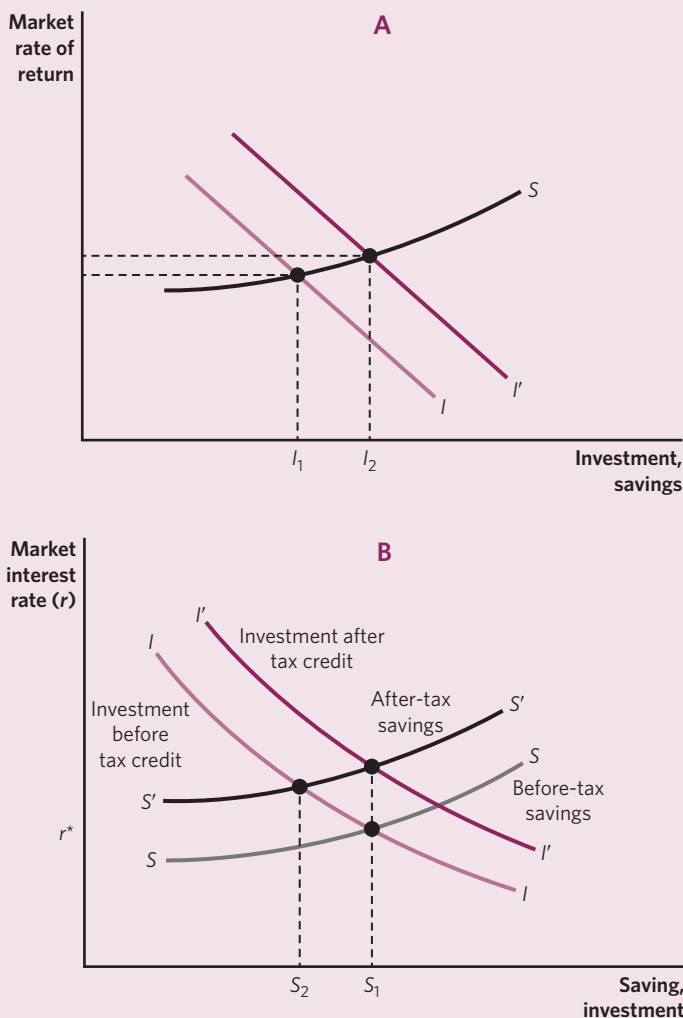
FIGURE 21.2

INVESTMENT TAX CREDITS
(A) An investment tax credit shifts the demand curve for investment, and thus increases the equilibrium level of investment. (B) With an appropriately chosen rate, an investment tax credit can offset the effect of a tax on the return to capital, so investment is left unchanged.

investment may be unaffected by this pair of policies, there may be significant other effects. For instance, not all the return to savings is derived from investment in plant and equipment in the United States; investors obtain returns from investments abroad and in real estate. Moreover, the investment subsidy affects only new capital; taxes on the returns to capital affect old capital. As a result, the combination of a tax on the return to capital and a subsidy to new investment has large redistribution effects, as the price of old capital falls relative to new investments, and can raise substantial amounts of revenue.

NATIONAL SAVINGS AND BUDGET NEUTRALITY

What matters for the nation's investment is, of course, not just the level of household or firm savings (private savings) but also the level of national savings, which includes government savings—its budgetary surplus or deficit. Reducing the tax on the return to capital typically will lower government revenues, and hence increase the deficit, unless some other taxes are increased. Under plausible conditions, the increased deficit more than offsets the increased private savings, so national savings—and, thus, investment—is actually reduced. If savings is relatively inelastic, then private savings will not be increased much, but the impact on the deficit can be considerable. Assume, for instance, that capital income amounted to 20 percent of GDP and that there was a 20 percent capital tax; then the *short-run* effect of cutting the tax on capital income in half (ignoring any impact on interest rates) will be to lower government revenues by an amount equal to 2 percent of GDP, and thus the government deficit will increase by an amount equal to 2 percent of GDP. If private savings currently equals 5 percent of GDP, and the interest elasticity is 0.1, then reducing the tax by 50 percent increases the return to capital by 12.5 percent, and increases savings by just over 1 percent (or 0.05 percent of GDP).[6] Thus, national savings—private savings minus the government deficit—actually falls significantly.

On the other hand, if the government substitutes the decrease in the tax on capital income with an increase in the wage tax, in such a way as to leave the individual just as well off as before, then savings is unambiguously increased. This can be seen in Figure 21.3, in which we have used a simplified model where a given individual lives for two periods (denoted by c_1 and c_2): working in the first and saving for retirement in the second. The BB is the individual's budget constraint before the interest income taxes, and $B'B$ is the budget constraint after; E is the point chosen, and EF is the tax revenue (realized in the second period); and DD is the budget constraint when a wage (or consumption) tax that leaves the individual just as well off is imposed. Now, there is only a substitution effect, and the individual clearly consumes less; the incentive to save is increased, and thus private savings increases. Moreover, government revenue, in present value terms, is increased: the new tax revenue is $E'F'$, clearly greater than EF[7]; thus, the

---

[6] If we ignore the change in interest rates, after tax returns increase from $0.8r$ to $0.9r$ (that is, by one-eighth). This induces an increase in private savings of 1.25 percent (or 0.0625 percent of GDP). The actual increase in private savings is somewhat greater: as the deficit increases, interest rates rise, and this elicits more savings by an amount that depends on the interest elasticity of savings. Standard estimates suggest that approximately one-third of the increased budgetary deficit might be offset by additional induced savings resulting from the higher interest rate, and hence, the overall reduction in national savings might be only about 1.33 percent of GDP—still a sizable sum. (When the induced flow of funds from abroad is also included in the analysis, the net impact on investment is reduced further, to approximately two-thirds of 1 percent of GDP.) If, on the other hand, the elasticity of savings is very large, increased private savings would more than offset the increased fiscal deficit.

[7] DD is parallel to BB, as there is no interest income tax. The distance between two parallel lines is the same everywhere; thus, FG equals $E'F'$.

COMPENSATED CHANGE IN TAX
(A) Shows that a reduction in the interest income tax compensated by an increase in a wage or consumption tax in such a way as to leave the individual just as well off, leads to increased private savings and lower government deficits. (B) Shows the limiting case of L-shaped indifference curves, in which neither savings nor government revenues are affected.

deficit is reduced. The magnitude of the increase in savings may be relatively small, however, depending on the shapes of the indifference curves. Figure 21.3B shows the limiting case of L-shaped indifference curves, in which private savings and government revenues are both unaffected.

These contrasting results emphasize how important it is to formulate the right question, being clear what is being held constant. Typically, the government is contemplating alternative ways of raising a given revenue. In that case, neither of the previous two formulations is quite correct, but the second provides a framework for arriving at the desired answer. Because a tax that generates equal utilities generates greater revenue with the consumption tax, the government could lower the tax rate.

Thus, a consumption tax that raises the same revenue as an income tax generates a higher level of utility and national savings.[8]

But that ignores the distributive consequences of capital taxation. Those with income from capital are disproportionately higher-income individuals. Some have acquired their wealth through inheritance taxation that avoided or evaded inheritance taxes, and some have acquired their wealth through taking advantage of the myriad of tax loopholes. Moreover, some of the seeming return to capital is really a return to labor—and even more would appear to be so if there were a larger gap in the tax treatment of labor and capital. For instance, those who manage other people's money through what are called private equity companies typically get paid by getting a share of the companies in which they invest. They frequently earn large capital gains, but this income is really a return on their labor and should be taxed as such, though currently (as this book goes to press) it is not.

EFFECTS OF REDUCED SAVINGS IN AN OPEN ECONOMY

These effects of capital taxation on investment may be greatly reduced in an open economy, however, where foreign savings can serve as a substitute for domestic savings. During the past five decades, a robust international capital market, allowing funds to flow from one country to another, has developed. In recent years, the United States has been borrowing hundreds of billions of dollars from Europe and Japan.[9] Slight increases in the U.S. rate of interest can draw large amounts of money into America. Many economists believe that, as a result, the supply curve for funds to the United States is close to horizontal (see Figure 21.4).

Suppose that there are no taxes, and consider the limiting case in which foreigners are willing to supply funds at an interest rate of r^*; that is, the supply of funds is infinitely elastic. The equilibrium interest rate would then be r^*, with S_1 being domestic savings, I_1 being domestic investment, and the difference, $I_1 - S_1$, being financed by borrowing from abroad.

---

[8] The timing of the government revenues may differ between the alternative regimes. The analysis compares taxes that generate the same present discounted value of revenues, using the before-tax interest rate.

[9] From 1980 to 1992, the United States borrowed so heavily from abroad that it went from being the world's largest creditor nation to the largest debtor. The U.S. net international investment position—the excess of foreign assets owned by U.S. banks, multinational corporations, and individuals over and above the value of U.S. assets held by foreigners—stood at $360 billion in 1980 and at *negative* $411 billion in 1992. (The main reasons for this change were the large federal deficits, discussed in Chapter 2.) By 2010, the net foreign indebtedness of the United States stood at $2471 billion. See U.S. Department of Commerce, Bureau of Economic Analysis, "International Investment Position of the United States at Yearend, 1976–2010," *International Economic Accounts*, Table 2.
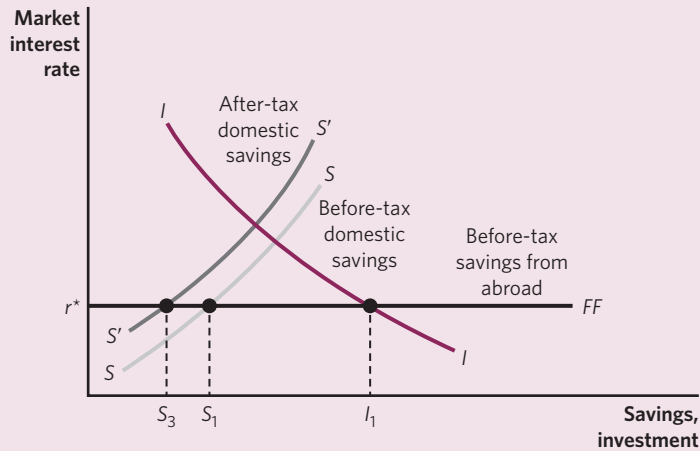
INVESTMENT AND SAVINGS IN AN OPEN ECONOMY
With well-developed international capital markets, foreign borrowing makes up the difference between domestic investment and domestic savings. If only the return to Americans is taxed, then investment will be unaffected, but domestic savings will be discouraged and borrowing from foreigners will rise.

Now, a tax on the return to savings *by Americans* does nothing more than effectively shift the supply curve of domestic savings up and to the left, as depicted in Figure 21.4. The level of investment remains unchanged, but now more of the investment is financed by savings from abroad. Although there is no effect on investment, and thus on the level of productivity in the United States, in the long run, there is an adverse effect on standards of living; for in the long run, Americans will owe more money to foreigners, and more of what is produced in America will have to be sent abroad to pay this indebtedness.

The real world is somewhere between the perfect global capital market just described and the closed capital market discussed in the previous section. Foreign capital is not a perfect substitute for domestic savings. Indeed, empirical studies show a high correlation between domestic savings and investment.[10]

As a result, changes in tax policy that affect domestic savings do affect domestic investment. Earlier, we looked at an example in which capital income amounted to 20 percent of GDP, the capital income tax rate was 20 percent, and the interest elasticity of savings was 0.1. We showed that, under those assumptions, a reduction of the tax by 50 percent would reduce national savings (at a fixed before-tax interest rate) by almost 2 percent of GDP. However, this shift in the national savings "curve" itself leads to higher interest rates,

---

[10] See, for example, E. L. Grinols, "The Link between Domestic Investment and Domestic Savings in Open Economies: Evidence from Balanced Stochastic Growth," *Review of International Economics* 4, no. 2 (June 1996): 119–140. The correlation may be partly because domestic conditions that promote savings tend also to promote investment; or it may be because much of investment is financed by firms themselves, so that when investment returns are high, they save more, or when their profits are high so they can save more, they invest more.

which themselves lead to more domestic savings and increased flows of capital from abroad. Standard estimates suggest that the increased domestic savings would amount to about 0.7 percent of GDP, and the increased capital flows would also amount to about 0.7 percent of GDP. Because

$$\text{Funds for investment} = \text{private savings} - \text{government deficit}$$
$$+ \text{ capital flows from abroad} = \text{investment},$$

this means that investment will be reduced by only 0.7 percent, much less than if there were no capital flows from abroad.

There are still other reasons that an increase in the tax on the return to capital may not have much of an effect on investment. One is that in an open global economy, especially rich individuals (who own a disproportionate share of the wealth) have a variety of ways by which they avoid paying such taxes—though by the same token, such taxes will raise less revenue than they would otherwise.

Moreover, much of the savings may not go to domestic investment, but to investments abroad. And even when savings remain at home, they may simply go to bid up the price of land, not to increase investments in capital goods.

IMPACT ON RISK TAKING

Without entrepreneurs undertaking risks, the capitalist economy would not have grown in the way that it has over the past two centuries, leading to immense increases in standards of living. No wonder, then, that there is alarm about any part of the tax system that might dampen entrepreneurial risk-taking. There is concern that taxes on capital discourage not only the overall level of savings and investment, but also, in particular, the amount of risk taking.

Although some individuals enjoy taking risks on a regular basis, and almost all individuals enjoy taking small risks occasionally (as evidenced by the popularity of state lotteries and the gambling casinos of Las Vegas and Atlantic City), when it comes to managing their wealth, most individuals are more conservative. They are willing to take risks—but only if they receive, as compensation, a sufficiently high expected return over what they could have obtained in a safe investment. There is widespread concern that taxing the return to capital is effectively taxing the return to risk bearing, the risk premium that individuals receive for bearing additional risks, and thus discourages risk taking.

The fruits of risk taking by entrepreneurs are all around us: major inventions such as the automobile, the airplane, and the computer were, in part, the result of investors and firms deciding to gamble their wealth on a new idea. These were the successes, and those undertaking the gamble earned huge rewards. For each of these successes, however, there were dozens of failures. Although the government takes away much of the fruits of success, it is often not as generous in helping to bear the costs of failure, a situation that exacerbates an already present market failure: the absence of insurance markets.[11] Normally, entrepreneurs can only partially divest themselves of the risks associated with their entrepreneurial activity. There is thus concern that even without the interference of the tax system, there would be too little (from a social perspective) investment in risk-taking activities.

WHY CAPITAL TAXATION WITH FULL LOSS DEDUCTIBILITY MAY INCREASE RISK TAKING

There is some controversy over the extent to which current taxes reduce risk taking. It is possible that they may actually increase it.

That the income tax might increase risk taking can be seen most easily by considering an extreme example. Assume that an individual has to decide between two assets: a safe asset yielding no return, and a risky asset that has a 50 percent chance of yielding a very large return and a 50 percent chance of yielding a negative return. The average return is positive, to compensate the individual for risk taking. The individual is conservative, so allocates a fraction of his or her wealth to the safe asset and the remainder to the risky asset.

We now impose a tax on the return to capital, but we allow a full deduction against other income for losses. The safe asset is unaffected. The risky asset has its return reduced by half, but the losses are also reduced by half. How does the individual respond to this? If he or she doubles the amount previously invested in the risky asset, the after-tax income when the return is positive is the same, and the after-tax income when the return is negative is also the same. The tax has left the individual

---

[11] There are several explanations for this market failure. For instance, investors typically have limited information concerning the potential risks and returns of investment projects, and it is costly to obtain more complete information; there is an infinite supply of charlatans willing to take money for hare-brained schemes; or those most willing to sell shares in their projects are those who believe the market has overvalued them. See, for example, B. Greenwald, J. E. Stiglitz, and A. Weiss, "Informational Imperfections in the Capital Market and Macroeconomic Fluctuations," *American Economic Review* 74, no. 2 (1984): 194–199; and S. C. Myers and N. S. Majluf, "Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have," *Journal of Financial Economics* 13, no. 2 (June 1984): 187–221.

TAX INCENTIVES FOR RISK TAKING

The Internal Revenue Code encourages risk taking by new entrepreneurs through two special provisions that allow for either exclusion or tax-free rollover of all or part of the gain from investments in qualified new enterprises.

Investors can exclude up to 50 percent of the gain from the sale or trade of stock in small businesses (C corporations with total gross assets of $50 million or less) that they have held for more than five years.* They can exclude up to 60 percent of their gain from the sale of empowerment zone† business stock. The total amount of gain eligible for exclusion is limited to the greater of ten times what the taxpayer paid for the asset or $10 million.

Investors can also benefit from a tax-free rollover of capital gain from the sale of qualified small business stock held more than six months by purchasing similar replacement stock; for empowerment zone business stock, the asset must be held for more than one year and the proceeds invested in the same empowerment zone.

These provisions are designed to encourage risk taking in new enterprises by engendering "patient capital." With these tax provisions, investors should have greater incentive to commit their funds for at least five years, and thus firms would enjoy an environment favorable to longer-run projects with higher risk. Many advocates of these provisions believe Wall Street to be dominated by the pursuit of short-run gains, and saw the tax incentives as a way to correct the problem.

---

*The exclusion ceiling rises to 75 percent for stock acquired after February 17, 2009, and no later than September 27, 2010, and to 100 percent for stock acquired after September 27, 2010, and before January 1, 2012.
† Empowerment zones are government-designated areas of high poverty and unemployment that qualify for tax incentives to encourage businesses to invest within these zones.

---

completely unaffected. Effectively, the government is sharing in the risks. By its willingness to share the risks—the losses as well as the gains—the government is acting as a silent partner, and because the government is willing to share the risk, the individual is willing to increase his or her risk taking.[12]

The importance of the government's sharing in risk taking depends on how well the private market does. For securities that are actively traded on the stock market, the risks are widely spread throughout the economy. For smaller firms, however, the government may be able to provide risk-sharing opportunities that the market cannot provide.[13]

---

[12] For an early discussion of the effect of taxation on risk taking, see E. D. Domar and R. A. Musgrave, "Proportional Income Taxation and Risk-Taking," *Quarterly Journal of Economics* 58 (1944): 388–422. The standard current view is presented in J. E. Stiglitz, "The Effects of Income, Wealth and Capital Gains Taxation on Risk Taking," *Quarterly Journal of Economics* 83 (1969): 262–283. See also A. B. Atkinson and J. E. Stiglitz, *Lectures on Public Economics* (New York: McGraw-Hill, 1980), Chapter 4; and A. Sandmo, "The Effects of Taxation on Savings and Risk-Taking," in *Handbook of Public Economics*, vol. 1, ed. A. Auerbach and M. Feldstein (Amsterdam: North Holland, 1985), pp. 293–309.
[13] The theory of asymmetric information provides an explanation for why markets do such a poor job in sharing/spreading risks.

This situation has one other interesting property: the tax yields, on average, a return to the government, but it has no effect on the welfare of the individual. The individual is unaffected because his or her after-tax position is the same as it was in the before-tax situation, whether the risky asset turns out to have a positive or negative return. This tax seems to do what no other tax seems capable of doing: it raises revenue (on average) without lowering welfare.

WHY CAPITAL TAXATION MAY REDUCE RISK TAKING

Before becoming too excited about the prospect of raising revenue without lowering welfare, one must keep several caveats in mind. First, with a progressive tax structure, returns to successful investments are taxed more heavily than losses from unsuccessful investments are subsidized. There is thus a built-in bias against risk taking.

Second, in the current system, there are limitations on the magnitudes of the losses that can be offset. Thus, the government, while sharing in the gains, shares in only some of the losses. Again, there is a bias against risk taking. The 1986 Tax Reform Act increased the severity of this limitation by providing that individuals cannot subtract losses from certain investment activities from their wage income in computing their income for tax purposes. (The limitation was imposed to reduce opportunities for tax avoidance through tax shelters. Although the provision achieved that goal, it added greatly to the complexity of the tax code and worsened the tax laws' bias against risk taking.)

A final difficulty is that we have assumed that the safe rate of return is zero. If there is a significant positive return on safe assets, and if it is taxed, there will be a wealth effect associated with the capital income tax. In other words, because the individual is worse off (in a sense, "less wealthy"), he or she is willing to take less risks; thus, the wealth effect may lead to the reduction in the demand for risky assets. However, this wealth effect would have occurred with a lump-sum tax as well. It is not a distortion, but simply a reflection of a lower willingness to take risks at lower levels of wealth.

CAPITAL TAXATION AND RISK TAKING

If the return on safe assets were zero and the government taxed gains and subsidized losses at the same rate, then capital taxation would encourage risk taking; the government would be, in effect, a silent partner. In practice, provisions for loss deductibility are limited, so the net effect is to discourage risk taking. In addition, the wealth effect of capital taxation—it makes individuals poorer—may result in less risk taking, because poorer individuals are less willing to bear risks.

Although the government in effect shares in the *financial* costs of investments, it does not share in the *effort* costs of entrepreneurs. The long hours put in by the innovators who contributed so much to the computer industry like Steven Jobs, the founder of Apple Computers, are now legend. Most of the returns they obtained from their efforts were in the form of gains on the sale of the companies that they started. There has been a real concern that high rates of taxation of capital gains will serve to discourage such risk taking and entrepreneurship. Concern about these adverse incentive effects of taxation was part of the motivation for the large reduction in capital gains tax rates in 1997, 2001, and 2003. Others contend, however, that innovators are driven by other than monetary incentives, and that the tax on capital gains has only minimal effects. In this view, there was little incentive benefit from the reduction in capital gains tax rates. The real point of the reduction in capital gains taxes was to reduce the overall degree of progressivity of the tax system, since most capital gains accrue to the very rich.

MEASURING CHANGES IN ASSET VALUES

Returns on capital come in two main forms—dividends and interest payments, and capital gains. An individual is better off when his or her assets increase in value, just as the individual would be better off if he or she received a dividend. If capital gains could be measured perfectly, there would be no reason to treat capital gains differently from any other return to capital. Indeed, for some assets—like holding gold—the only return is the capital gain. Thus, *if* the returns to capital are to be taxed, then it makes sense to tax the returns in whatever form they take, including capital gains. By the same token, however, a decrease in the value of an asset constitutes a negative return, and that loss needs to be subtracted from the other returns (such as dividends) to ascertain the net return. The problem is that it is often difficult to measure either the increase or the decrease in the value of an asset. Tax codes throughout the world have dealt with this problem by, in effect, giving the taxpayer the benefit of the doubt: in the case of capital gains, the individual typically does not have to pay any tax until the gain is *recognized*—that is, until the asset is actually sold. For assets for which there are active markets, like widely traded stocks and bonds, it would in fact be easy to

tax the capital gains on an annual basis—that is, by simply comparing the value of shares at the end of the year to their value at the beginning of the year. The difference would constitute the capital gain. Such a system is called *marking-to-market*. However, for many assets, such as real estate, it is impossible to tell the value with any accuracy except when a deal is consummated, and there is a worry that treating some assets one way and others another way would not only be confusing, but it could also create biases in the choice of assets. There is a way, admittedly imperfect, of dealing with the capital gains associated with non-marketed assets: treating the returns *as if* they occurred smoothly over the holding period, and calculating the present discounted value of the taxes that would have been paid had the taxes been paid as the gains actually accrued.

In the case of depreciation, the taxpayer is allowed to deduct an estimate of the loss of value as a result of aging or obsolescence. The estimates are based on simple rules, which typically are overly generous: they allow a larger loss of value in the early years of the asset, so the present discounted value of the depreciation allowances is greater than it would be with "true" depreciation—that is, the decrease in value that would have occurred if there were a perfect competitive market for used capital.

Inflation poses one more set of problems for capital taxation. We measure gains and losses in dollars, but the value of a dollar changes over time, as a result of inflation (discussed later).

CAPITAL GAINS

Because capital gains are taxed only upon realization, an individual who owns a security that has increased in value may be reluctant to sell it. The individual knows that if sold, a tax will have to be paid on it. If the individual continues to hold the asset, he or she can postpone the tax until some later date. The present discounted value of the tax liabilities is reduced by the postponement of the tax. The individual is thus induced to hold on to the securities rather than sell them. This distortion is referred to as the **locked-in effect**.

The consequences of this may easily be seen. Assume that an individual bought a security last month at $1, and that it suddenly rises to $101. The individual now expects that it will earn a return lower than the return he or she could obtain elsewhere. Assume, for instance, that the individual believes that there is another investment opportunity that could earn

EQUITY AND THE REDUCTION IN CAPITAL GAINS TAXES

The U.S. long-term capital gains tax rates were reduced substantially in 1997: for high-income individuals, from 28 to 20 percent; for those in lower tax brackets, from 15 to 10 percent. Rates were further reduced in 2003, to 15 percent for high-income individuals and 0 percent for those in lower tax brackets.

A major issue in the debate has been equity: the ownership of assets is much more concentrated than income. Thus, a reduction in capital gains tax rates benefits mainly those at the very top.

Various proposals have been put forward to limit the inequalities generated by reductions in capital gains taxes. For instance, it has been proposed that only up to $100,000 of capital gains be eligible for the special treatment in any year, or that individuals be allowed the special treatment on $1 million over their lifetime. These proposals have been rejected, however, and as a result, the overall progressivity of the income tax has been substantially reduced.

a return of 10 percent. In the absence of taxation, the individual would simply sell his or her security and buy the new investment.

Consider now what happens if the individual sells the security. The individual must immediately pay a capital gains tax.[14] Assuming the individual has to pay a 20 percent capital gains tax, he or she would thus have only $81 to reinvest.

Assume the individual believes he or she will need money in one year's time. The individual's after-tax yield in the new investment, assuming it earns a return in capital gains, is: $(1 - 0.20) \times 10\% = 8\%$. In one year's time, the individual will thus have: $81 \times 1.08 = $87.48. On the other hand, if the individual keeps the $101 in the old investment for one more year, and it increases in value at only 8 percent, he or she will have: $101 \times 1.08 = $109.08. The individual must pay a capital gains tax of 20 percent on the gain (that is, the tax is $0.20 \times $108.08 = $21.61); thus, after tax, he or she has $87.47. The individual is essentially just as well off with the money yielding a return of only 8 percent in his or her current asset than the individual would be if he or she sold the asset and purchased an asset yielding a much higher return of 10 percent—and with any return greater than 8 percent the individual would be better off.

---

[14] The individual's capital gains tax rate for short-term gains (less than 12 months) is equal to his or her personal income tax bracket; on long-term gains, the tax rate is 20 percent.

CONSEQUENCES AND IMPORTANCE OF THE LOCKED-IN EFFECT

There is considerable debate about the consequences and importance of this locked-in effect. Martin Feldstein, former chairman of the Council of Economic Advisers under President Reagan, has claimed that the effect is so large that reducing the capital gains tax would lead individuals to sell securities that they previously had refused to sell, to such an extent that government revenues would actually increase.[15] More recent estimates,[16] however, suggest that a permanent reduction in the capital gains tax rate would have little effect. This is in contrast to a temporary reduction. Clearly, if individuals believe that the tax during, say, the next two years will be substantially lower than subsequently, they will sell during this period; it is as if the government is having a sale. This effect is particularly pronounced if the tax reduction has been anticipated, for then individuals who might have sold their assets shortly before the lower tax takes effect decide that it is worthwhile postponing the sale for a short period.

Even if the locked-in effect is significant, however, there may be only a short-run revenue gain: the taxes that individuals pay now will not be paid later, so government revenues in the long run may change little. Moreover, because the reduction in the tax makes individuals better off, the tax may lead to an increase in current consumption and a decrease in aggregate savings at the same time that current government revenues are increased.

Interestingly, the 1997 tax changes were actually designed to encourage individuals to hold on to their assets longer. Under the legislation, assets held for more than twelve months were subjected to a rate of 20 percent (10 percent for those in the 15 percent income tax bracket); assets held for less than twelve months were subjected to a rate of 28 percent. Differential tax rates for long-term and short-term capital gains have been maintained, but at lower rates under the 2001 and 2003 Bush tax cuts: there is now a maximum 15 percent tax rate on long-term capital gains for most taxpayers (0 percent for those whose regular tax rate would be less than 25 percent, and for special types of capital gains, 25 or 28 percent); short-term gains are subject to the same tax rates as ordinary income.

Even those who argue that there is a significant locked-in effect generally agree that it is largely the result of a special provision in the U.S. tax system that allows assets that are held until death to escape capital gains taxation. Thus, whereas younger people simply save on the *timing* of tax payments by postponing realization—and when interest rates are low, the resulting dis-counted value of tax savings is relatively low—for elderly individuals, the savings from postponing realizations may be very high, as they may be able to

[15] M. S. Feldstein, J. Slemrod, and S. Yitzhaki, "The Effects of Taxing on Selling and Switching of Common Stock and the Realization of Capital Gains," *Quarterly Journal of Economics* 94 (1980): 777–791.
[16] See L. Burman and W. Randolph, "Measuring Permanent Responses to Capital Gains Tax Changes in Panel Data," *American Economic Review* 84, no. 4 (September 1994): 794–809.

avoid the tax completely. The remedy for this problem is not, of course, to lower the capital gains tax rate, but to eliminate this special provision.[17]

EQUITY   Those advocating special treatment of capital gains often point out that the tax is levied not on real capital gains—taking into account the effects of inflation—but on nominal capital gains. This, it is argued, is unfair. Unlike other forms of capital income, though, the tax is imposed only on realization, and this provides a significant benefit. One can ask, would an investor have been better off if the tax were levied on real capital gains, as they accrue, than under the current system? The answer depends on the period over which the asset was held: for many investors, the benefits of postponement more than offset the costs of taxing nominal income; this is especially true for investors who finance a significant part of their investments by borrowing, particularly so in an era of low inflation.

There are further debates about the welfare consequences of the locked-in effect. Much of the discussion has focused on individuals' purchases of securities. Economic efficiency requires that each security be held by the individual who values it the most—the one who thinks that it will yield the highest return.[18] The locked-in effect means that an individual may retain a security even though there is someone else who values it more. This results in what is referred to as **exchange inefficiency**. Some economists believe, however, that the economic consequences of this should not be taken too seriously. They argue that the stock market is essentially a gambling casino for the rich, and that, although the locked-in effect may impair the efficiency of this gambling casino, it has few further repercussions for the economy. In their view, there is not a very direct or strong relationship between the effect of the capital gains tax on the performance of the stock market and the decisions made by the managers and owners of firms concerning, for instance, investment and production.

The one area in which the capital gains tax may have a significant effect on the production efficiency of the economy is in smaller, owner-managed firms. There comes a point in the life cycle of such firms when the original owner–manager's skills and talents become less appropriate for the development of the firm. In the absence of capital gains taxation, the original owner–manager might want to sell the firm to some other entrepreneurs, but the high cost imposed by the capital gains tax may discourage the owner from doing so.

---

[17] Technically, this provision is referred to as a "step-up in basis at death." An individual who inherits a security and then sells it is taxed on the increase in the value from the time the individual inherited it. By contrast, if someone gives a security to another as a gift, and the recipient subsequently sells it, the recipient is taxed on the capital gain *from the time the asset was originally purchased*. The cost to the U.S. Treasury from the "step-up in basis" is estimated at billions of dollars per year.

[18] We define economic efficiency in the usual sense of Pareto efficiency. In the presence of risk, however, there is some question about the appropriate way of measuring the welfare of each individual. The sense in which we use the term here is according to the individual's own expectations concerning the outcome, regardless of the objective reality of those expectations.

DEPRECIATION

Not all assets increase in value over time. Typically, machines become less valuable as they get older. This reduction in value is called **depreciation**, which causes machine owners to incur a capital loss. Conceptually, again, there is no problem in the tax treatment of depreciation: just as capital gains should be added to income, capital losses should be subtracted from it. The problem is an operational one: How do we measure the capital losses? The tax code provides for **depreciation allowances**, which are meant to be estimates of the decrease in value.

The reason that it is so important to make some provision for depreciation is illustrated by the following example. Consider a machine that lasts for, say, five years, after which it dies. The machine generates a revenue stream of $100 a year. The net income is clearly not $100 a year ($500 over the five years). Some account must be taken of the fact that each year, the machine is older, and that eventually it will wear out.

True economic depreciation is the actual decrease in the machine's market value. However, because markets for most types of used machines are not well developed, the government has no easy way of ascertaining what the true decrease in market value is. Therefore, instead of using the true value of economic depreciation, the government uses simple procedures that are supposed to approximate actual depreciation, giving some

DISTORTIONS FROM DEPRECIATION

A favorite pastime of economists is to look for unintended distortions arising from seemingly innocuous tax provisions. One such example is the movable wall, which has become popular in the United States. There are markedly different depreciation rates for buildings and equipment: today, nonresidential buildings depreciate over thirty-nine years, equipment over five to ten years. The boundaries between the two are often not well defined. Clearly, a wall is part of a structure and should depreciate with the rest of the structure. But when is a wall not a wall? If the wall is movable, it could be called equipment—after all, it could be moved from one building to another; to be part of the structure, it must be *attached*. There are, to be sure, other reasons for making walls detached from the structure; for example, it allows for more flexible use of space, in response to changing circumstances. No doubt, though, one of the determinants of the move to movable walls was the large tax advantage.
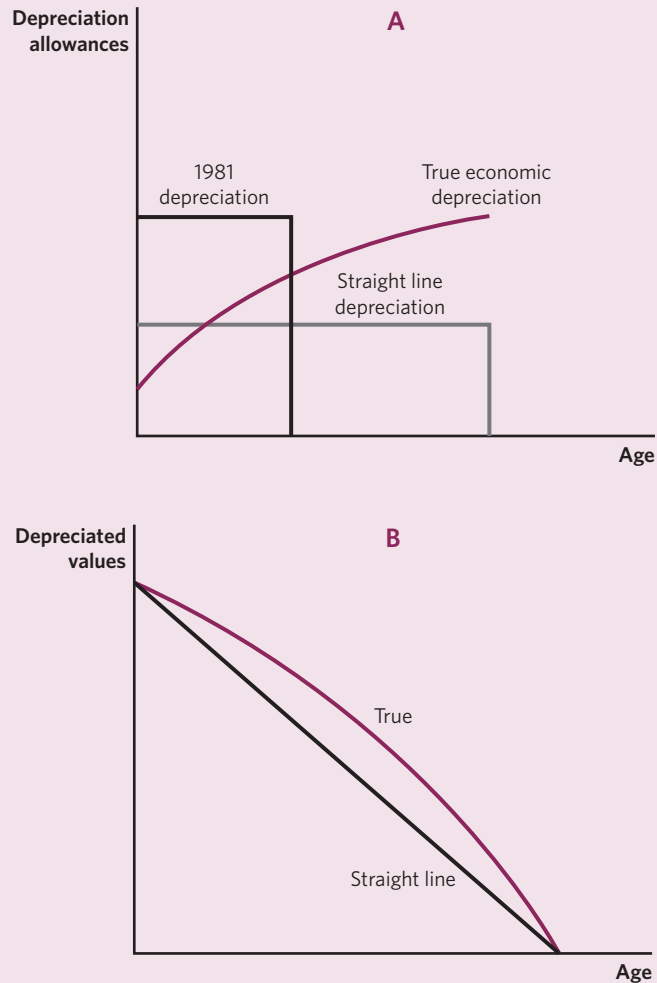
benefit of the doubt to the investors. The procedures entail first estimat-
ing the average life of the machine. For example, cars live, on average, for
six years; commercial buildings, for more than thirty years; office equip-
ment, for five or ten years. The depreciation allowances are then spread
out over the life of the machine. The simplest procedure, called **straight-
line depreciation**, allows the investor to deduct, for instance, one-tenth
of the purchase price for a ten-year machine, one-fifth for a five-year
machine (see Figure 21.5).

NEUTRAL TAXATION

To achieve neutrality in the choice of investment projects, the government has two options. One option we have already described: it would allow true economic depreciation allowances (or at least would attempt to devise rules that more closely approximate true economic depreciation).

The second method entails the government's allowing a 100 percent deduction for the cost of the investment. Then, the government would be reducing the costs of the project by exactly the same amount that it is reducing the benefits (the returns that the investor receives). The government, in effect, would be entering as a silent partner into the enterprise. A project for which the present discounted value of returns exceeds the cost—which therefore would have been undertaken in the absence of the tax—would still be undertaken.

Whereas the first method corresponds to a neutral capital income tax (one that does not distort the choice of investment projects), the second method corresponds to a neutral pure profits tax: the difference between the present discounted value of the returns to an investment project and its costs can be thought of as pure profits.[19]

In practice, many governments do not seek to achieve neutral taxation, but actually use the tax system to encourage investment in capital by **accelerated depreciation**—that is, by allowing depreciation even faster than straight-line depreciation. Accelerated depreciation is of special value to long-lived assets, so tax provisions that allow for accelerated depreciation distort the economy's pattern of investment, encouraging investments in industries and technologies with long-lived assets. In Chapter 23, we look at some of the ways that this has been done in the United States.

INFLATION

Inflation presents several difficult problems in the definition of income from capital. One wants to tax *real* returns to capital, not nominal returns. If an individual owns an asset, and it increases in value by 10 percent but prices in general have gone up by 10 percent, the individual is no better off. This individual's **real capital gain** is zero, even though the **nominal capital gain** is positive. Similarly, consider an individual who puts $1000 in a savings account and receives $100 in interest. If the rate of inflation is 10 percent, the *real return* is zero. The $100 in interest is just enough to compensate the

---

[19] Some of the returns may be attributed to managerial efforts, in which case the difference between the present discounted value of the returns and the *direct* costs (excluding those associated with management) is a mixture of pure profits and return to management and entrepreneurship.

<div style="border:1px solid">

KEY PROBLEMS IN IMPLEMENTING CAPITAL TAXES

- Measuring capital gains: increases in value
- Measuring depreciation: decreases in value as a result of machines wearing out or becoming obsolete
- Inflation: separating out real gains from inflationary gains

</div>

individual for the decrease in the real value of the savings account.

By the same token, inflation reduces the *real value* of the depreciation allowances, which are tied to the nominal price that the individual or the firm paid for the asset, and it does so much more for long-lived assets than for short-lived assets. In the late 1970s and early 1980s, the United States went through a period of high inflation, with annual rates of price increases exceeding 10 percent. (This was low by standards in other countries, where in extraordinary cases inflation rates have exceeded 100 percent *per month*.) It became clear that when inflation rates were high, our tax system did not treat returns to capital in a fair or efficient manner.

The tax system taxes nominal returns, not real returns. Thus, the first consequence of the presence of inflation is that individuals who have a small positive before-tax real return on capital find that they have a large negative after-tax real return.

Consider an individual in the 33 percent tax bracket, receiving a 12 percent return (say, in the form of interest) with an inflation rate of 10 percent. In periods of high inflation, most of the return is just an adjustment for the decreased purchasing power of money. The real return is only 2 percent. (The real rate of return on an asset is the nominal return minus the rate of inflation.) However, present tax laws do not take account of this. This individual would have to pay 33 percent of the nominal return to the government (ignoring state and local income taxes), leaving a net after-tax return of 8 percent. With inflation, the real return is: 8% – 10% = −2%. The individual loses 2 percent of his or her ability to consume simply by postponing consumption by one year. One would expect that this would serve as a strong disincentive to savings.

Between 1974 and 1982, the price of stock was doing little more than keeping up with inflation (and, in many cases, not even doing that). Because the price level doubled, however, individuals found themselves paying a large capital gains tax if they sold their shares. Again, this seemed unfair.

There has been strong support for **indexing** tax brackets—that is, adjusting them to offset the effects of inflation; since 1986, tax brackets, the standard deduction, and personal exemptions have all been indexed. As a result, the income levels at which taxpayers become subject to higher tax rates increase with the price level. Far more complicated than indexing the tax brackets, however, is designing the tax system so only real returns to capital are taxed. Not only must capital gains be indexed, but so must interest payments, as well as interest receipts and depreciation.

Thus, investors who have borrowed money would be able to deduct only their real interest payments, not their nominal interest payments. If capital gains were indexed but interest payments were not indexed, individuals would find it profitable in inflationary times to borrow to purchase capital assets that were increasing in value at the same rate as the rate of inflation.

The fact that much of capital gains are illusory—that is, they do not represent real increases in value—has been one of the motivations for the special treatment of long-term capital gains, such as the lower tax rates enacted in 1997. As we noted earlier, however, the benefits of postponing taxes (capital gains are taxed only upon realization, not as they occur) have, at least in the past, more than offset the "unfair" taxation of inflation gains. More generally, taxing capital gains at a lower rate provides a very imperfect substitute for indexing: especially given the low inflation rates today, assets held for only a short period have almost no inflationary component, and thus the much lower rate at which they are taxed cannot be justified by inflation alone. Moreover, indexing capital gains, without indexing other forms of return or borrowing, creates large distortions.

> ### CAPITAL GAINS, DEPRECIATION, AND INFLATION
>
> - Capital gains typically are taxed only upon the sale of the asset. This gives rise to the locked-in effect. The significance of the locked-in effect is, however, open to question.
> - Most tax systems provide overly generous allowances for depreciation. Such provisions not only encourage investment, but also distort investment patterns—for instance, favoring longer-term investments.
> - The tax system taxes nominal, not real, returns. Even though there is now more indexing in the U.S. tax system than there used to be, returns to capital, including capital gains, are not indexed. Partial indexing—indexing capital gains but not debt—may be more distortionary than the current system. At the inflation rates that prevailed in the 1990s, the distortions associated with the failure to index taxation are small.

Indeed, there has been some controversy over whether the tax system encouraged or discouraged investment in the inflationary period of the late 1970s. On the one hand, the fact that nominal interest payments were fully tax deductible while 60 percent of capital gains was tax-exempt created, in some cases, effective subsidies to capital. On the other hand, the fact that depreciation allowances were not indexed served to discourage investment.

It is apparent that our current tax system is not inflation-neutral; as a result of inflation, in some circumstances, assets with a positive before-tax return have a negative after-tax return, discouraging investment, whereas in other cases, the tax system encourages investment.

Full indexation is required to obtain an inflation-neutral tax system. Partial indexation of capital income (such as indexing capital gains but not debt) would exacerbate some of the distortions and would leave other distortions unchanged.[20]

---

[20] There are other distortions in the tax system that may partially offset the distortions associated with the inappropriate treatment of inflation. For instance, the fact that failure to index depreciation results in excessively high taxation of especially long-lived assets is offset (with high inflation only partially; with very low inflation, more than completely) by the fact that depreciation formulae—even straight-line depreciation—are typically accelerated relative to true economic depreciation.

In the debate leading up to the capital gains tax cut in 1997, the inequity associated with taxing nominal returns was emphasized by advocates of a reduction. However, they failed to note the favorable treatment that results from the fact that capital gains are taxed only upon realization; no one proposed full indexation—in particular, allowing only real interest payments to be deductible—although a proposal to tax only real capital gains actually passed the House of Representatives. The lower tax rates that were enacted are, at best, a rough correction for the effects of inflation, which would have eroded the value of an asset held for twenty years far more than the value of one held for five.

The decline in the inflation rate in the 1990s has taken much of the heat out of the issue. Moreover, there is a growing consensus that the way inflation is measured probably overstates the inflation rate, and possibly by a considerable amount (between 1 and 2 percentage points a year). Still, if the inflation rate again increases to the level attained in the 1970s, there will be a renewed concern about the distortions and inequities associated with the taxation of nominal returns to capital.

REVIEW AND PRACTICE

SUMMARY

1. There are both equity and efficiency reasons for arguing that income from capital should not be taxed. Some argue against the taxation of capital on the grounds that it involves heavy administrative costs; capital taxation accounts for much of the complexity of the tax code. But there are perhaps even stronger equity arguments in favor of the taxation of capital, since without such taxes, much of the income of the richest individuals in society would escape taxation.

2. The taxation of the return to capital tends to reduce savings and investment. In a small open economy, in which only the returns to domestic investors are taxed, investment is unchanged but there is increased borrowing from abroad. In the U.S. economy today—a large open economy—taxation of savings does lead to some lowered investment, but less than would be the case if the United States could not borrow from abroad.

3. A tax on capital with full loss offset provisions (so the government, in effect, subsidizes losses at the same rate that it taxes gains) would—with a zero safe return to capital—typically increase risk taking; the government would be acting as a silent partner. Generally, loss offset provisions are very limited, so the taxation of capital may reduce risk taking.

4. Capital gains—the increases in the value of assets over time—are just another form in which individuals receive a return to capital and should be taxed the same as other returns. There are, however, serious problems in the measurement of both capital gains and losses.

5. The fact that capital gains are taxed only when the asset is sold gives rise to the locked-in effect; individuals may retain an asset when, in the absence of taxation, they would have sold it. In the United States, however, the locked-in effect arises mainly because assets held until death

completely escape capital gains taxation. The locked-in effect would be much reduced if capital gains were taxed on the basis of accrual—that is, as they occur, rather than only upon realization (when the asset is sold).

6. With the United States having the highest level of inequality among advanced countries, there is a special concern about tax provisions that contribute to inequality. The special provisions for capital (lower rates for capital gains and dividends and the step-up of basis upon death) are benefits that accrue mostly to those at the very top of the income distribution.

7. Because the actual decrease in the value of an asset as it wears out or becomes obsolete cannot be easily measured, governments use simple rules to estimate depreciation (called depreciation allowances). Even the simplest rules, such as taking off one-tenth the value of an asset each year for an asset that lasts ten years, tend to be excessively generous; that is, they provide allowances in early years that exceed true economic depreciation (the decrease in the value of the asset in a perfect competitive capital market). As a result, they introduce distortions, with longer-lived assets typically being favored. Tax neutrality requires either that depreciation allowances correspond to true economic depreciation, or that the total value of the asset be depreciated in the year of purchase, in which case the tax becomes a tax on pure profits, not a tax on the return to capital.

8. Ideally, the tax system would tax real returns, not nominal returns; there would be full indexing for inflation. But inflation is hard to measure. Partial indexing—indexing of capital gains but not of debt—may result in even greater distortions than no indexation.

KEY CONCEPTS

Accelerated depreciation
Depreciation
Depreciation allowances
Exchange inefficiency
Indexing
Locked-in effect
Nominal capital gain
Real capital gain
Straight-line depreciation
True economic depreciation

QUESTIONS AND PROBLEMS

1. It is difficult to ascertain precisely the decline in the value of most assets as they grow older. An exception is automobiles. Assume that a new car costs $5000; that its value at the end of one year is $4000, at the end of two years $3000, at the end of three years $2000; and that it loses $250 in value for each of the following eight years. What is the true economic depreciation? What is the present discounted value of this, assuming a 5 percent after-tax interest rate? What will be the depreciation allowances under the current system? What is the present discounted value of these depreciation allowances? (Cars are treated as five-year assets.)

2. Supporters of accelerated depreciation in 1981 acknowledged that it favored heavy industry ("smokestack America") but argued that this was desirable. Why do economists tend to look askance at such arguments? Can you identify any major market failures? If it were decided to subsidize these industries, in what other ways might it be done?

3. The government has tried to encourage savings by allowing individuals to save a limited amount for their retirement, without facing taxes on interest. Assume individuals can put, say, $2000 a year in a retirement account (called an individual retirement account, or IRA), and that the interest would not be taxed. Draw the individual's budget constraint (between consumption today and consumption at retirement) with and without the IRA. Describe the income and substitution effects for (a) an individual who was planning to save a little, and (b) an individual who was planning to save a great deal. In each

case, what difference might it make if the individual has other assets, such as a savings account? Discuss the equity and efficiency consequences of changing the rules so that only amounts in excess of $2000 per year are afforded special tax treatment.

4. In the debate concerning repeal of the provision allowing capital gains on assets passed on to one's heirs to escape taxation, some have reasoned that death is not voluntary, and therefore one should not tax capital gains upon death. Evaluate.

5. Assume there are two "states of the world": in the good state, a risky asset yields a high return, whereas in the bad state, it yields a loss; the safe asset yields a zero return in both. Let C_g denote consumption in the good state, and C_b in the bad. Draw a figure with consumption in the good state on the vertical axis and consumption in the bad on the horizontal axis, then draw a 45° line. In the figure, let S represent the individual's consumption in the two states if he or she invests only in the safe asset (consumption in the two states is the same), while R represents the individual's consumption in the two states if he or she invests only in the risky asset (a higher consumption in the good state, a lower one in the bad). Explain why the line SR shows the individual's consumption possibilities—consumption in the two states depending on the proportion of the assets that he or she invests in the safe or risky asset. Now draw an indifference curve showing the bundles of consumption in the two states among which the individual is indifferent. Mark the point of tangency between, the indifference curve tangent and the consumption possibilities curve with the letter E.

   a. If E is halfway between S and R, what does this imply for how the individual allocates his or her portfolio?

   b. Now assume a 50 percent tax is imposed, with full loss offset. What happens to point S? Point R? Draw the new consumption possibilities locus, and describe what happens to E, and to the portfolio allocation.

   c. Assume now that losses are not deductible. What happens to point R? Draw the new consumption possibilities locus, and explain what happens to the portfolio allocation.

   d. Assume now that there are no taxes, but the safe asset yields a positive return. Show what happens to point S. Now, assume that there are taxes. What is the new point S? Use the diagram to analyze the impact of taxes on portfolio allocation with and without loss offsets.

6. In the text, we explained what happened to equilibrium investment in an open capital market, when a tax was imposed on the returns to capital received by Americans. Analyze what happens if a tax is imposed on the returns to capital whether received by Americans or by foreigners.

7. The text discussed the problems and distortions posed by inflation. In the late 1990s, the possibility of deflation has loomed large; some countries have actually seen falling prices. Describe the distortions and inequity associated with an unindexed tax system in the presence of deflation.

TAXATION IN PRACTICE

The next four chapters apply the general principles of taxation developed in Part Five to the analysis of taxation in the United States, placed in an international comparative perspective. Chapters 22 and 23 explain the major provisions of the personal and corporate income taxes and their implications for both capital and labor, and discuss some of the major policy issues that have confronted the United States and other countries over the past three decades. Chapter 24 focuses on tax avoidance, and Chapter 25 looks back at reforms since the 1980s and forward to future reforms.

22

THE PERSONAL INCOME TAX

The personal (or individual) income tax is the single most important source of revenue for the federal government. It is also the tax that impinges most on our lives. So important is the personal income tax that tax changes have headed the political agenda no less than six times during the past three decades: in 1981, 1986, 1993, 1997, 2001, and 2003. Before evaluating these tax reforms, however, we must understand the basic structure of the U.S. income tax, the principles underlying it, and the major problems of administering it.

OUTLINE OF THE U.S. INCOME TAX

There are four steps in the calculation of an individual's tax liability. The first is to calculate *gross income*. One adds up the total of wages and salaries, dividends and interest received, net income from one's business,

net rent (after expenses) from rental properties, and net gains from the sale of assets. Unemployment compensation is now taxed fully, and pensions are taxable to the extent that receipts exceed contributions on which tax has already been paid. Alimony received is also included in income. Gambling earnings, reduced by gambling losses, are included. Illegal earnings (such as from drug dealing), from whatever source, are taxable, although not generally reported. Several sources of income, though, are not taxable at all—among them child support, gifts and inheritances, interest on state and local bonds, some Social Security benefits, interest on life insurance, welfare, and veterans' benefits. None of these amounts is included in gross income. Benefits that employees receive from employers, such as health insurance and contributions to pension funds, are the most important exclusions from income for most individuals. Were these required to be reported, the amounts would be considerable. For instance, had individuals reported as income their employer-provided health insurance payments, tax revenues in fiscal year 2011 would have been $163 billion, or 7.1 percent, higher. Tax savings from the exclusion of pension contributions and earnings amounted to another $119 billion. In recent years, the tax savings associated with many other benefits, such as life insurance, have been reduced.

To get from gross income to **adjusted gross income (AGI)**, one subtracts contributions to certain tax-exempt savings plans, alimony paid, and a few other items (see Table 22.1).

To get from adjusted gross income to **taxable income**, there are two alternatives. One can either itemize personal deductions for—large medical expenses and casualty losses, mortgage interest, state and local taxes, charitable contributions, and moving and other job-related expenses—and then subtract the sum from adjusted gross income, or one can take what is referred to as the **standard deduction**, which is a set amount for different categories of taxpayers.[1] The point of the standard deduction is to simplify tax reporting for the majority of taxpayers, who have a limited amount of deductions. Thus, an individual whose itemized deductions were less than the appropriate standard deduction would simply use the standard deduction.

Both those who itemize and those who take the standard deduction are entitled also to deduct one or more personal exemptions. A taxpayer is allowed personal exemptions for himself or herself (and spouse, if filing jointly) and the family members he or she (or the couple filing jointly)

---

[1] In 2011, the standard deductions were $11,600 for a married couple filing jointly, $8500 for a head of household, and $5800 for a single individual or spouses filing separately. These amounts increase with inflation (they are indexed), so their real value remains fixed, although their nominal amounts change from year to year.

TABLE 22.1 CALCULATING TAX LIABILITIES

  Wages and salaries
  Interest income, dividends
  Net business income
  Net rental income
+ Other income

GROSS INCOME

− IRA contributions (when eligible), and contributions by self-employed to pension plans

− Alimony

− ½ of self-employment tax

− Part of health insurance premiums paid by self-employed for themselves and family

ADJUSTED GROSS INCOME

Alternative 1	Alternative 2: Itemized deductions
− Standard deduction	− Mortgage interest
	− State and local income and property taxes
	− Medical expenses in excess of 10% of adjusted gross income
	− Charitable contributions
	− Moving expenses (connected to relocation for employment)
	− Employee expenses (in excess of 2% of income)
	− Casualty losses
− Exemptions	− Exemptions

TAXABLE INCOME

× Tax rate

TAX LIABILITY

− Taxes previously withheld

− Tax credits (child care expense, foreign taxes paid, earned income tax credit, college tuition)

TAXES DUE

supports.[2,3] Personal exemptions had been phased out for high-income individuals. However, the 2001 tax act included the gradual elimination of the phaseout beginning in 2006, with complete elimination in 2010, after which the phaseout was to be reinstated in 2011 with expiration of the 2001 tax act.[4]

[2] In 2011, the amount was $3700 per exemption claimed.
[3] Difficulties in determining who should be allowed to take the exemption arise when individuals are supported in part by more than one taxpayer, such as children of divorced parents, or earn part of their support themselves. Writing down rules on how such issues should be resolved contributes greatly to the length and complexity of the tax code.
[4] In 2011, the personal exemption would have become zero for married couples filing jointly with AGI exceeding $376,850 and for single individuals with AGI exceeding $292,050, if the 2001 tax act had not been extended temporarily. The American Taxpayer Relief Act of 2012 included resumption of the personal exemption phaseout as scheduled, but raised thresholds above those set under previous law.

TABLE 22.2 FEDERAL TAX RATES, 2011

RATE	SINGLE		MARRIED FILING JOINTLY		HEAD OF HOUSEHOLD	
	FROM	TO	FROM	TO	FROM	TO
10%	0	8,500	0	17,000	0	12,150
15%	8,500	34,500	17,000	69,000	12,150	46,250
25%	34,500	83,600	69,000	139,350	46,250	119,400
28%	83,600	174,400	139,350	212,300	119,400	193,350
33%	174,400	379,150	212,300	379,150	193,350	379,150
35%	over 379,150		over 379,150		over 379,150	

The exemptions do not, of course, represent the additional cost of support for an additional person, which are typically far greater. Rather, they are intended, combined with the standard deduction, to ensure that no taxes are imposed on the very poor. Historically, changes in the minimum income below which no tax is imposed have roughly followed the poverty level.

Subtracting itemized deductions or the standard deduction, and the personal exemption, from adjusted gross income gives us taxable income (see Table 22.1). The basic tax liability may then be calculated. Like the standard deduction, the tax will depend on whether the individual is single, married filing jointly with a spouse, married filing separately, or a head of household.

The extra tax that an individual must pay as a result of earning an extra dollar of income is called the **marginal tax rate**. The income tax is derived from six marginal tax rates for ordinary income. These are shown in Table 22.2, as applied in 2011 to different levels of income. For example, a single person with taxable income of $50,000 paid a tax of 10 percent on the first $8500, 15 percent on the next $26,000, and 25 percent on the remaining $15,500. The tax applied to the last dollar earned—the marginal tax rate—is 25 percent.

The true marginal tax rate is somewhat more complicated than Table 22.2 indicates, because, as we noted, personal exemptions and a portion of itemized deductions are normally phased out as AGI increases. Had the 2001 tax act expired on schedule at the end of 2010, personal exemptions would have been reduced in 2011 by 2 percent for each $2500 (or fraction thereof) of AGI over $254,350 for couples and over $169,550 for single persons. As a result of the phaseout of exemptions, the top true marginal rate is approximately 39 percent (rather than 35 percent), and in the 33 percent bracket of Table 22.2, the true marginal tax rate is 37 percent.[5] Note that once the exemptions are

---

[5] The total value of exemptions for the family of four is $14,800. For each $2500 earned, the taxpayer loses $296 of exemptions. If the taxpayer is in the 33 percent tax bracket, that means he or she pays an extra $98 tax, which is equivalent to a marginal tax rate of 4 percent—thus, the marginal tax bracket is 37 percent.

completely phased out, the marginal rate falls back to the level shown in Table 22.2. As a result, higher-income individuals may face lower marginal tax rates than lower-income individuals.

In addition, for high-income taxpayers, itemized deductions are normally reduced by 3 percent of the amount by which income exceeds a specified threshold, up to a maximum of 80 percent of itemized deductions. This limitation on itemized deductions is commonly referred to as "Pease," named after the congressman who helped create the statute. If the 2001 tax act had expired at the end of 2010 as scheduled, the threshold in 2011 would have been $169,550—the threshold is the same for single filers, heads of household, and married couples filing jointly. Thus, a filer with taxable income of $300,000 and deductions of $30,000 faces a marginal tax rate of approximately 36 percent.[6]

To determine the ultimate size of the tax bill, another set of adjustments have to be made. These involve **tax credits**, which are direct deductions from the taxes paid to the government. Thus, if an individual owed $10,000 in taxes, but had tax credits for $900, he or she would send the government a check for only $9100. The 1997 Taxpayer Relief Act introduced a **child tax credit** amounting to $400 per child in 1998, which had risen to $1000 per child in 2011 (with the credit still phasing out for married couples filing jointly with incomes exceeding $110,000). The 1997 law also introduced a college tuition tax credit, which has evolved into two tax credits to help offset the costs of postsecondary education: the American opportunity tax credit, for up to $2500 per year per eligible student during the first four years of postsecondary education (degree or other recognized education credential studies); and the lifetime learning tax credit, totaling up to $2000 per year per tax return for all years of postsecondary education and job skills training (part-time studies allowable).[7] There are two other important tax credits: the earned income tax credit (discussed in Chapter 15) allows a low-income family a credit of up to 45 percent of earnings; and the **child and dependent care tax credit** allows a low-income family a credit of up to 35 percent of child and dependent care expenditures that enabled the taxpayer to work (or actively look for work), with the total amount capped at $3000 per year for one qualifying individual and $6000 for two or more qualifying individuals.

There is one more complexity to the tax calculations: a capital gain, the increase in the value of an asset between the time it was purchased and the time it was sold, is taxed at a special rate. In 2011, capital gains were taxed at 15 percent for upper-income individuals (above $69,000 for

---

[6] The loss of deductions increases the effective marginal tax rate by $0.03 \times 35 = 1.05$.

[7] The American opportunity tax credit has a $180,000 limit on modified adjusted gross income (MAGI) and the lifetime learning tax credit has a $122,000 MAGI limit, in both cases only if married and filing jointly; if single or head of household, the MAGI limits drop to $90,000 and $61,000, respectively.

married couples filing jointly and $34,500 for single filers) and 0 percent for those with lower incomes, on gains on assets held more than twelve months. Beginning in 2012, those with income over $400,000 faced a tax of 20 percent, plus a new 3.8 percent Medicare surtax. Those with an income between $200,000 and $400,000 faced a combined tax rate of 18.8 percent.

LEGISLATED VERSUS ACTUAL TAX RATES

Because of all the deductions, credits, and special provisions, the actual tax paid by individuals is markedly lower than the legislated (statutory) rate. The **effective tax rate** is defined as the ratio of tax payments to income. Effective tax rates increase with income, but because rich individuals typically have more opportunity to take advantage of the special provisions, the discrepancy between the "official" rate and the effective rate is larger for them. As a result, the actual degree of progressivity (the extent to which the ratio of taxes to income increases as incomes increase) of the income tax is less than that suggested by the tax schedule. Table 22.3 shows how the average effective tax rates have changed over time. The major tax reform in 1986 increased progressivity, lowering tax rates for the bottom four quintiles and raising the rate for the top quintile. The structure of the 2001 and 2003 tax cuts, however, decreased the degree of progressivity.

TABLE 22.3 AVERAGE EFFECTIVE INDIVIDUAL INCOME TAX RATES BY INCOME CLASS AND TAX YEAR (INCOME CLASS IN QUINTILES + TOP 1%; RATE IN PERCENT)

INCOME TAX QUINTILES	1985	1987	1996	1998	2000	2004	2011 (ESTIMATE)
Lowest	0.5	−0.6	−5.1	−5.4	−4.6	−6.2	−5.8
Second	4.0	3.2	1.8	1.5	1.5	−0.9	−2.9
Third	6.6	5.8	5.4	5.0	5.0	3.0	3.2
Fourth	8.8	8.1	7.9	7.9	8.1	5.9	7.0
Highest	14.0	14.9	16.1	16.5	17.5	13.9	14.9
All families	10.2	10.3	10.7	11.0	11.8	8.7	9.3
Top 1%	18.9	21.5	24.2	23.4	24.2	19.7	20.3

NOTE: Income class quintiles are based on comprehensive household income, which equals all pretax cash income, including taxes paid by businesses and employees' contributions to 401(k) retirement plans, plus all in-kind benefits. The 2011 estimate excludes in-kind benefits.

SOURCES: Congressional Budget Office, *Historical Effective Federal Tax Rates, 1979 to 2005,* (December 2007); and Urban-Brookings Tax Policy Center Microsimulation Model, *Average Effective Federal Tax Rates By Cash Income Percentiles, 2011* (February 2012).

A LOOPHOLE IN THE EARNED INCOME TAX CREDIT?

The expansion of the earned income tax credit was a major achievement of the 1993 tax law, benefiting more than 19 million low-income families. It went a long way toward achieving the goal of making work pay, of ensuring that all those who work full-time, even at a minimum wage, are able to work their way out of poverty.

One aspect of the expanded credit, however, though not yet seeming to have caused any problems, has given rise to considerable worry among some economists. Very–low-income individuals receive a credit of 45 percent of what they earn. Thus, if they earn $12,750, they receive a credit—a check in the mail from the government—for $5751. Consider the incentive this provides for two unemployed individuals. They hire each other to clean their houses, paying each other $12,750. They are honest folks, so they pay their social insurance contributions of 15 percent; each sends a check to the government of $1913. In return, they each receive back from the government a check for $5751, for a net gain of $3838. With a 45 percent subsidy rate, it is hard, in this scenario, to see why anyone does not report an income of at least the amount to make them eligible for the maximum earned income tax credit. Although the transaction may *look* fraudulent, in fact, the system encourages them actually to clean each other's houses, to convert what would have been an activity simply inside the house into a "market" activity.

When an individual earns an extra dollar, his or her tax goes up. The amount by which the tax goes up—the effective marginal tax rate—depends, of course, on all the deductions, exemptions, and credits, which, in turn, depend in part on how the individual receives the money and how he or she spends it.

OTHER TAXES

This chapter focuses on the federal individual income tax, but it is important to remember that there are other taxes, and that behavior is affected by the net effect of all taxes together. Many states impose a state income tax. The marginal tax rate that an individual faces includes both the extra taxes he or she pays to the federal government and those to the state government. In Chapter 16, we discussed the Social Security tax. Because for many individuals benefits increase with taxes, not all the Social Security contributions should be viewed as a tax. Medicare benefits do not depend on contributions, however, and thus Medicare "contributions" are taxes. Moreover, as we saw in Chapter 18, it makes little difference who actually sends the check to the government; both employer and employee contributions should be treated

**LEGISLATED MARGINAL
AND AVERAGE TAX RATES**

Marginal tax rates change
by jumps, as shown in Table
22.2, but average tax rates
increase gradually.
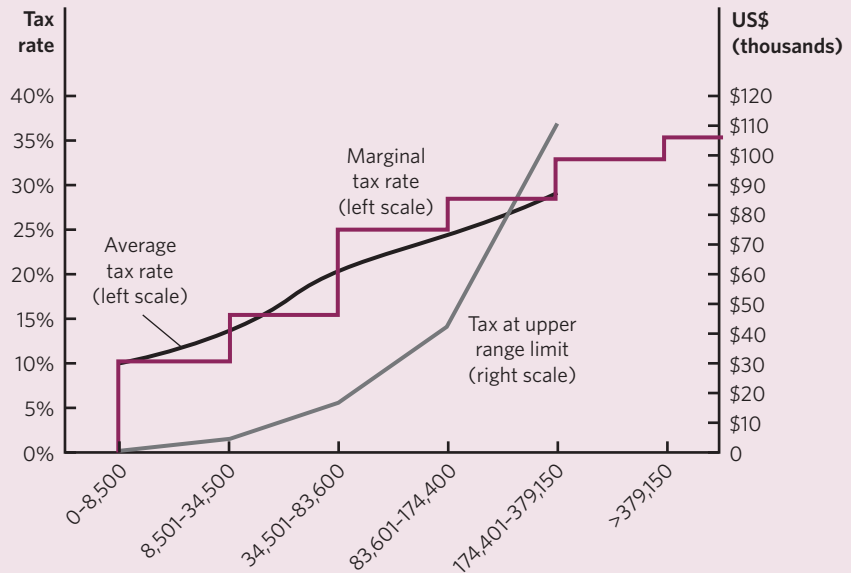
Note: "Single" filing status.

the same. Thus, the 2.9 percent Medicare tax (half paid by the employer) increases the marginal tax rate on an upper-income individual to as high as 42 percent, and when combined with state taxes, to 49 percent.[8]

For an individual's behavior, what is relevant is the impact of all the taxes and subsidies together—with all their special provisions. For instance, what matters for labor supply is the net marginal tax rate—how much a taxpayer's consumption can go up if he or she earns an extra dollar. Even focusing only on the principal taxes—the basic income tax combined with the earned income tax credit and Social Security—yields a marginal tax rate that varies markedly with income. Whereas the *legislated* marginal and average income tax rates for ordinary income appear schematically in Figure 22.1, the *effective* marginal and average income tax rates for adjusted gross income gives a far more complicated picture (see Figure 22.2). For example, the earned income tax credit results in a negative 45 percent marginal tax rate for incomes below $12,750, but then a positive marginal tax of 21 percent during the phaseout, between $21,800 and $49,078 (for a married couple with three children filing jointly). The picture is even more complex in Figure 22.3, which depicts average effective rates for all federal taxes for the same year (2003), including

---

[8] If the individual lives in a state that imposes a state income tax, such as California or New York, marginal tax rates may run at 10 percent or higher, but the state income tax is deductible from the federal income tax. Taking all this into account, a family making $300,000, living in a state with a 10 percent state tax, faces an effective marginal rate of approximately 49 percent: $39.5 + 10 (1 − 0.35) + 2.9$.
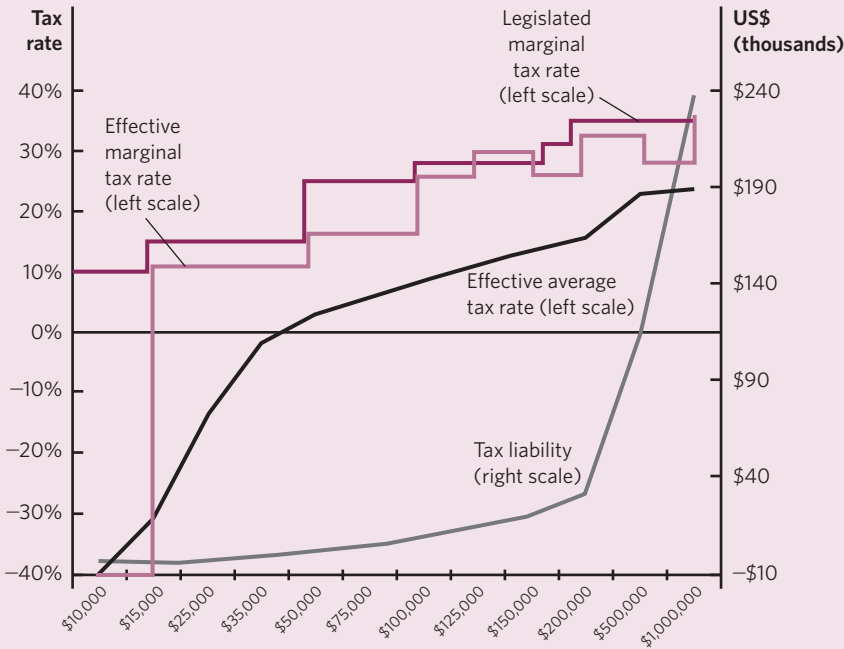
Note: "Married filing jointly" filing status with two children.
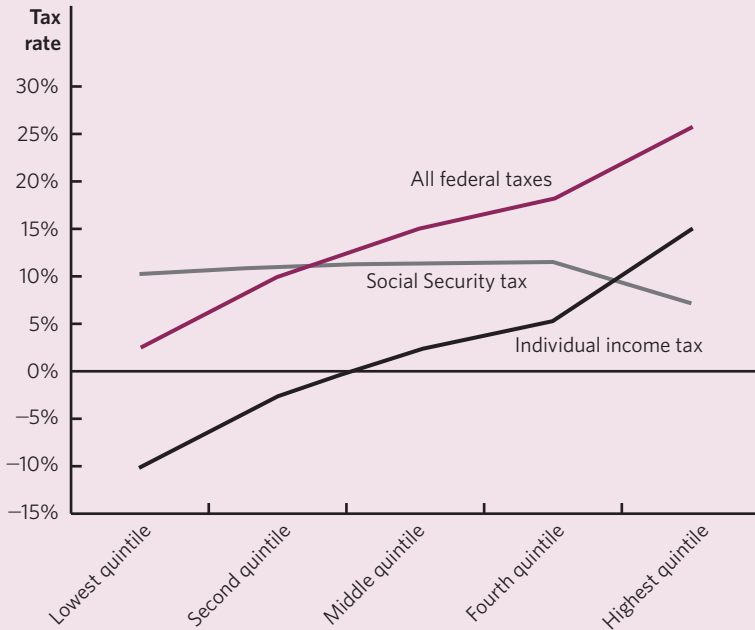
Note: Data for households with children.

TABLE 22.4 TAX LAWS MAKE AVERAGE EFFECTIVE TAX RATES LOW, WHEREAS LEGISLATED MARGINAL TAX RATES ARE HIGH

CALCULATION OF TAXES	INITIALLY		AFTER $3001 RAISE	
Adjusted gross income		31000		34001
Medical Expenses	5000		5000	
Less 7.5% of AGI	2325		2550	
Equals deduction for medical expenses		2675		2450
Job expenses	750		750	
Less 2% of AGI	620		680	
Equals deduction for job expenses		130		70
Interest deduction	3500		3500	
Total deductions		6305		6020
Personal exemptions	3700 × 3	11100	3700 × 3	11100
Taxable income		13595		16881
Tax (10 percent)		1360		1688
Child care tax credit				
26% of $5800		1508		
25% of $5800				1450
Earned income tax credit		2093		1461
Net tax liability		−2242		−1223
Average effective tax rate		**−7.2%**		**−3.6%**
Legislated marginal tax rate		**10%**		**10%**
Change in tax liability				1019
Ratio of change in tax liability to change in income				34%

social insurance (payroll) taxes, and uses the broader comprehensive household income base, rather than AGI, for calculating these rates.

The special provisions can have the effect of lowering the *average* effective rate at the same time as they raise the effective *marginal* rate. Table 22.4 illustrates how this can happen. It looks at an individual (in 2011) with two children and an initial income of $31,000 in the 10 percent bracket, with mortgage interest and medical, job, and child care expenses. Because of the deductions, initially the average rate is only −7.2 percent. After a $3001 salary increase, the average rate remains low at −3.6 percent, but the marginal tax rate is 34 percent—far higher than the legislated marginal tax rate of 10 percent.

PRINCIPLES BEHIND THE U.S. INCOME TAX

Several basic principles underlie the U.S. tax system—although because the tax law is so complicated and has evolved so much over the decades since it was first introduced, not every provision is fully consistent with these principles.

THE INCOME-BASED PRINCIPLE AND THE HAIG–SIMONS DEFINITION

The current U.S. tax code rests on the premise that the appropriate basis for assessing tax liability is the household's income (net of expenses personally incurred on the job).

For the most part, economists have argued that a *comprehensive* definition of income should be used that includes not only cash income (net of expenses required to earn the income) but also capital gains, whether the gain is realized or simply accrued. A number of other adjustments must be made to convert "cash" income into the "comprehensive" income that, in principle, should form the basis of taxation. This comprehensive definition of income is referred to as the **Haig–Simons concept**, after two early-twentieth-century economists who advocated its use. They believed that such a comprehensive income measure most accurately reflects "ability to pay."

There are three major differences between how our present tax system measures income and the Haig–Simons concept of "comprehensive" income.

1. *Cash-basis market transactions.* For the most part, only **cash-basis** market transactions are taxed. The tax thus rests on a notion of income that is narrower than that which most economists would ideally like to see employed. Certain nonmarket (noncash) economic activities are excluded, although activities that appear to be identical and are marketed *are* subject to taxation. For instance, a housekeeper hired to clean house has his or her compensation taxed, whereas a spouse who performs exactly the same services in his or her home (and whose support by the spouse working outside the home can be thought of, at least partially, as compensation for the services performed) is not taxed. If an individual owns a house and rents it out, the net rental income is subject to taxation; if the owner lives in the

house, no tax is due. The primary reason for this is the difficulty of determining appropriate values in the absence of market transactions; when there is a market transaction, there is an observable variable, the *transaction price*, which we can (and do) use to value the service.[9]

Some noncash transactions are listed in the tax code but are difficult to enforce. **Barter arrangements** (e.g., Sally Housepainter paints Joe Carpenter's house in return for his building her a new garage) are subject to tax. Also, when employers provide *in-kind payments* to their employees (e.g., making an automobile available for personal use), then, in principle, the employees are required to assess the value of these payments and report them on their 1040 forms. In fact, though, noncash payments often are not reported.

A major category of income that the tax system fails to trap is *unrealized capital gains*. Capital gains (the increase in the value of an asset) are taxed only when the asset is sold. Capital gains are taxed, in other words, only upon *realization*, rather than on an **accrual basis**—that is, as they actually occur from year to year.

To see why economists have argued that income should include capital gains, consider two individuals: one puts $100 savings in a bank and earns 10 percent interest, for $10 income; the other buys $100 worth of gold. During the course of the year, the price of gold rises by 10 percent, so at the end of the year, the individual could sell the gold for $110. The capital gain increases the wealth of this individual just as the interest payments do. From an economic point of view, the two individuals appear to have an identical ability to pay. But their tax liabilities differ: the individual who purchases gold has to pay a tax only on the capital gain when he or she sells the asset.

2. *Equity-based adjustments.* A second difference between our tax system's measure of income and true "comprehensive" income is that our tax system allows individuals who have large medical expenses or casualty losses to deduct a portion of those expenses from their income. The rationale here is fairness. These individuals are less able to pay taxes than someone with the same income, but without those expenses.

3. *Incentive-based adjustments.* Finally, the tax code is used to encourage certain activities, such as charitable contributions, by allowing tax credits or deductions for them. The tax code also allows the exclusion

---

[9] Some countries such as Sweden have attempted to tax the imputed "rent" on owner-occupied houses, as if the individual rented the house to himself or herself.

from income of most health insurance and life insurance premiums provided by employers to employees, presumably to encourage employers to provide these benefits.

Thus, whereas some capital income, such as owner-occupied housing, receives favorable treatment because of the absence of a cash market transaction, other capital income, such as money set aside in qualified retirement programs, receives favorable treatment because of a desire to encourage savings for retirement.

Those who advocate that consumption, not income, should be the basis of taxation argue that we should move away from the Haig–Simons attempt at a comprehensive definition of *income* toward a comprehensive definition of *consumption*. For consumption tax advocates, the failure of the current tax system is not so much that capital income escapes taxation but that so much of capital income is taxed at all.

THE PROGRESSIVITY PRINCIPLE

Our tax structure is based on the premise that those with higher incomes not only should pay more in taxes, but should pay a larger fraction of their income. The effect of the differences between comprehensive income and the tax definition of income is to reduce significantly the effective degree of progressivity, as we have seen. To limit the extent to which individuals can avail themselves of these loopholes, Congress passed a **minimum tax**, the intent of which is to ensure that upper-income individuals pay a tax at least equal to 26 percent of their income (in excess of a basic exemption level).

Whereas the 1986 tax reform pared back many of the special provisions, the 1997 reform brought back some of the old ones and introduced some new ones. The 1997 law provided something special for almost everyone: a child tax credit for those with young children; a tuition tax credit for those with college-age children; and lower taxes on capital gains, a provision of special benefit to the very wealthy. The result is that every income category saw its effective tax rates going down, by between 0.9 and 1.7 percentage points. However, whereas the percentage point reduction for someone in, say, the $50,000 to $75,000 bracket was larger than for someone in the $100,000 to $200,000 bracket, the dollar value of the tax savings for the latter was obviously much greater. A millionaire would see his or her taxes going down $15,000, whereas someone with a $25,000 income would see his or her tax bill cut by $350.

Interestingly, the 1997 law undid what had been viewed as one of the major achievements of the 1986 tax reform, which had basically eliminated the preferential tax rates for capital gains. Indeed, prior to 1986, the value of these benefits and other tax loopholes that the rich had taken advantage of was so great that those with incomes over $1 million actually faced a lower effective tax rate than those with an income between $500,000 and $1 million.

THE FAMILY-BASED PRINCIPLE

The basic unit of taxation in the United States is not the individual, but the family. Two individuals who decide to get married (and thus change their family status) will find that their tax liabilities are altered. The tax code attempts to make some limited adjustments for families in different circumstances. Families in which there is only one adult are taxed at a rate halfway between the rate of an individual and the rate of a two-adult family. Families with children are allowed exemptions for each child, as mentioned earlier. Families in which both parents work outside the home are allowed a credit for child care. Under the current tax code, individuals who get married have the choice of still filing as if they were single ("married filing separately"), so marriage will not make them worse off (see the example in the next section).

Although the tax system is essentially family based, it is not completely so: children can file their own tax returns so that their income, whether earned or investment income, could be taxed at a lower rate than it would be if included in their parents' tax return. The United States is now one of the few countries still employing a family-based tax system. Other countries, such as Canada, have an individual-based system, under which each individual is taxed on his or her own income.

Divorce presents problems for a family-based tax system: Which parent should claim a deduction for supporting the child, when both provide *some* support? How should the payments from one divorced parent to the other (alimony) be treated? (If the couple were still married, a transfer from one to the other would not affect taxes.) Under present provisions, alimony (but not child support) is deductible by the party paying it and taxable to the party receiving it.[10]

---

[10] Thus, with the progressive tax structure, if ex-husband and wife are in very different tax brackets, it pays to label payments that are really child support as alimony. Like everything else in modern life, getting divorced in a manner that minimizes tax liabilities requires care and thought.

UNIT OF TAXATION The unit of taxation makes a difference because of progressivity. In effect, our tax system used to impose a tax on marriages between couples of similar incomes and a subsidy on marriages between couples of disparate incomes. The increased tax payment resulting from getting married is sometimes referred to as the *marriage penalty*, or **marriage tax**; conversely, the reduced tax payment can be thought of as a *marriage bonus*, or **marriage subsidy**. The current tax code eliminates this "marriage penalty" for couples at similar income levels, although it continues to provide "marriage bonus" for couples at significantly different income levels. To see this, we contrast the effects of marriage on two different couples; the results are summarized in Table 22.5.

Abigail and Billy currently are living together but are not married; each earns $30,000, and in 1997, they would have paid a total of $6960 in income taxes. Had they married in 1997, their joint tax liability would have increased by $1068 (to $8028). If they anticipate remaining married for, say, fifty years, and do not anticipate any change in their salaries (after adjusting for inflation) over that period, the present discounted cost to them of getting married would exceed (with a 5 percent real interest rate) $20,000. If Abigail and Billy have some doubts about whether to get married, this calculation might well resolve them.

By contrast, when Amy marries her low-paid boyfriend, Bradford, they find that their total tax liabilities are reduced by $1084. For this couple, the tax system acts to encourage marriage. Assuming fifty years of marriage, and that their income ratios remain constant, the value of the government's subsidy to this couple's tying the knot is more than $20,000.

Was it the intent of Congress, in enacting the tax code, to encourage marriages between individuals with very different incomes and to discourage marriages such as that between Abigail and Billy? Probably not. Thus, the current tax code tries to be "marriage neutral."

Consider now what happens after we have changed the tax code to maintain the option of filing either separately or jointly but now try to

TABLE 22.5 TAX EFFECT OF MARRIAGE: "MARRIAGE PENALTY" OR "MARRIAGE BONUS"?

FILING STATUS		EARNINGS	TAX ON INDIVIDUAL		TOTAL TAX OF A&B	
			1997 Tax Law	2011 Tax Law	1997 Tax Law	2011 Tax Law
Single	Abigail	30,000	3480	4075	6960	8150
	Billy	30,000	3480	4075		
Single	Amy	12,000	780	1375	9112	9500
	Bradford	48,000	8332	8125		
Married	A&B	60,000			8028	8150

ensure that individuals neither benefit from nor are penalized by marriage. This does eliminate the discrimination against those who choose to live together under the bonds of matrimony for couples with similar incomes, but there is still an advantage for couples of very different incomes who file jointly.

Abigail and Billy now have the same tax liability of $8150 whether or not they are married. However, Amy and Bradford now pay $1350 more when not married because Amy's income places her in the 25 percent tax bracket when filing individually, whereas their combined income is in the 15 percent tax bracket when filing jointly.

Not surprisingly, while Amy and Bradford are better off married than not married, they think the current tax system is unfair. Shouldn't the family's total tax burden depend simply on family income, not on how much each member of the family earns? Why should Amy and Bradford pay more taxes than Abigail and Billy? And their friends Mark and Allen, who live together but are not allowed to marry in the state in which they live, also feel aggrieved. They have the same incomes as Amy and Bradford, respectively, but have to pay much more in taxation.

No tax arrangement appears to be "fair" in all circumstances. However, do the general theories of fairness we discussed in Chapter 17 provide any guidance? The ability-to-pay approach suggests that two families (with the same number of children and both parents working) with the same income ought to pay the same taxes. Because the costs of two individuals living together are much lower than twice the costs of two individuals living singly, the ability-to-pay approach would suggest that whenever two individuals cohabit, they should be subjected to higher taxation than if they live singly.

Unfortunately, the tax authorities cannot easily monitor cohabitation; as long as the vast majority of cohabitators are married, and as long as most married individuals live together, basing taxes on whether individuals are married (which is easier to ascertain) rather than on whether they cohabit (which is not easy to ascertain) does not create too many inequities. At the time the tax code was first adopted, it clearly reflected the vast majority of cases. Moreover, the inequities may not have been too large when most American households had a similar structure—with one wage earner. Today, however, most women work outside the home, and there is a wide variety of household structures.

The utilitarian approach attempts to ascertain how the family circumstances in which individuals find themselves affect their marginal utility of income.

In both the utilitarian and the ability-to-pay approach, one might want to distinguish between families with two wage earners and those

with one. Assume the families have the same total income. The current tax system treats them alike. However, the family with both individuals working may have to purchase many services that the nonworking spouse provides free. Both those who believe in ability to pay and those who believe in utilitarianism might well argue that a family with two working individuals should pay a lower tax than a family with one worker.

Because the marriage penalty seems so patently inconsistent with "American values," it has long been a major focus of criticism, and advocates of tax cuts have repeatedly put forward proposals for its elimination or reduction. While nothing was done, the Tax Relief Act of 2001 mitigated many of the effects of the marriage penalty by making adjustments for married couples of the standard deduction; the 10, 15, and 25 percent tax brackets; and the child and earned income tax credits.

THE ADVANTAGES OF A FLAT-RATE TAX SCHEDULE   There is one—and only one—way to avoid the inequities surrounding the choice of unit of taxation: impose a **flat-rate tax schedule**. If all individuals pay a proportional tax on income in excess of a basic exemption level (and if those with an income below this exemption level receive a cash payment from the government), there is no penalty and no reward for marriage, and no reward for divorce. However, this comes at a high cost, because a flat rate schedule reduces the tax system's progressivity.

THE ANNUAL MEASURE OF INCOME PRINCIPLE

The U.S. income tax is based on annual, not lifetime, income. Consequently, two individuals with the same lifetime income may, over their lifetimes, pay quite different taxes. The individual who decides to postpone more of his or her consumption until retirement will, for instance, pay more in taxes than his or her less frugal counterpart. Or consider two individuals with the same (before-tax) present discounted value of income, one of whom is a late bloomer, earning most of his or her income in later life. The present discounted value of this indivudal's tax payments will be lower.

Because of the progressivity of the tax structure, the use of an annual measure of income also affects differently those with stable incomes and those with fluctuating income. Middle-income families with variable income are adversely affected. Consider a family of four whose average adjusted gross income is $28,600, but in half the years it has an income of $38,600 and in the other years it has an income of only $18,600. Using the 2011 tax law, in a good year, the family will be taxed at a marginal rate

of 15 percent; in a bad year, it will be taxed at a marginal rate of 10 percent. Its total tax liability will be greater than that of a family with a stable income of $28,600 that always faces a 10 percent marginal tax rate; the *average* additional annual tax payment of the family with a fluctuating income is $250.[11]

Prior to the 1986 Tax Reform Act, taxpayers were allowed to average income over periods when their income varied widely. By eliminating the privilege of income averaging, the tax reform aggravated the distortions produced by the annual basis of taxation. The main reason for dropping the averaging provisions is that they do cost the Treasury money; as the government tried to maintain budget neutrality as it reformed the tax code—in a way that would have some degree of popularity—provisions like income averaging that made good tax policy sense, but did not have large political constituencies behind them, were sacrificed.

PRACTICAL PROBLEMS IN IMPLEMENTING AN INCOME TAX SYSTEM

In translating the basic principles of the income tax into workable tax law, there are three extremely difficult problems: determining what "income" is; determining when somebody has received some income; and deciding what deductions from income to allow.

DETERMINING INCOME

For most wage earners, determining income for tax purposes is a simple matter: they add up their paychecks, interest, dividends, and so on. For taxpayers who run their own businesses, however, it is not. There are two central problems. The first has to do with determining depreciation (the loss in value of machines and buildings as they age) and adjusting the cost

---

[11] We assume that the family has no adjustments to income and takes the standard deduction. Then, in the good year, the extra tax payment is $0.15 \times \$10,000$; in the bad year, the reduced tax payment is $0.10 \times \$10,000$. The total extra tax payment over a two-year period is thus $500; dividing by 2, we obtain the average annual extra payment.

of inventories for inflation. The second problem is differentiating between consumption expenditures and legitimate business expenses.

The tax code recognizes that legitimate expenses required in order to earn a living ought to be deductible from an individual's income. The principle seems clear. Surely, a store owner who sells candy should not be taxed on the total value of sales; owner's expenses—the rent for the store, the purchase of candy from the candy manufacturer, the salaries paid to employees—should all be deducted from sales to calculate gross income. What about the candy the owner consumes while working? The owner may claim that the consumption of candy is a form of advertising; when customers see him or her eating candy, they increase their purchases. However, what about the candy that the owner consumes when no one is around? The owner may claim he or she is "testing" various samples, to ensure the quality of the candy he or she sells. One might suspect that neither of these explanations is the owner's real motive for eating candy. He or she simply likes candy.

Similarly, in many businesses there is very little difference between advertising and entertainment expenses. Taking clients to dinner is a method of persuading them to buy one's product, just as putting an advertisement in the newspaper is an attempt to persuade customers to buy one's product. On the other hand, there are other instances in which "business entertainment" is purely a matter of having a good dinner partly at Uncle Sam's expense.

These examples illustrate the two central problems:

1. In many instances, it is impossible to ascertain what are legitimate business expenses and what are not.

2. Even when the distinctions between legitimate and illegitimate business expenses are conceptually clear, performing the required monitoring is often impossible. Returning to our earlier example, it is difficult to imagine the kinds of records that would be required to isolate the owner's consumption of candy (if we decided that consumption is not a legitimate business expense).

CONSEQUENCES OF ALTERNATIVE BUSINESS-EXPENSE RULES

It is impossible to devise a system of distinguishing between legitimate and illegitimate expenditures in a way everyone would consider to be fair. Someone always either is unfairly burdened or benefits unfairly, no matter what rule is devised.

Either the government can allow a fairly generous treatment of expenses—for instance, for travel—in which case the individual who is really traveling for recreation purposes is unfairly receiving a tax benefit. On the other hand, the government can be fairly restrictive—for instance, by not allowing first-class travel or meal deductions above a certain amount—in

which case the individual who has no recreational motive may be unfairly burdened. There is no way the tax code can be fair to both these individuals. Moreover, any rule induces economic distortions. If deductions for travel expenses are restricted, businesses requiring travel will be discouraged; if travel expenses are not restricted, businesses in which there is scope for hidden pleasure travel may be encouraged. The deductions are a form of tax-exempt income. Furthermore, if deductions for travel expenses are restricted, businesses may substitute less efficient communications methods for travel. This is because the relative after-tax price of travel will rise if travel expenses are not deductible, but other communication expenses (telephone, fax, etc.) are still deductible (see Figure 22.4A).

In Figure 22.4B, we consider a self-employed individual who is able to claim business entertainment as a deduction. The deductibility of these expenses shifts the individual's budget constraint—the alternative combinations of "entertainment" and "other consumption goods" the individual can purchase. His or her before-tax budget constraint is E_0C_0; with no deductibility, it is E_1C_1. When entertainment is deductible, it is E_0C_1. Entertainment becomes relatively less expensive; if the individual is in the 35 percent tax bracket, he or she has to give up only 65 cents' worth of other goods to get a dollar's worth of entertainment. The individual's consumption decisions are clearly distorted.

In the 1986 Tax Reform Act, Congress took an intermediate position: tax deductions for luxury cars were reduced, and only 80 percent of entertainment expenses and meals were deductible. In 1993, this was reduced further to 50 percent, which still applies in 2011.

WHAT CONSTITUTES A BUSINESS? Not only is it difficult to determine what are legitimate business expenses, but in some cases, it is even difficult to determine what is a business. For instance, individuals who raise horses could be raising horses as a business. On the other hand, they could be keeping the horses simply for their own pleasure. If they buy a horse, keep it for several years, sell it, and take a loss, the loss is really not on a business activity or on an asset, but on an ordinary pleasurable activity. One could argue that there is no reason why their capital loss should be deducted from their income tax. On the other hand, there are individuals who do earn their living raising horses—buying them at a lower price, feeding them, and then selling them at a higher price. Not to allow these individuals who are in the business of raising horses for profit to deduct their losses would seem to be grossly unfair. However, it is virtually impossible to distinguish between the two situations.

The government attempts to combat this kind of tax avoidance by insisting that serious businesses make a profit. The rule of thumb is that an individual should make a profit in at least three years out of five; in the case

of horse breeding, this rule is relaxed to at least two years out of seven. Such rules obviously do succeed in reducing the amount of tax avoidance. At the same time, of course, there are individuals who are seriously in business who make losses year after year. Setting up a business often takes three or four years in order to establish a reputation—and a profit.

EMPLOYEE BUSINESS DEDUCTIONS   In principle, the "necessary costs" of working should be deductible. The difficulty is ascertaining what are necessary costs. Because of the impossibility of doing this on a case-by-case basis, the government has set up certain basic rules—for example, some

educational expenses are deductible, but most are not; moving expenses connected with a job are deductible, but commuting expenses are not.

In drafting the 1986 Tax Reform Act, Congress in effect decided that the old law had resulted in too many individuals' claiming employee business expenses for what were actually ordinary consumption expenditures. Accordingly, they decided that only taxpayers with large unreimbursed employee business expenses—exceeding 2 percent of adjusted gross income—could deduct them.

To the extent that expenses incurred as a result of going to work are not deductible, not only is an inequity created, but incentives to work are reduced. If a job requires clothing expenditures that the individual would not otherwise have incurred, for example, the net return to working is reduced. Let's say Bill earns $15,000 a year, and has work-related expenses of $5000 that are not deductible, for a net income of $10,000. Because he has inherited a modest-sized fortune, he is in the 28 percent marginal tax bracket, and the government requires him to pay 28 percent of the $15,000 total—that is, $4200—in taxes. This amounts to 42 percent of his net earned income of $10,000; the $5800 he receives yields a return below the minimum wage, and is hardly enough to motivate him to go to work.

CHILD CARE EXPENSES   Child care expenses present a similar problem. In one sense they are voluntary—expenses resulting from a family's decision to have children. For a family that has had children, however, they are expenses that can be avoided only by having one parent stay at home.

Present tax law allows a credit against child care expenses (up to $3000 for one child or $6000 for two or more children), paid to allow the taxpayer (and spouse, if they file jointly) to work or to look for work. The amount of the credit ranges from 35 percent of expenses for working parents with an adjusted gross income (AGI) under $15,000, to 20 percent of expenses for parents with AGI over $43,000. Therefore, the maximum credit is: $0.35 \times \$6000 = \$2100$. If a woman with a child goes to work and has to pay someone to take care of her child, the family's net income is just the difference between what she receives and what she must pay out. If she receives $15,000 and must pay out $5000 in child care expenses, the family's net income is only $10,000. Not allowing the deduction of these expenses creates strong distortions. Assume the woman is married, and as a result of her husband's income, 33 percent of what she earns goes to the IRS and her credit is limited to 20 percent. If she pays $5000 for child care, the total increase in what the family can spend on other things (after taxes) is only $5050.[12] She may well be discouraged from taking the job. Allowing a credit of $600 ($0.2 \times \$3000$) goes only a little way toward fully

---

[12] The additional amount the family has to spend will be further reduced by Social Security taxes and, perhaps, state and local income taxes.

alleviating the distortions or correcting the inequities. On the other hand, the child care tax credit creates a distortion itself. By lowering the price of child care, it encourages a greater consumption of child care.

This is an example of the more general problem arising from the fact that the income tax is based only on *market transactions*. Activity that occurs within the household—and is essentially identical to that purchased in the market—is not taxed.[13]

Both the inequities and the distortions arise from the inability to measure (and therefore to tax) the household services provided within the family. The failure to tax the "imputed" value of household services discriminates against the purchase of those services in the market; it encourages the production within the household.

Decisions concerning the tax treatment of child care expenditures not only have economic consequences; these decisions also reflect, and have consequences for, social values and family structure. Family life when both parents work is different from that when one parent (usually the mother) remains at home. A tax system that penalizes women who enter the marketplace may be thought to reflect or perpetuate a particular set of attitudes concerning the role of women.

EMPLOYEE BENEFITS   For many individuals, a significant fraction of their compensation comes in forms other than direct cash payment, especially in employee benefits. The most important of these is medical insurance. Some workers receive thousands of dollars' worth of medical and dental benefits that are exempt from taxation. Most economists believe that such benefits should be included within taxable income. As Congress and the administration have debated how to slow down the rising costs of medical care, even though it has been recognized that these special provisions have the effect of encouraging health care expenditures, political pressures (both from unions, which had fought hard to achieve these benefits, and from the health care industry) have kept the issue largely off the agenda. This is a vicious circle: given the high cost of medical care, having good health insurance is viewed as essential for survival.

SOME CONCLUSIONS   Three lessons can be learned from this discussion of the practical problems of defining "income" for tax purposes. First, what may seem like minor details of a tax law can have important consequences. Second, many (but by no means all) of the provisions of the tax code that seem unfair and distortionary are not the result of politicians' representing special interest groups, or of bureaucrats' incompetence. There are real difficulties in determining what is income and

---

[13] In spite of this, there is a strong trend for goods that used to be produced inside the household to be produced in the market. The reason why the distortion does not appear larger is that there are sufficient economies from market production that overcome the tax distortion.

what are legitimate business expenses. Third, whatever rules are chosen will entail both some inequities and some inefficiencies. Designing the tax code necessitates weighing one inequity against another, one distortion against another. The objective of our discussion has been to clarify these trade-offs and to help explain the all-too-frequent situations in which lawmakers discover that in the process of correcting one inequity or distortion, they have created a new one, as bad or worse than the first.

TIMING

The second important practical problem of implementing an income tax is determining when somebody has received some income. Again, for most wage income, there is no problem. However, for instance, consider an author writing a book. Authors commonly are paid a royalty, a certain fraction of the revenues generated by sales of the book. Usually, the publisher provides an advance payment prior to the publication of the book in anticipation of future royalties. In principle, if the book fails to sell, the advance must be returned. Should the advance be treated as income to the author at the time he or she receives it? Or should it more properly be treated as a loan, which will be repaid with the proceeds of the book? In the latter case, the author would have to pay the tax when the book has been sold.

There are, in fact, many transactions that are, or can be made, to take on this form. Consider a contractor building a building. The contract is not fulfilled unless, and until, the building is completed. However, the contractor receives payments as the building progresses. Are these payments to be treated as loans, in which case the builder records the income only when the building is completed? Similar issues arise in any long-term project, such as defense contracts to develop a new airplane.

People care about timing because a dollar today is worth more than a dollar tomorrow. The **present discounted value** of tax liabilities is reduced by postponing the tax.

On the other hand, these issues of timing are of less consequence for those who believe that consumption, and not income, is the appropriate basis of taxation. Timing is important because, in effect, it allows individuals to escape interest income taxation, which, from the perspective of consumption taxation, should not be taxed in the first place.

PERSONAL DEDUCTIONS

The third practical problem in implementing an income tax is deciding which deductions from income to allow. Thus, in arriving at taxable income

from adjusted gross income, the government allows deductions that are designed to make a more equitable tax system and to encourage certain socially desirable activities. There are five important kinds of expenditures for which deductions are allowed: medical expenses, mortgage interest, state and local taxes, charitable contributions, and casualty losses.

MEDICAL EXPENSES   The motivation for allowing medical expenditures to be deducted from income seems clear: Health problems lead to costly medical bills and often reduce the individual's earnings as well. An individual who is spending all of his or her income on doctors' bills simply to stay alive has a lesser ability to pay than an individual with the same income but no medical expenses. Ability to pay is measured best not by total income but by *discretionary* income, the amount in excess of the amount required to survive.

This argument has been criticized on two grounds. The first is that there are other categories of expenditures such as food that, at least at some level, are equally necessary. However, differences in the necessary amount of food are likely to be smaller than differences in the necessary amount of medical expenditures. The second is that a significant fraction of medical expenses are discretionary (e.g., staying in a private room rather than a semi-private room, having a television set in one's room, plastic surgery to stay young-looking, etc.), and the law does not distinguish between "necessary" and "discretionary" expenditures. Again, however, this is understandable, as the distinction, though clear in principle, is virtually impossible to make in practice. The tax rules now allow deductions for medical expenses only to the extent that they exceed 10 percent of adjusted gross income. This seems to reflect the judgment that significant inequities in ability to pay arise only with significant medical costs, and that these large medical costs are likely (though not always) to be nondiscretionary.

The provisions for deducting medical expenses effectively reduce the price of medical services. For someone in the 28 percent marginal tax bracket whose medical expenses exceed the 10 percent minimum, the private cost of an extra $100 of medical services is only $72. The distortions this introduces are obvious: to the extent that medical expenditures are discretionary, the individual has an incentive to spend too much on medical services (relative to other commodities). The amount by which the effective price of medical services is reduced depends on the individual's marginal tax bracket. For an individual in the 15 percent bracket, an extra $100 of medical services costs $85. Whenever individuals face different prices for the same commodity, there is inefficiency.[14] On the other hand,

---

[14] There are further distortions associated with health expenditures. Employer-provided health insurance is not taxed, and self-employed individuals can also deduct health insurance premiums. This encourages the purchase of health insurance, and insurance, by lowering the cost of obtaining medical services, leads to excessive consumption of certain health services (as we saw in Chapter 13).

most individuals will not increase their demand for heart replacements simply because the price is lower.

Aside from the inefficiencies introduced by the medical expense deduction, this deduction has been objected to on grounds that it is unfair. The reduction in the tax liability as a result of, say, $1000 of medical expenses for an individual at a higher income level is greater than that for an individual at a lower income level. Thus, if an individual in the 28 percent bracket incurs a $1000 medical expense (ignoring, for the moment, the provision limiting the amount that can be deducted), his or her tax liability is reduced by $280. On the other hand, an individual in the 15 percent bracket would have his or her tax liability reduced by only $150. The *value* of the provisions for deductibility of expenses is thus much greater for the individual in the higher bracket.

The efficiency and equity arguments against deducting medical expenses (which apply to other categories of deductions as well) have led many economists to conclude that credits are preferable to deductions. With a tax credit, an individual with a $1000 medical expense would have his or her tax reduced by the same amount, regardless of income. If there were a 20 percent tax credit, the government would, in effect, be paying 20 percent of medical expenditures. Not only does this seem fairer to many individuals, but it also seems more efficient. Recall that one requirement for the Pareto efficiency of the economy is that all individuals' marginal rates of substitution between different commodities be the same. (This was called the *principle of exchange efficiency*.) This is ensured if all individuals face the same price for every commodity. With tax credits, as the government is, in effect, paying part of medical costs, individuals do not pay net (after tax) the true marginal cost of the medical services they obtain, but all individuals face the same effective price. With deductions, as we have noted, upper-income individuals face a lower price than lower-income individuals.

The preceding arguments are not completely persuasive. Recall that the motivation for allowing the deduction was to base taxes on some measure of ability to pay. It was believed that medical expenses reduce the individual's ability to pay. If this is true, gross income by itself does not provide an appropriate basis for judging ability to pay. Gross income, net of "involuntary" medical expenses, provides a better measure. Although we know we cannot separate out "voluntary" from "involuntary" medical expenses, we may believe that it is better to take account of medical expenses than to ignore them; that is, we may believe that gross income, net of all medical expenses, is a better measure (though admittedly imperfect) than gross income alone.

It is, of course, true that the deductibility of medical expenses produces distortions, particularly for individuals in the 28 to 35 percent marginal brackets. As we discussed earlier, there is an important equity/efficiency trade-off. The fraction of medical expenses that should be tax deductible

would presumably depend on the elasticity of demand for medical services. As we noted earlier, if the elasticity is low, the distortion is low, and there will be little gain in efficiency from making medical expenditures only partially deductible. On the other hand, if the elasticity of demand for medical expenditures is high, there are potentially great gains in efficiency, and then we may wish to have only a fraction of medical expenses deductible.

INTEREST The motivation for the tax deductibility of interest is simple: income (as usually defined) includes wages plus net interest receipts—that is, the difference between interest paid and interest received. If we believe that an individual who has positive net interest receipts has a higher ability to pay than someone with no interest receipts, and that an individual with negative net interest receipts has a lower ability to pay, interest paid should be tax deductible.

Those who believe that consumption is the appropriate basis of taxation argue that interest should not be taxed and that interest payments should not be tax deductible. Those who believe that income is the appropriate tax base, however, have been concerned about the inequities and inefficiencies to which interest deductibility gives rise. First, because some types of capital income receive favorable tax treatment, borrowing to finance favored types of investments provides one of the major classes of tax avoidance devices. Second, deductibility of interest payments encourages borrowing, and thus discourages saving. In the aftermath of the financial crisis, another concern has been raised: excessive debt can lead to economic vulnerability. Tax deductibility of interest encourages debt, and thus can be destabilizing.

The 1986 Tax Reform Act took a compromise position that is still in effect today. It continued the tax deductibility of mortgage interest (for up to two homes), but eliminated the deductibility of all consumer interest. Money, however, is fungible; money borrowed allegedly for one purpose can be used for another. An individual who is buying a house and a car, and who was planning to borrow money to pay, say, 80 percent of the cost of each, can obviously borrow a little more against the house, using the additional funds to pay cash for the car. Banks also make home equity loans, enabling an individual to borrow against the current market value of his or her house, which is often considerably more than the purchased price. These are treated as mortgages, and interest on these loans is tax deductible. The restriction on the deductibility of consumer interest is thus likely to encourage this trend.

Congress was aware of these problems, and partially addressed them by limiting the size of the mortgage for which interest is deductible to the original purchase price plus borrowing for medical and educational purposes. Other limits include a maximum debt of $1 million if the loan is for home acquisition and the taxpayers are married and filing jointly,

and up to a debt of $100,000 for home equity loans if married and filing jointly). The effectiveness of these provisions on the limitation of interest deductibility is questionable, however.[15]

STATE AND LOCAL TAXES   State and local taxes are deductible. The primary motivation for this provision is the concern that without such a provision, the imposition of federal taxation would seriously impair the states' ability to raise revenues. Indeed, without such a provision, during World War II, when federal marginal tax rates reached as high as 94 percent, some individuals would have faced total marginal tax rates (combining federal and state taxes) in excess of 100 percent. Some individuals also have expressed a concern that without such a provision, there would effectively be double taxation of the same income. Whether such double taxation is inequitable is, however, another question. If the taxes are thought of as being associated with the benefits of living in a particular locale, it is not obvious that these expenditures should be treated any differently from expenditures on other goods and services.

Indeed, the deductibility of local taxes may give rise to an important source of distortions and inequities. Many services provided by local communities—garbage collection, sewage disposal, education, tennis courts—differ little from services that can be purchased privately. Such services provided by local communities are known as "local publicly provided goods": all members of the community benefit from these goods, but those outside it do not. Thus, the deductibility of local taxes encourages the public provision of these goods and services (regardless of whether the services might be more efficiently provided privately) and encourages the consumption of those goods and services that can be provided through local communities.

Not all economists agree on the significance of these distortions or inequities. Although many goods provided by local communities are much like privately provided goods, many are basically public goods, little different from those associated with private charities for which a deduction is allowed; for instance, the elderly do not benefit directly from the expenditures on education (although in some instances they may benefit indirectly, from the increased value of their house). In the voting models discussed in Chapter 9, the outcome of the political process depends critically on the median voter, the one who is such that half the voters want more public expenditures, and half less. In that case, what is critical is how the tax system affects the median voter. In most states and communities—but not all—the median voter

---

[15] The provision limiting the size of the mortgage to the original purchase price plus borrowing for medical and educational purposes has an additional potential distortion: an individual whose house has increased in value will have an incentive to sell the house and buy a house of equivalent value. This is especially true as a result of the current tax law provision that allows the capital gain to go untaxed (up to $500,000 for a married couple).

does not itemize; that is, according to the median voter model, deductibility is essentially irrelevant. Still, many economists, and almost all politicians, were concerned that the elimination of the deductibility of state and local taxes would decrease the demand for state and local goods.[16]

CHARITY   The deduction for gifts for charitable purposes—for education, religion, health, and welfare—is one of the more controversial provisions of the tax code. The motivation for allowing the deduction of these expenditures is clear. As we discussed at greater length in Chapter 5, there are insufficient private incentives for spending money on goods that generate benefits to others. Money spent to develop a polio vaccine, for instance, may yield little direct benefit to the giver but may provide great benefits for mankind. Similarly, gifts to educational and other cultural institutions may contribute much to the welfare of society, but relatively few of the benefits accrue directly to the benefactors.

Opponents of the deduction for charity argue that:

1. Many of the expenditures are not really for public goods.

2. The public—that is, the government—should determine directly how expenditures on public goods should be allocated.

3. The provision for deductibility of charitable gifts benefits mainly the rich and thus reduces the redistributive impact of our tax system.

4. Eliminating the provision of the deductibility of charitable gifts would have little effect on charitable giving.

It is difficult to assess the validity of these various claims. Many of the most important advances in medicine have been the outcome of research supported by private foundations. The Rockefeller Foundation's development of "miracle seeds" brought on a "green revolution" in developing countries that has greatly increased the availability of food in these countries. There is also some evidence that elimination of the deductibility of charitable donations would have a substantial effect on gift giving.[17]

Whether the deductibility provisions result in the tax system's being unfair is also not clear. To the extent that wealthy individuals create foundations that pay high salaries to their officers but spend little on true

---

[16] The fact that either one or the other is deductible implies that states should either levy sales taxes or income taxes, but not both. In fact, many states do levy both kinds of taxes.

[17] See, for instance, C. T. Clotfelter and R. L. Schmalbeck, "The Impact of Fundamental Tax Reform on Non-profit Organizations," in *Economic Effects of Fundamental Tax Reform,* ed. H. J. Aaron and W. G. Gale (Washington, DC: Brookings Institution Press, 1996), pp. 211–243; G. E. Auten, J. M. Cilke, and W. C. Randolph, "The Effects of Tax Reform on Charitable Contributions," *National Tax Journal* 45, no. 3 (September 1992): 267–290; and Y. S. Choe and J. Jeong, "Charitable Contributions by Low- and Middle-Income Taxpayers: Further Evidence with a New Method," *National Tax Journal* 46, no. 1 (March 1993): 33–39.

public goods, the provision for charitable donations may well be thought to be inequitable. So, too, for expenditures in goods that are largely valued by the rich. To the extent that the expenditures are really for public goods, though, the giver gets no more enjoyment out of the expenditure than do many other members of society. There is no more reason to include these expenditures in his or her income than in that of any other individual (who benefits equally by it).

If the appropriate basis of taxation is "income available for spending on private goods" (discretionary income), the appropriate tax treatment is to allow a full deduction (just as we argued earlier that the appropriate tax treatment of medical expenses was a deduction, not a credit). However, the consequence of this is that the marginal cost of charity is less for those in the 28 percent or higher brackets than for those in the 15 percent bracket.

The charitable contribution has also sometimes been abused, with individuals giving away property (such as paintings), and assigning a value to the gift far in excess of the true value. Recent legislation has imposed severe penalties for those who get caught doing this.

The deductibility of charitable contributions raises basic questions concerning the manner in which decisions regarding the supply of public goods should be made. Critics claim that the tax deductibility of charitable contributions gives, in effect, undue power to the rich in deciding which public goods should be provided. However, some argue that political processes do not provide a very good mechanism for registering ordinary individuals' attitudes toward different public goods. Individuals vote for representatives and have little opportunity for expressing their views on the relative allocation, say, of educational and health expenditures.

On the other hand, the provision for charitable deductions has encouraged a system in which public goods are provided by a variety of institutions. Individuals can express their views about the importance of different categories of public goods in a variety of ways. If the government were the only source of funds for, say, health research, the views of that bureaucracy would exclusively determine the direction of health research; as it is, these decisions can be made independently in a variety of institutions. The arguments for the decentralization of decision making for public goods closely parallel those for the decentralization of decision making in other areas: having competing (or at least alternative) organizations providing similar services leads to greater efficiency, and it allows for diversification, so that the consequence of mistakes will be smaller.[18]

---

[18] The provision for the deductibility of expenditures on religious charities has another motivation: some argue that the constitutional prohibition against legislation interfering with the free exercise of religion prohibits taxation of religious organizations. Although this may provide a justification for exempting religious institutions directly from taxation, it does not seem to provide justification for deducting gifts to them for purposes of the income tax.

CASUALTY LOSSES   Individuals are allowed to deduct losses from theft, fire, accident, and other casualties that exceed 10 percent of their adjusted gross income. The motivation for these provisions is again clear. These losses reduce the individual's ability to pay; they represent "expenditures" that were not voluntary and from which the individual presumably got no enjoyment.

This provision, however, has some important consequences. In particular, it means that the government effectively provides a kind of insurance against these casualties. The magnitude of the insurance depends (as we noted in our discussion of the medical deduction) on the individual's marginal tax rate. Once individuals have suffered a loss, they are obviously much better off than if the government did not provide this kind of insurance. On the other hand, these provisions may seriously distort individuals' behavior. Insurance, in general, reduces individuals' incentives to avoid the losses in question. Thus, if the government pays 33 percent of the loss from theft, the individual may not make as much effort to avoid being robbed.

EDUCATION EXPENSES   Two key tax credits exist to help defray educational expenses: the American opportunity credit, for tuition, enrollment fee, and course material expenses incurred in the first four years of postsecondary education; and the lifetime learning credit, for tuition and fees paid during all years of postsecondary education, including non-degree and vocational programs. The American opportunity credit is up to $2500 per eligible student per year, and was originally designed to make it possible for students in most states to go to a community college at almost no cost. (With an increase in tuition, this is no longer the case.) The lifetime learning credit is up to $2000 per tax return per year,[19] and is designed to encourage both college enrollment and professional "retooling" for workers who need to acquire new job skills or upgrade existing skills. It reflects a concern about growing wage inequality between those with secondary school education and those with college education, as there has been a consistently high correlation between years of education and household income over the past fifty years. College enrollment rates among children of poorer families have been markedly below those of children of the country's higher-income families. Without access to higher education, there is a vicious cycle perpetuating inequality, which seems so much in contradiction to American ideals. Poor families lack the resources to send their children to college, so their children will have low

---

[19] Eligibility for both tax credits is determined by *modified* adjusted gross income (MAGI), which excludes the deduction for contributions to IRAs, a deduction that is allowed in the calculation of adjusted gross income. The American opportunity credit's MAGI limit is $180,000 if married and filing jointly, and $90,000 for other filers; and the lifetime learning credit's MAGI limit is $122,000 if married and filing jointly, and $61,000 for other filers.

incomes. The tuition tax credits represent an effort to enhance equality of educational opportunity.

The tuition tax credits have been criticized on several grounds. First, critics argue that the much higher earnings of college graduates already provide sufficient private incentives to attend college. Moreover, they claim that the major expansion of student loan programs enacted in 1993, combined with the increase in Pell grants (tuition grants for children from lower-income families), already largely eliminated the financial barrier to attending college. To the extent that any financial barrier remained, it was mainly among lower-income children; and for these, the tuition tax credit would be of little benefit—it would be far better to expand the Pell grant program, and if necessary, the student loan program. As student debt has increased to a level in excess of $1 trillion, and tuition has increased faster than inflation, it has become clear that there are significant barriers to higher education that the tuition tax credit did not address. Second, there has been concern that colleges and universities would respond by increasing tuition, and although that would benefit the university system and ease pressure on state treasuries, it would not lead to much of an increase in student enrollment.

DEDUCTIONS VERSUS CREDITS

Adjustments to income for tax purposes can take the form of deductions or credits. A 20 percent tax credit is the same as a deduction for a person in the 20 percent tax bracket. By and large, most economists argue that *if an adjustment is to be made*, it is preferable to use credits rather than deductions. Deductions give bigger benefits to those in high tax brackets, and are, in that sense, inequitable; furthermore, because deductions result in different individuals' facing different prices, they are distortionary. However, as we have seen in the case of medical expenses, in some circumstances there can be an argument based on equity for deductions. If one believes that taxes should be based on ability to pay, and if ability to pay is best measured by income *net of expenditures on medical expenses*, then a deduction for medical expenses seems appropriate. If one believes that taxes should be based on private consumption—what one takes out of the economy—then expenditures on public goods (charity) should be subtracted.

IMPLEMENTATION PROBLEMS

- Determining income
- Timing
- Deductions

TEMPORARY TAX CHANGES

In 2001 and 2003, President George W. Bush faced a challenge: he wanted large tax decreases, but he did not want it to seem that the tax decreases were beyond the country's ability to afford them. Congress does its budgeting with a ten-year "window"—that is, it asks, how much, say, a new tax law would cost over the next decade. There are costs, of course, that extend well beyond ten years. One way to get a big tax decrease with a seemingly small cost is to make the tax change temporary (e.g., for the next five years) in the hope that when five years passes, there will be great pressure to renew the tax decrease. How later administrations pay for that tax decrease is a problem that others will have to solve. Thus, the 2001 and 2003 tax reductions were made only temporary.

One provision was especially peculiar: the estate tax was abolished, but for only one year, 2011.

The descendants of the very rich who died in that year would get a windfall gain—and some worried that it might even have perverse incentives.

Temporary tax changes can have particularly large distortionary effects, with effects that are often not those that were originally intended. For instance, the temporary reduction in dividends provided incentives for firms to pay out dividends while the tax rate remained low, and that was, in fact, what happened. However, the original hope that lower taxes would lead to higher savings and investment did not happen. Indeed, some argued that the effect on investment may have gone in the other direction: as firms paid out more in dividends, they had less cash on hand when good investment projects came along.

SPECIAL TREATMENT OF CAPITAL INCOME

Among the most complicated provisions of the tax code are those that relate to savings and capital income. Some of the problems arise from the desire to encourage savings—in particular, savings for retirement. However, any preferential treatment in one category of income gives rise to problems as individuals seek to convert income into a form that receives preferential treatment. The tax authorities respond by imposing rules that hinder these conversions. Thus, many complexities of the tax code arise from attempts to reduce the scope for these tax avoidance activities. Some of the problems, however, are inherent in the attempt to tax the return on certain assets, most notably on housing.

HOUSING

The most important investment for the majority of individuals is a home. The return to this asset—the housing service it provides—is not taxed in the United States.

In spite of the fact that the returns to housing (the housing services or imputed rent) are not taxed, money that individuals borrow to purchase houses (the interest on their mortgages) is tax deductible; indeed, since 1986 this has been the only form of consumer interest that is tax deductible.

Imputed rent on owner-occupied housing is not taxed because of the difficulty of ascertaining what the appropriate "rent" should be. This is not an insurmountable problem, however: we "impute" the market value of houses for purposes of property taxes. Once property values are known, it is fairly easy to estimate what the appropriate rent should be. There is, perhaps, another reason: the tax laws reflect the values of our society, and the strong belief that it is good for individuals to own their own home (perhaps an extension of the Jeffersonian ideal that America should be a country of small landholders). Individuals who own their own homes may be more likely to feel like members of the community and participate as constructive citizens. Considerations such as these—as well as the lobbying of the new-home construction industry—have been dominant in retaining the favorable treatment of housing.

Today, a family filing a joint return can make a gain of $500,000 every two years without incurring any tax liability. One of the reasons for this provision was that the tax on the capital gains in owner-occupied housing raised relatively little revenue (as it was not effectively enforced), but required extensive bookkeeping on the part of households.[20]

The favorable treatment of owner-occupied housing obviously has both equity and efficiency effects. It leads to an overinvestment in housing relative to other assets. In addition, there has been a general presumption that it generates benefits to homeowners relative to renters. How significant these effects are, though—and, indeed, whether they even exist—has varied over time with tax rates and benefits. The reason is that the tax law also gives considerable benefits to renters, though indirectly: by providing accelerated depreciation, tax deduction of

---

[20] Households were taxed on their capital gain—the difference between what they received when they sold their house and what they paid for it—*plus any expenditures on home improvements*. Thus, in principle, they needed to keep detailed records over a long period of time. Investments in owner-occupied housing have long benefited from another special provision: capital gains can be "rolled over." That is, if an individual sells a house, making a capital gain, but reinvests the proceeds in a larger house, the capital gain is postponed. The individual can continue to use the rollover provision until he or she dies, then the special provisions relating to taxation of capital gains at death mean that the capital gain completely escapes taxation.

interest, and taxing capital gains only upon realization and at favorable rates, the tax laws provide landlords with incentives to invest in rental housing; as this market is relatively competitive, in the long run, these benefits are passed on to renters. Thus, during 1981–1986, it appears that rental housing actually received more benefits than did owner-occupied housing, although since then, the traditional presumption appears to have been restored.[21] The 1997 law, by effectively making capital gains in housing tax exempt, exacerbated the incentives for overinvestment in owner-occupied housing.

SAVINGS FOR RETIREMENT

A variety of special provisions relate to savings for retirement, including special treatment of pensions and IRAs (individual retirement accounts).

OLD-STYLE IRA ACCOUNTS   The simplest of these to analyze is the old-style IRA accounts. Although they were discontinued in 1986, there is constant discussion of reintroducing them. In computing their adjusted gross income, individuals were allowed to deduct contributions to these accounts of up to $2000 (or $2500 for an individual with a nonworking spouse) annually, and neither the contribution nor the interest income was taxed until it was withdrawn.

The implications for savings can be seen by looking at what happens if one puts $100 aside for retirement. If Abigail saves $100, her current tax is reduced by $100 \times t$, so her consumption falls by $(1 - t)100$. Upon retirement, her consumption will increase. Assume the interest return is r—say, 20 percent. When she cashes in her investment, she must pay a tax on the total amount: $t \times (1 + r)100$. Her consumption next period thus goes up by $(1 - t)(1 + r)100$. The ratio of the change in consumption next period to the change in consumption this period is $1 + r$. *Allowing individuals to deduct savings from their income (but then taxing the entire amount, principal plus interest, upon retirement) is equivalent to exempting interest from taxation.* For those who save little, the preferential tax treatment has an ambiguous effect: an income effect that reduces savings (increases consumption) and a substitution effect that increases savings. For those who save a lot, however, the preferential tax treatment unambiguously reduces savings, as there is only an income effect; for those with savings beyond $2000, there is no marginal incentive.

There is a further criticism: individuals who already have assets do not have to save at all to receive the tax benefit. All they need to do is

[21] See M. King and D. Fullerton, *The Taxation of Income from Capital* (Chicago: University of Chicago Press, 1984).

FIGURE 22.5

THE EFFECT OF IRAS
ON SAVINGS

For individuals who save
more than the limit, IRAs
simply have an income
effect and no substitution
effect; they actually
reduce savings.
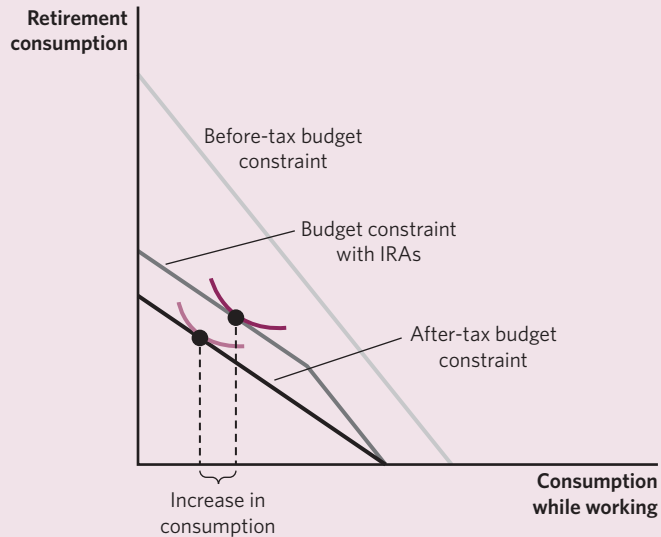
transfer savings from other accounts into an IRA account. For these individuals, there clearly is no marginal incentive effect, other than on the form of savings (see Figure 22.5).

NEW-STYLE IRAs   The 1986 tax law limited the deductibility of contributions to IRAs to lower-income individuals and to individuals not covered by employer-sponsored plans. Subsequent legislation has expanded IRAs and provided more flexibility of withdrawal. Today, there is no income limit to make nondeductible contributions of up to $5000 ($6000 if the individual is 50 or older) to an IRA and still defer the tax on the interest income. However, we saw earlier that the old-style IRAs were equivalent simply to allowing an exemption from interest, with one important difference: the timing of tax payments to the government. With the new-style IRAs, individuals pay more taxes now and less taxes later, although the present discounted value is the same. With the government facing major budgetary constraints, the "ruse" of moving tax receipts earlier was irresistible. Note that doing so has no effect on the long-run budgetary position of the government, as it will be receiving less taxes later.[22]

---

[22] The value to the government may differ if the discount rates used by the government and by the investor differ. Because, typically, the government can borrow at lower rates than the private sector, the cost of postponement is less than the value of postponement to the private sector.

In recent years, however, the government has focused on reducing the deficit in the short run, and thus the backloaded IRAs have gained in favor. There have even been cynical inducements to encourage individuals to convert old-style IRAs (in which taxes are not paid when money is put into the account, but when the money is withdrawn from the account) into new-style IRAs, with the "bribe" of greater flexibility in withdrawal. When the individual made the conversion, he or she would have to pay the accrued tax liability, so that current revenues would be enhanced.

The 1997 tax law greatly expanded the flexibility of the non-deductible IRAs through the Roth IRA, which became available in 1998. After a five-year holding period, distributions from a Roth IRA are tax free for those over age 59½ and first-time home buyers using the funds to purchase a house.[23] Even with an ordinary IRA, under current provisions there is no penalty for early withdrawal of up to $10,000 for a first-time home buyer, as well as for higher-education costs and unreimbursed medical expenses.[24]

In addition, self-employed individuals and small businesses can take advantage of SEP (Simplified Employee Pension) IRAs, SIMPLE (Savings Incentive Match Plan for Employees) IRAs, and Keogh accounts, all of which are actually tax-deferred pension plans.

PENSIONS   Savings that firms put aside for individuals for their retirement typically are not taxed until the individual receives the money. At the same time, the firm can deduct the amount it has set aside, just as it would if it had paid the money directly to the individual. Clearly there is a tax preference. There are two reasons for allowing the preference. First, it may be difficult for the individual (or the tax authorities) to know how to value the benefit under a *defined benefit scheme,* under which the employer specifies retirement benefits. Typically, these depend on how long the worker stays with the company, and what his or her pay is in the final years of working for the firm. (Under a *defined contribution* scheme, there is no problem: the firm simply sets aside a given amount of money into an account with the individual's name on it; it is as if the individual had received the money and put it into an IRA account.) The second reason may be more important: government wanted to encourage employers to put aside money for their employees' retirement, so there would be fewer destitute old people—fewer people to be a burden on the state in their old age.

INTEREST ON STATE AND MUNICIPAL BONDS

Interest on state and municipal bonds is tax exempt. These may include bonds used by municipalities to finance schools and by states to finance roads; industrial revenue bonds, which raise funds that are re-lent to businesses located in the town; and community-issued bonds that raise

---

[23] Individuals may contribute up to $5000 per year to a Roth IRA; eligibility phases out between $95,000 and $110,000 for an individual, and between $150,000 and $160,000 for a married couple.
[24] The 1997 law also raised the limits on deductible IRAs: for single individuals, the phaseout range is $30,000 to $40,000; for married persons filing jointly, it is $50,000 to $60,000.

funds to be re-lent for mortgages for lower- and middle-income individuals. Some states have set up special agencies to borrow funds to finance dormitories at private as well as state universities, to build sports complexes, and to construct hospitals.

The original motivation for the interest exemption was a concern about the constitutionality of taxing interest on state and local bonds. However, the expansion of this provision to include bonds issued by municipalities for money to re-lend to private individuals was clearly viewed as a form of federal subsidy to the localities. Many municipalities took up this federal subsidy with such enthusiasm that severe curbs have been imposed on the issuance of tax-exempt bonds.

Tax-exempt bonds are an inefficient way of aiding states and localities: the cost to the U.S. Treasury exceeds the benefits to the states and localities. The reason is that the interest rate on the bonds adjusts to make the marginal buyer of the bonds indifferent between buying taxable or tax-exempt bonds. Currently, there are sufficiently few individuals in the 39.6 percent tax bracket that the marginal buyer is at a much lower tax bracket. If the marginal buyer is in the 25 percent bracket, and if a taxable bond yields 10 percent, a tax-exempt bond (of the same risk) would yield 7.5 percent. It is this lower cost of capital that makes the tax-exempt status of value to states and localities. However, many of the bonds are purchased by individuals in the 39.6 percent tax bracket. If these individuals switch from a taxed to a tax-exempt bond, their tax payment goes down by 39.6 percent. If someone had $1 million to invest, he or she would have received $100,000 in interest and paid $39,600 in taxes, leaving $60,400. Now, the individual buys $1 million of tax-exempt bonds, receives $75,000 in interest, and is thus $14,600 better off. The community issuing the bond saves $25,000 in interest; it is better off than if the bonds had been taxable. The Treasury is worse off by $39,600. Thirty-seven percent of this amount goes not to benefit the community, but to benefit upper-income taxpayers.[25]

CAPITAL GAINS

In Chapter 21, we saw that returns to capital could accrue in the form of capital gains—increases in the value of an asset—and that these were typically treated differently from other returns. As we noted, the preferential

---

[25] One proposal for addressing this problem is to create "tax credit" bonds for which taxpayers would receive a credit at a particular rate for interest received from state and local bonds. These could be designed so that the interest paid by states and localities would be the same as with current tax-exempt bonds. Higher-income individuals, however, would not receive a windfall benefit, as they do currently.

treatment is in part a consequence of the difficulty of measuring the increase in value on an accrual basis—that is, as it occurs. As a result, capital gains are typically taxed only upon realization—when the asset is sold as opposed to "marking to market" (that is, taxing the increase in market value on a year-to-year basis). To be sure, there are assets, such as publicly traded stocks, for which it is easy to ascertain the market value; however, taxing those differently from assets such as real estate would itself introduce a distortion. In Chapter 21, we described a way by which this distortion could be largely eliminated, by imputing the gains as if they occurred smoothly over the holding period of the asset.

Although the taxation of capital gains upon realization is a natural outcome of the difficulty of measuring capital gains, the other forms of preferential treatment are largely a result of political pressures by wealthy individuals who receive much of their income in the form of capital gains. To be sure, they often argue that such preferential treatment is fair (otherwise, they pay taxes on the spurious increases in value attributed to inflation), or that it increases efficiency by spurring risk taking and entrepreneurship.

The two main arguments for preferential treatment—that it is unfair to tax nominal, instead of real, returns, and that it is important to encourage risk taking—are arguments for special treatment of all returns to capital, or returns to investments in risky ventures; they are not arguments for the preferential treatment of capital gains alone, regardless of the asset in which they are invested. Moreover, we saw in Chapter 21 that the equity argument itself was dubious, once the benefits of taxation upon realization are taken into account, and once it is noted that nominal (not real) interest payments on borrowing to finance the debt are deductible.

Whether the preferential treatment of capital gains leads to more risky investments remains a subject of controversy. The 1993 tax law attempted to target capital gains tax relief at new enterprises, which typically are quite risky; the 1997, 2001, and 2003 capital gains tax cuts, by contrast, were across the board, and applied as much to capital gains on unproductive investments, like gold, as it did on high-risk new-technology ventures.

A further objection to the preferential treatment of capital gains—and one of the reasons why the preferential treatment was eliminated in 1986—was that it gives rise to a host of tax avoidance schemes, as taxpayers work to convert ordinary income into capital gains. Even though the scope for such schemes was greatly reduced in 1986, with the reintroduction of large preferences, they emerged (see Chapter 21).

There is intense debate over lowering the capital gains tax rates. Critics point out that much of the benefit would accrue to investments that have already been made. Because the investments have been made,

SPECIAL TREATMENT OF CAPITAL INCOME

The U.S. tax system provides special treatment—lower rates or postponed taxes—to a large fraction of the total returns to capital. There are special provisions relating to:

- Housing
- Savings for retirement
- Pensions
- Capital gains

The preferential treatment distorts the allocation of resources and often introduces important inequities.

incentive does not matter. The lower taxes would be just a windfall gain, giving these investors a far higher return than they had expected when they made their investments. Indeed, the reductions in capital gains and dividend tax rates during the Bush administration may have led to lower savings and investments. Moreover, as wealth ownership is far more concentrated even than income, the benefits of the tax cuts would accrue disproportionately to the very rich. The reason why, in the end, the capital gains tax cuts have been so attractive is that the short-run budgetary cost is low. In the first few years, many individuals who have been "locked in" sell their assets, and have to pay capital gains taxes (though at reduced rates); this is especially true if they worry that capital gains taxes may rise later. However, although the short-run budgetary cost may be low (or, in some estimates, revenues may actually increase), in the long run there are adverse effects. The tax revenues today are partly at the expense of revenues that would have been realized in the future, when the asset would in any case have been sold.

CONCLUDING REMARKS

We ask much of our income tax: it should be simple, fair, and easy to collect, and should encourage economic efficiency. The tax is much maligned. It does not achieve any of its goals perfectly. It is the product of compromises: economic compromises between competing objectives, as well as political compromises. Were equity considerations of less concern, a far less complex and less distortionary tax system would be easy to design. However, compared with the tax systems in many other countries, there is less tax avoidance and greater compliance in the U.S. tax system, and perhaps even fewer distortions. Still, there remain demands for reform of the tax system. The directions that these reforms might take are discussed in Chapter 25. First, however, we will take a look at the other major tax, the corporate income tax, and will study some of the ways in which people avoid—legally—paying taxes under our current tax system.

REVIEW AND PRACTICE

SUMMARY

1. The U.S. income tax system is based on the principle that taxes should be related progressively to the family's cash (marketed) annual income. The tax code discriminates in favor of nonmarket transactions and against those with fluctuating income.

2. There are problems in implementing an income tax system, both in defining income and in determining the time at which the tax should be imposed. A principal difficulty encountered in defining income is distinguishing activities that are motivated by business considerations from ordinary consumption activities.

3. Many of the problems associated with designing a workable income tax system arise from the unobservability of (or costs of observing) the essential variables—for example, of knowing whether some medical procedure was really "necessary."

4. The tax code allows a number of adjustments to income and personal deductions, motivated both by equity and by incentive considerations. Regardless of the motivations behind them, however, the deductions have both incentive and equity effects that must be taken into account. For instance, the medical deduction effectively lowers the price of medical care, and it lowers it more for high-income individuals.

5. The basic unit of taxation in the United States (unlike most other countries) is the family. The tax system has a number of provisions that are intended to ensure that those in different family situations face equitable taxation. There is a limited tax credit for child care expenses, and different rate schedules for married couples and single individuals. Still, although the current system no longer imposes a marriage penalty on individuals with similar incomes, it still provides a marriage bonus for individuals with very dissimilar incomes.

6. The tax bills of 1986, 1993, 1997, 2001, and 2003 all affected the degree of progressivity of the federal tax code. The 1986 act reduced progressivity but eliminated many of the special provisions that were of special benefit to the rich. However, state and local income, sales, and property taxes remain deductible, as does mortgage interest. Employee benefits, including medical benefits, are not taxed. The 1993 act substantially increased progressivity, by greatly increasing tax rates on upper-income individuals (from 28 percent to almost 40 percent). Subsequent tax reforms have given substantial benefits to the very rich, by enacting large cuts in capital gains, but at the same time providing a child tax credit. The net effect of these changes is that upper-income salaried workers are worse off, upper-income individuals who live off capital are better off, and middle-income individuals with children are better off. In addition, the simplification of the tax code, the great achievement of the 1986 bill, has been reversed, as numerous special provisions have been introduced, for example for education and capital gains.

KEY CONCEPTS

Accrual basis

Adjusted gross income (AGI)

Barter arrangements

Cash basis

Child and dependent care tax credit

Child tax credit

Effective tax rate

Flat-rate tax schedule

Haig–Simons concept

Marginal tax rate

Marriage subsidy

Marriage tax

Minimum tax

Present discounted value

Standard deduction

Taxable income

Tax credits

QUESTIONS AND PROBLEMS

1. For each provision of the tax code listed below, which provides the best explanation: incentives, horizontal equity, vertical equity, administrative reasons, or special interest groups?

    a. Deductibility of mortgage interest

    b. Deductibility of casualty losses

    c. Deductibility of medical expenses

    d. Deductibility of charitable contributions

    e. Child care tax credit

    f. Credit on taxes paid to foreign governments

    g. Tuition tax credit/deduction

    Which of these provisions can be justified as a response to a market failure? Compare this response—the use of the tax system—with alternative responses (such as direct government expenditures).

2. Discuss the arguments for and against using a tax credit rather than a deduction for medical expenses, charitable contributions, and child care expenses.

3. If you were asked to write the regulations concerning the deductibility of business expenses, how would you treat the following items, and why? Discuss the inequities and inefficiencies associated with alternative possible rules.

    a. Educational expenses required to maintain one's current job

    b. Educational expenses incurred to obtain a better job

    c. Moving expenses arising from a reassignment by one's present employer

    d. Moving expenses incurred in obtaining a new job

    e. Business suit worn by an individual who does not wear suits except for business

    f. Business lunches costing more than $25

    g. Expensive cars

    h. Commuting costs

    i. Car expenses of a traveling sales representative

    j. Subsidized cafeteria lunches

4. Under the old tax law, money that scientists received from winning the Nobel Prize (or similar prizes) was not taxable. Now it is. Which treatment do you think is appropriate? Under both the old and the new tax laws, money received by lottery winners (as well as gambling receipts) is taxable, but losses are not deductible. Is this fair? What distortions does this introduce?

5. Consider a divorced individual who earns $26,000, has two children, and pays $4800 in child care expenses, $2000 in mortgage interest payments, and $4500 in medical expenses. Medical expenses in excess of 7.5 percent of one's income are deductible. When the individual's income is $26,000, he or she gets a 22 percent child care expense credit; when the income is $28,001, the credit is only 20 percent. The individual's base marginal tax rate is 15 percent. What is the actual marginal tax rate?

6. Explain why federal tuition tax credits might lead states to increase the tuition they charge. Many states currently charge a tuition in state community colleges that is substantially below $2500, the level of the federal credit. What might one expect to happen in those states?

7. In the debate over the tuition tax credit, one administration economist pointed out that in Georgia, which had provided these scholarships for its students, tuition had not increased after the credit had been provided. Why might a federal tuition tax credit or deduction be expected to have a different impact than a credit or deduction provided by the state for state universities?

23

THE CORPORATION INCOME TAX

The corporation income tax has been the subject of considerable controversy, with some (such as Barack Obama, during the 2008 presidential campaign) arguing that corporations were escaping their fair share of taxes, and others (such as John McCain, during the same presidential campaign) calling for a cut in corporate taxes. There is even controversy about who actually pays the corporation income tax—its incidence; as we shall see, many economists believe that the tax is actually borne largely by consumers and workers, not by the owners of the corporation. Indeed, one of the reasons why the tax has remained so controversial is this widespread debate about who bears the tax. Economists agree on one thing (which is not well understood by others): *The corporation does not bear the tax*. People—shareholders, customers, workers, managers—bear taxes. The question is, which people?

Following a decline in importance as a source of federal revenue for almost three decades, average effective tax rates and the share of total federal receipts increased briefly after 1983; although share of revenue

has stabilized at between 10 and 12 percent in most years since then, the effective federal corporate tax rate has continued its steep decline from 34 percent in 1986 to 18 percent in 2010 (see Figure 23.1).

But these numbers do not fully convey what has happened. The effective tax rate is the ratio of taxes to reported profits. Corporations have been very effective in redeeming reported profits. Changes in aggregate revenues may not, however, convey much information about changes in the economic impact of the corporate tax. The corporate tax causes important distortions in the allocation and overall level of investment. The nature and magnitude of the distortion relate, in part, to specific provisions of the tax, such as the differences in tax treatment of interest payments and dividends, and the allowances for depreciation.

This chapter first describes the U.S. corporate income tax, then analyzes alternative interpretations of the tax—different theories ascribing the burden of the tax to different groups. The analysis includes not only how the tax affects economic efficiency, but also how it affects financial decisions. The chapter concludes with a discussion of the most fundamental issue: Should there be a tax on corporations?

FIGURE 23.1

FEDERAL CORPORATE INCOME TAX: EFFECTIVE RATE AND SHARE OF TOTAL FEDERAL RECEIPTS

From 1955 through 1982, the average effective tax rate of the corporate income tax and its share of federal receipts fell steadily. Even though the share has stabilized since then, the rate has continued its precipitous decline and both numbers are still considerably below levels in the period 1955 to 1970.

SOURCE: U.S. Department of Commerce, Bureau of Economic Analysis, *National Income and Product Accounts, 1929–2011*, Tables 3.2, 6.17B, 6.17C, 6.17D.
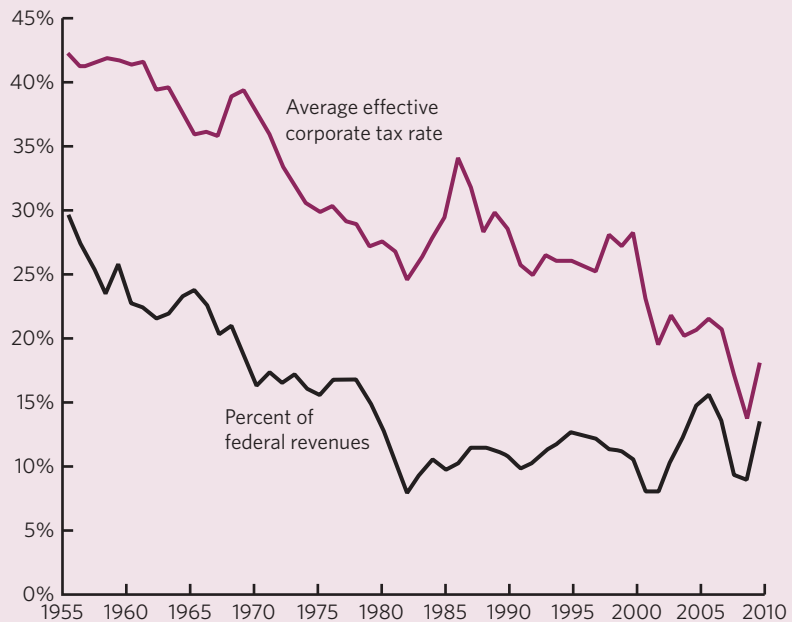
THE BASIC FEATURES OF THE CORPORATION INCOME TAX

The corporation income tax applies only to *incorporated* businesses. The essential difference between an incorporated business and an unincorporated business is the liability of investors for the debts of the corporation. Corporations have limited liability; that is, investors can lose only the amount of money they have invested in the firm. In contrast, if an unincorporated business has debts it cannot pay, the creditors can attempt to recover their losses from the owners. Because of the protection provided to investors by the corporate form of organization, almost all large firms in the United States are incorporated. More recently, limited liability partnerships have enabled some organizations to benefit from limited liability, acting essentially as corporations, except that they avoid paying corporate income taxes. In the case of partnerships, profits are attributed back to the owners in proportion to their "share" in the partnership, whether the profits are distributed or not.

The corporate tax is essentially a flat-rate tax, currently at 35 percent for most businesses. However, as shown in Table 23.1, the average tax rates for corporations with taxable income below $18.3 million are somewhat lower, as a concession to small business. The **tax base** is corporate taxable income. In general, taxable income is defined in the tax law to be gross revenues less wages, materials, interest paid, and depreciation allowances. Dividend payments, however, are not deductible. Thus, when Congress legislates more generous depreciation allowances, as it did in 1981, a corporation's taxable income is reduced, and so is its tax.

TABLE 23.1 MARGINAL AND AVERAGE TAX RATES ON CORPORATE TAXABLE INCOME

TAXABLE INCOME	MARGINAL RATE	AVERAGE RATE
$0–50,000	15%	15%
$50,000–75,000	25%	15%–18.3%
$75,000–100,000	34%	18.3%–22.25%
$100,000–335,000	39%	22.25%–34%
$335,000–10,000,000	34%	34%
$10,000,000–15,000,000	35%	34%–34.3%
$15,000,000–18,333,333	38%	34.3%–35%
Over $18,333,333	35%	35%

As with the individual income tax, after the tax liability is calculated, a firm's tax is reduced by *tax credits*. Before its repeal in 1986, the most important corporate tax credit was for new investment. Firms still obtain a credit for a portion of their research and development expenditures and for most taxes paid to foreign governments.

Table 23.1 presents the statutory tax rates for corporations as listed in the official IRS tax tables. However, corporations actually pay much less. Although nominal corporate income tax rates in the United States are higher than in many other countries, the effective tax rate—the ratio of tax payments to corporate profits—is lower because firms have mastered the art of tax avoidance. The U.S. statutory marginal rate of 39.2 percent is much higher than the average OECD rate (excluding the United States) of 25.5 percent, but the gap in effective tax rates is much smaller, with the United States at 27.1 percent and the OECD at 23.3 percent. The same holds true when comparing the United States with the world's 15 largest countries, whose statutory rate is about 10 percent lower than the U.S. rate, at 29.8 percent, but whose effective rate is only 2 percent less than the U.S. rate, at 25.3 percent.[1] In fact, as a share of GDP, corporate tax revenues in the United States are less than half that of the average of the advanced industrial countries due to tax loopholes and the shifting of profits abroad.

This is partly because U.S. corporations, especially multinationals, have excelled in the science of tax avoidance. For example, General Electric (GE) is often given accolades for its good management and innovation. But one area of innovation that has been called into question is its skills at tax avoidance—in many years, it has succeeded in paying zero taxes. Of the 288 Fortune 500 companies that were profitable every year from 2008 through 2012, GE and 25 others paid no federal income tax over the five-year period, and 111 did not pay any federal income tax in at least one of these five years.[2]

This provides the basis of an important set of reforms: if one eliminated the loopholes and special provisions, one could simultaneously raise more money and lower tax rates.

In the discussion that follows, we shall see that the impact of the corporation income tax depends on several features of the tax code—the fact that interest payments are tax deductible but dividend payments are not; the fact that depreciation allowances are typically more generous than true economic depreciation unless they are repatriated; and the fact that the government taxes profits but does not provide a

---

[1] J. G. Gravelle, *International Corporate Tax Rate Comparisons and Policy Implications*, Congressional Research Service Report R41743, December 28, 2012.

[2] Robert S. McIntyre, Matthew Garner, and Richard Phillips, *The Sorry State of Corporate Taxes: What Fortune 500 Firms Pay (or Don't Pay) in the USA and What They Pay Abroad—2008 to 2012* (Washington, DC: Citizens for Tax Justice and Institute on Taxation and Economic Policy, February 2014).

symmetric treatment for losses[3]—and on the relative rates at which individual incomes and corporations are taxed. We shall also see that the impacts of the tax extend beyond whether investments in corporations are favored over investments in the noncorporate sector; they extend to how production in the economy is organized, the kinds of investments that are made, and how investments are financed.

THE INCIDENCE OF THE CORPORATION INCOME TAX AND ITS EFFECT ON EFFICIENCY

There are several different views on the effects of the corporation income tax. The popularity of the tax is attributable in part to the fact that its incidence is unclear. Politicians like to give the impression to voters that someone else pays the corporate tax. The reality, however, is that households, workers, consumers, managers, and investors pay the tax, just as they pay any other tax. The question is how the burden is shared among these groups. Under some plausible conditions, for instance, the corporate income tax is effectively borne by workers, consumers, managers, and not by investors at all; in other cases, it is borne not just by investors in the corporate sector, but also by all investors.

THE CORPORATION INCOME TAX AS A TAX ON INCOME FROM CAPITAL IN THE CORPORATE SECTOR

One of the earliest views of the corporation income tax, and still one of the most prevalent, is that it is a tax on the return to capital in the corporate sector. With wage payments and purchases of other inputs tax deductible, what remains is profits and the return on capital. In a perfectly competitive economy, there would presumably be no pure profits, so the tax is *just* a tax on the return to capital.

---

[3] That is, the government is like a silent partner that shares profits but not losses. The government does allow a firm to carry forward losses for a limited number of years. However, because the benefits are not received until some time in the future, they are worthless. A loss of $100 today used to offset a profit next year is worth only approximately $85 if the firm pays a 15 percent interest rate, $70 if the firm has to wait two years before taking advantage of it.

Assume, for instance, that the firm did not finance any of its investment by borrowing (or that interest was not tax deductible). The corporation income tax reduces the after-tax return to capital in the corporate sector. *This is the short-run effect.* Typically, though, as we saw in Chapter 18, matters do not stop here. In an ingenious model, Arnold Harberger, then of the University of Chicago and now of UCLA, traced out the longer-run effects, assuming that the overall supply of capital was fixed but that capital was perfectly mobile between the corporate and the unincorporated sectors. With a lower return to capital in the corporate sector, capital shifts out of that sector and into the unincorporated sector. Firms in the corporate sector substitute labor for capital, and thus labor shifts from the unincorporated sector into the corporate sector. Goods in the corporate sector become more expensive (thus, the tax is like an increased cost of production). When matters are fully equilibrated, the after-tax return on capital is the same in the corporate and the unincorporated sectors: thus, to the extent that capital bears a cost, it is borne equally by those in all sectors of the economy. Consumers bear some of the cost in the form of higher prices. Workers, too, may bear some of the cost: if, for instance, the corporate sector is far more labor intensive than the unincorporated sector, the reduced demand for the output of the corporate sector reduces the overall demand for labor; this effect is reinforced if capital is easily substituted for labor in the unincorporated sector, for then firms in that sector may substitute capital for labor as the cost of capital falls. Although the precise manner in which the burden is shared by consumers, owners of capital, and workers depends in complicated ways on demand elasticities, on elasticities of substitution (which measure how easy it is to substitute capital for labor), and on labor intensities in the two sectors, the fundamental point is that the long-run effect can be markedly different from the short-run effect.

John Shoven of Stanford University has solved explicitly for the effect of the corporation tax in the intermediate run, comparing the present equilibrium with what it would have been in the absence of the distortionary corporate tax on the return to capital. He estimated that the deadweight loss from the corporation income tax was roughly 12 percent of the tax revenue it generated.

Shoven also estimated the extent to which the burden of the corporate tax was shifted from capital owners to consumers and workers. The share of the burden borne by owners of capital is measured by the change in the income of capital divided by total corporate tax revenues. If the burden exceeds 100 percent, then the income of capital is reduced by *more than* the tax. The corporate sector is relatively capital intensive (that is, much more capital is used for each worker than in the unincorporated sector). Hence a shift in demand toward the unincorporated sector

indirectly reduces the demand for capital and, thus, the returns to capital. The smaller the elasticity of substitution in the unincorporated sector, the greater the extent to which the return to capital must be lowered to absorb all the capital that is released as a result of the change in the composition of demand. This explains why Shoven calculated that capital bears 162 percent of the burden of the tax in the case in which the elasticity of substitution in the unincorporated sector is very low relative to that in the corporate sector and when the consumer demand elasticities are low.

Without knowing the precise value of the demand elasticities and elasticities of substitution, we cannot even tell whether capital bears more or less than 100 percent of the burden. No wonder, then, that there is no agreement about the extent to which the corporate tax is, effectively, a tax on capital or a tax on consumers. However, details of the tax code matter—depreciation allowances and deductibles of interest mean that the consequences of the corporate incentives may be markedly different from that suggested by the calculations.

SHIFTING OF THE CORPORATE TAX IN THE LONG RUN

The Harberger model has been important in focusing attention on the *general equilibrium* effects of taxes—in emphasizing that a tax imposed on capital in one sector has ramifications throughout the economy; that with perfect capital mobility, the tax must be borne equally by capital in both sectors of the economy; and that the general equilibrium incidence may be quite different from the direct impact, in some cases with capital bearing more than 100 percent of the tax.

However, the Harberger model has been extensively criticized on two grounds. First, its assumption that the capital stock is fixed but that capital can be fully shifted between sectors seems peculiar. By contrast, if one assumes that the capital stock itself is variable, quite different (and sometimes simpler) results emerge. For instance, for a small country facing a perfectly elastic supply of capital at the international interest rate $r*$,[4] the after-tax return remains unaffected by the tax. Thus, (a) the tax increases the cost of production in the corporate sector; (b) production thus shifts to the unincorporated sector; and (c) as this happens, whether the demand for labor falls or increases depends in part on whether the unincorporated sector is less or more labor intensive than the corporate sector. Accordingly, owners of capital bear none of the tax, consumers of

---

[4] Similar results obtain if the supply of savings within a country is highly elastic.

USER COST OF CAPITAL

- The tax system affects how much it costs a firm to use capital.
- With accelerated depreciation and full deductibility of interest payments, investment may actually be encouraged relative to a no-tax situation.
- Investment is discouraged in assets which are financed by equity and those in which the benefits of accelerated depreciation are limited (such as short-lived assets).

the goods produced by the corporate sector always bear some of the tax, and workers may or may not be adversely affected.

The second objection is more difficult to deal with. It is that the tax itself is not just a tax on the return to capital. Interest is tax deductible, and there are typically accelerated depreciation allowances. Obviously, more favorable tax provisions, such as accelerated depreciation, lower the cost of using capital, and thus encourage its usage. In the paragraphs that follow, we derive some detailed formulae relating the after-tax cost of using capital, called, not surprisingly, *the user cost of capital*, showing how a variety of tax provisions affect the costs of using capital. Not only is the effective tax rate often markedly lower than the official tax rate, but under some plausible assumptions capital is also actually subsidized *at the margin*.

USER COST OF CAPITAL   The magnitude of the effective tax on capital in the corporate sector depends on several provisions of the tax code—on investment tax credits, depreciation allowances, and the fraction of investment that is financed by debt (which is tax deductible).

Suppose the cost of a machine is p. If a firm rented the machine for a year, in a competitive capital market, the owner would charge rp in interest, where r is the interest rate, and δp in depreciation, where δ is the percentage reduction in the value of the machine as it has grown older. (Recall our discussion of depreciation from Chapter 21: as machines get older, they become less valuable both because they wear out and because they become obsolete; δ summarizes both effects.)[5] If R is the yield of the machine, the firm will rent the machine so long as

$$R > (r + \delta)p.$$

The effect of a tax imposed at the rate t on the corporation depends on what is deductible. If the firm were actually renting the machine, the full rental price would be deductible. If rental rates and depreciation charges, $(r + \delta)p$, remained unchanged, then both sides of the inequality would be multiplied by $(1 - t)$. The tax would induce no distortions, because the firm would still rent the machine if $R > (r + \delta)p$, just as it does in the absence of the tax.

Now, suppose that the firm buys its own machine, as is usually the case. The interest on the money the firm borrows to buy the machine is

---

[5] Changes in prices may at times result in capital gains on machines, in which case this, too, would be reflected in the rental price.

deductible. Assume a fraction α is financed by borrowing (for buildings, firms often borrow 80 percent or more of the costs; for machines, they often can borrow only a third), then the interest deduction is $\alpha r p$.

The true economic depreciation is δp. In Chapter 21, however, we pointed out that, typically, firms are allowed more generous depreciation allowances. The 1981 tax law provided for greatly accelerated depreciation, and although these provisions were scaled back in 1986, depreciation still is at a greater rate than true economic depreciation. Let β denote the extent of acceleration—that is, the tax deduction for depreciation is $\beta\delta p$, with $\beta > 1$.

Finally, the government has, from time to time, provided an investment tax credit. If a machine costs $100 million, a 5 percent investment tax credit means that the firm's taxes are reduced by $5 million. It is as if the government has paid 5 percent of the cost, or the price has been reduced by 5 percent. Let c denote the tax credit rate, which effectively just reduces the price. Thus, the annual after-tax cost of capital is now

$$(r + \delta)p(1 - c) - t(\alpha r + \beta\delta)p(1 - c) =$$

$$(r + \delta)p(1 - t) - p\{c[r(1 - \alpha t) + \delta(c - \beta t)] - t[(1 - \alpha)r + (1 - \beta)\delta]\}.$$

The first term is the annual cost, after accounting for the investment tax credit. The second term is the value of the income tax deductions.

Recall that, after taxes, the machine yields $R(1 - t)$. If the cost is multiplied by the same factor, then the tax will have no effect on the firm's decision whether or not to buy the machine. However, the preceding expression may be more or less than gross costs times $(1 - t)$, depending on whether the term in curly brackets is negative or positive. If the investment is totally debt financed ($\alpha = 1$), there is no accelerated depreciation ($\beta = 1$), and there is no investment tax credit ($c = 0$), the cost of capital is reduced by exactly the same amount as the return is reduced; there is no distortion. Today there is no investment tax credit. What matters for firm decisions is the marginal user cost, since many if not most firms finance marginal investments by debt ($\alpha = 1$). Hence, the tax law encourages investment at the margin. The benefits of accelerated depreciation are typically greater for long-lived assets. The corporation income tax can lead to overinvestment in long-lived assets at the same time that it leads to underinvestment in short-lived assets.

THE CORPORATION TAX FOR A FIRM WITHOUT BORROWING CONSTRAINTS

The preceding analysis treated the firm as if it had a fixed debt-to-equity ratio—that is, it could finance only a certain fraction of its investment, α, with debt. In fact, one of the important decisions firms face is how to

finance new investments they wish to undertake; for most firms, borrowing is an option. What, then, is the effect of the deductibility of interest payments on such decisions? We answer this question by assuming that depreciation allowances correctly reflect the decrease in the value of aging plant and equipment, thereby isolating the implications of the deductibility of interest payments.

If there is no accelerated depreciation and no investment tax credit, if a firm finances all of its investment with debt, there is no distortion. Returns are reduced by $(1 - t)$, but the cost of capital is reduced proportionately. In economic decision making, what matters is how returns and costs are affected *at the margin*. At the margin, firms in well-functioning capital markets can choose to finance additional investment out of borrowing. If they do, then their marginal costs are reduced by $(1 - t)$, by the same proportion that their returns are. Investment decisions are thus unaffected by the corporate tax. Consider, for example, the simplest case of an infinitely lived asset (there is no depreciation). If the firm borrows to buy such an asset costing p dollars, its before-tax cost is rp, but its after-tax cost is $r(1 - t)p$. When the firm asks itself whether is it worth, at the margin, borrowing a little bit more to invest a little bit more, its answer is the same after the tax is imposed as before.[6] Thus, in the standard theory of the firm (sometimes called the *neoclassical theory*), when the firm faces no borrowing constraints, at the margin, it can be thought of as borrowing to finance investment, and with the marginal investment debt financed, and with debt tax deductible, the corporation income tax causes no distortion in investment.

In practice, much, if not most, of investment is financed by debt at the margin. Typically, firms first use retained earnings (earnings less what they distribute as dividends) to finance investment, and investment beyond that amount is financed by borrowing.

INCIDENCE OF THE CORPORATION INCOME TAX WITH CREDIT-CONSTRAINED FIRMS

The model presented earlier in this chapter assumed, as we have noted, that at the margin, investments were financed the same way that they were on average. We have just seen that if firms are not credit constrained, at the margin, investment can be thought of as financed by borrowing, in which case there is no distortion.

---

[6] Equivalently, a firm with outstanding debt that is not currently borrowing from the market but is doing some investment can ask itself: Is it worth investing a little bit less, and using the extra funds to repay some of our outstanding debt obligations?

Many firms, especially smaller ones, however, are credit constrained. In that case, the impact of the corporation tax may be markedly different. If a firm can neither borrow nor issue equity, and if it is investing all its retained earnings, taxes reduce the funds available for investing, and thus reduce investment. Note that, in this case, the impact of the corporate income tax depends on the total amount by which the funds available for investment are reduced—that is, it depends completely on *average* tax rates, not marginal tax rates, as in the earlier analysis.

There is a typical life cycle to the firm. Newly established firms sell shares to raise capital; it is often the only source of capital they can obtain—banks find long-term loans too risky, and these firms are too small to issue long-term bonds. Firms are reluctant to make long-term investments based on short-term loans. The original entrepreneur usually takes his or her return largely in the form of stock ownership (rather than wage payments). Thus, the corporation income tax, which exempts interest payments, can be viewed effectively as a tax on these new, credit-constrained, entrepreneurial firms; it has adverse effects on the investment of new firms that cannot raise funds by borrowing additional amounts.[7]

If this view is correct, the long-run effects of the tax are not so much those associated with the reallocation of resources between the corporate and noncorporate sectors. Rather, they have to do with the degree of innovativeness of the corporate sector and the rate of technical progress. The magnitude of these effects depends on the elasticity of supply of entrepreneurship and risk taking. It was precisely these concerns that led President Clinton, in his 1993 tax bill, to include provisions for preferential treatment of very long term capital gains in new firms, which still remain in effect today.

This analysis[8] makes clear that the impact of the corporation income tax depends critically both on the special features of the tax (such as the tax deductibility of interest) and the situation of the corporation

> ## CORPORATION INCOME TAX WITH AND WITHOUT CREDIT CONSTRAINTS
>
> - If firms can borrow to finance investment at the margin, then a corporation income tax with true economic depreciation does not distort investment.
> - In effect, the corporation tax can thus be viewed as a tax on credit-constrained firms, such as new firms that cannot borrow easily.
> - In these cases, the distortion may be more closely related to the average tax rate than to the marginal tax rate.

---

[7] In addition, we noted in Chapter 21 that the imperfect risk sharing (limitations on loss offset provisions) discourages risk taking, which is particularly important for these new enterprises.

[8] These alternative views of the corporation income tax were put forward in J. E. Stiglitz, "Taxation; Corporate Financial Policy, and the Cost of Capital," *Journal of Public Economics* 2 (February 1973): 1–34; and "The Corporation Income Tax," *Journal of Public Economics* 5 (April–May 1976): 303–311.

(whether it is or is not credit constrained).[9] Many empirical studies of the effects of the corporation income tax simply assume that it increases the marginal cost of capital to the firm—the amount it would cost the firm to invest an additional dollar—by an amount equal to the average tax payments per unit of capital. That is, they assume that the marginal and average costs of capital are the same. We have seen that if firms can finance their marginal investment by borrowing (and if depreciation allowances are equal to true economic depreciation), there may be no marginal distortion caused by the tax system, and the marginal cost of capital may differ markedly from the average cost. Regardless of whether one holds the view that the marginal cost of capital is equal to (or less than) the marginal cost of funds raised by borrowing, there is no justification for the hypothesis that marginal and average costs of capital are the same.

THE CORPORATION TAX AS A TAX ON MONOPOLY PROFITS

When there are monopolies, there are monopoly profits. Pure profits, sometimes called *excess profits,* are total revenues less total costs, including a normal return on capital. (In long-run equilibrium, with competition, and with constant returns to scale, there are no pure profits.) The part of the corporate income tax corresponding to a tax on pure monopoly profits is, in effect, a lump-sum tax on monopolists. The reason for this is simple. A monopolist maximizes its profits, π. If there is a tax at the rate t on its profits, its after-tax profits are $(1 - t)\pi$. But whatever the monopolist does to maximize π is exactly what it does to maximize $(1 - t)\pi$.

Whether the distortionary effects of the corporation income tax are greater (per dollar raised) with monopoly or competition is uncertain. On one hand, to the extent that the tax is partly a pure profits tax, it is nondistortionary. On the other hand, to the extent that it acts as an excise tax—a tax on what is produced in the corporate sector—and, because of

---

[9] From this perspective, it makes little difference whether the firm actually cannot borrow or has to pay what it views as an exorbitant marginal interest rate. There is a large recent literature showing the prevalence of credit rationing, as well as providing theoretical rationale for its existence. See J. E. Stiglitz and A. Weiss, "Credit Rationing in Markets with Imperfect Information," *American Economic Review* 71, no. 3 (June 1981): 393–410; also see T. Hellman and J. Stiglitz, "A Unifying Theory of Credit and Equity Rationing in Markets with Adverse Selection," *European Economic Review* 44, no. 2 (February 2000): 281–304; G. Hubbard, ed., *Asymmetric Information, Corporate Finance, and Investment,* A National Bureau of Economic Research Project Report (Chicago and London: University of Chicago Press, 1990); and J. Edwards et al., eds., *Recent Developments in Corporate Finance* (New York and Melbourne: Cambridge University Press, 1986).

monopoly, production of the sector is already lower than the socially optimal level, the tax causes a greater distortion.[10]

If a good is produced by a monopolist, the price may rise by more than the tax payments per unit output. As we saw in Chapter 18, under monopoly, with a demand curve of constant elasticity (that is, a 1 percent change in price has the same percentage effect on demand, regardless of the level of output), price is just a fixed markup over the marginal costs of production—including taxes. If the markup is, say, 20 percent, then if marginal costs increase by \$1 from the imposition of the corporation tax, price will increase by \$1.20.

Thus, if the corporate sector consists of a large number of industries, each controlled by a monopolist, the tax is largely shifted onto consumers, with the observed increase in prices being greater than the tax payments to government.

MANAGERIAL FIRMS: AN ALTERNATIVE PERSPECTIVE

To predict how firms will be affected by—and respond to—the corporate income tax, we need a model of firm behavior. That is, we need to make assumptions about how firms act and what they try to achieve. The analysis so far has assumed that firms maximize their after-tax returns. There are a number of aspects of firms' behavior that seem hard to reconcile with this view. For instance, later we shall show that firms should not pay dividends; there are better (from a tax perspective) ways of distributing funds from the corporate to the household sector. The fact that firms have continued to distribute as much of their earnings as they have is called the "dividend paradox." Firms' dividend policy is not the only inexplicable aspect of corporate behavior.

Accelerated depreciation provides another "tax paradox." At times, the government has given firms and individuals the right to depreciate their assets at an accelerated rate. Again, the total nominal depreciation allowances are unaffected (they equal the cost of the machine, minus its salvage value, if any). However, more rapid depreciation reduces reported income and, hence, taxes, in the early years of the asset. All firms should want to take advantage of this opportunity, yet firms were very slow to do so.

---

[10] We noted in Chapter 19 that the deadweight loss of a tax increases with the square of the tax. The effect of monopoly is similar to that of a tax. Indeed, if the demand curve has a constant elasticity of, say, 2, monopoly has the same effect on output and consumer prices as a 50 percent tax on the output of a competitive industry. Imposing a 10 percent tax on the output of the monopoly thus has the incremental deadweight loss associated with increasing the tax on a competitive industry from 50 to 70 percent (taking into account the fact that the monopoly price rises by twice the magnitude of the tax when there is a constant elasticity demand curve with elasticity of 2). This is substantially larger than the incremental deadweight loss from increasing the tax from 0 to 10 percent on the output of competitive industry.

This is true even though firms can make it clear to their shareholders that *reported* earnings are lower than they otherwise would be because the firm has made use of accelerated depreciation; indeed, one might have thought that shareholders would take such a report as a positive signal of good management, and the absence of such a report as a negative signal.

Firms have a choice of how to treat their inventories. Assume a firm that is selling steel beams bought some steel at $40 a ton and some at $100 a ton, a few months later, as a result of rapid inflation in the industry. Both kinds of steel beams are in its inventory. When it sells some steel beams for, say, $110 a ton, does it say its cost of purchase was $40 or $100? The Internal Revenue Service allows the firm to choose what to say, as long as it does so in a consistent manner. It can say either that it is always selling the item most recently acquired (this is called the *last-in, first-out system*, or LIFO) or that it is selling the first item acquired (*first-in, first-out*, or FIFO). In inflationary periods, LIFO has a decided advantage over FIFO. Current tax liabilities are lower (although future tax liabilities are increased by the same amount). However, the general principle that a dollar today is worth more than a dollar tomorrow implies that firms are better off with lower current tax liabilities. Amazingly, though, firms were very slow to switch to LIFO, and even today, many firms continue to use FIFO.

EXPLANATIONS FOR TAX PARADOXES Two explanations are offered for such seeming irrationalities. One is that managers of the firms have considerable discretion in managing their firms, sometimes pursuing their own interests at the expense of shareholders' interests. In particular, they may not pursue value-maximizing strategies. They may, for instance, wish to maximize the rate of growth of the firm; they may do this either because they believe that the larger their firm and the faster its rate of growth, the larger their salary, or simply because they enjoy the excitement that accompanies expansion and the personal recognition that it affords. In this view, the discipline of the marketplace simply is not strong enough to ensure that managers act in a profit- or value-maximizing manner. Firms that pursue the interests of managers, rather than maximizing profits, are called **managerial firms**. The second explanation is that firms (or their managers) are rational but that shareholders are irrational. Shareholders do not understand how the tax system (or corporations) work. It is unlikely that they would notice a firm's switch to the LIFO system, but they would see the firm's current reported profits decline, and they would believe that the firm was not doing as well as it was. As a result, the price of the firm's shares would decline. Managers, whose compensation often depends partly on the market value of the firm, thus prefer to keep shareholders "happy" by engaging in policies that do not minimize the firm's tax liabilities. Both explanations probably are partially correct.

These theories of firm/managerial behavior are important for several reasons. For instance, they yield quite different predictions concerning the effects of changes in corporate taxes on firms' behavior and government revenue. Consider a simple change, such as allowing more-rapid accelerated depreciation. Standard theories might predict a strong investment response, with a concomitant reduction in revenues to the Treasury. However, if managers are worried that the accelerated depreciation will make this year's profit statements look bad, many firms may fail to avail themselves of the tax advantages—with less loss of revenue to the government, but also less stimulation of investment.

Increasing recognition of the differences in interests between managers and investors (and other stakeholders in the firm) has led to calls for tax policies directed at better aligning those interests. For instance, the view that managers succeed in getting the board of directors (which they arrange to be elected) to vote themselves outlandish salaries resulted in the Clinton administration proposing, and Congress adopting, a tax on salaries of top management in publicly owned corporations in excess of $1 million *that were not based on performance criteria.* This had some unintended consequences. In response, corporations paid executives with stock market bonuses—this seemed to make the compensation performance based. But to a large extent this was not true, since a major determinant of stock market prices are the overall macroeconomic conditions, such as the interest rate and whether there is a boom in the economy. Thus, managers were rewarded for increases in stock prices for which they could in no way claim credit. Worse, it encouraged CEOs to try to manipulate the information they released to the market, so that the stock market would go up in the short run, and so would their pay. Less than a decade later, there would be a rash of corporate scandals based on "creative accounting." This debate has been rekindled and expanded to a discussion of both the magnitude of compensation and the nature of performance criteria for top corporate managers in the aftermath of the financial sector's excesses and subsequent meltdown. (The ratio of the salaries of top managers to that of the average worker in the United States is reportedly more than five times that in Japan, and considerably larger than in virtually all other advanced countries.)

THE CORPORATE VEIL   More generally, our analysis of firm behavior is predicated on the assumption that individuals can understand what is going on inside the firm: that they are indifferent, for instance, when choosing between owning 10 shares in a firm with 1000 shares or 9 shares out of 900 shares in the same firm; that if the firm reduces its debt obligations by $1000, the market will see that the net worth of the firm is now $1000 greater, and its share prices will correspondingly increase; that if the firm invests $1 million of retained earnings, and a shareholder owns

1 percent, it is as if the shareholder himself or herself has invested $10,000 directly. We assume, in other words, that individuals can see through the **corporate veil** to what is really going on.

It is important to realize that all investors need not be well informed about firms' savings and investment for the corporate veil to be pierced. All that is required is that enough investors realize that a firm that has saved and invested $1 million should have a market value $1 million larger than it was before for the stock price to rise by the requisite amount. Uninformed shareholders may not know why the firm's shares have increased in value. All they know is that they have done so. There is a theory, called the *efficient markets theory*, which argues that all information is efficiently and rapidly reflected in stock market prices. This theory has been hotly contested. Grossman and Stiglitz (1980) showed that if the theory were true, there would be no incentive for anyone to spend money to acquire information, since it would instantaneously become available to others; so, the market would actually reflect only costless information. In 2013, the Nobel Memorial Prize was given to both the strongest advocate of this theory, Eugene Fama of the University of Chicago, and one of its most ardent critics, Robert Shiller of Yale University, who presented strong empirical evidence that it was not true.[11]

Whether shareholders see through the corporate veil has important consequences both for the behavior of the firm and for the economy. At the level of the firm, shareholders' seeing through the corporate veil means that, in the absence of taxation, they would be completely indifferent about how the firm financed itself.[12] If the firm borrowed more, shareholders would treat the indebtedness as if the firm had borrowed on their behalf. If shareholders did not like that amount of indebtedness, they could simply reduce the level of debt they had on their own accounts (or they could buy offsetting corporate bonds). If the firm invested its retained earnings, shareholders would treat the investment as if they had made it directly in the firm themselves. In this view then, in the absence of taxation, whether firms paid dividends or retained earnings would make no difference. Accordingly, if taxes give preferential treatment to retained earnings and debt finance (as we will show shortly that they do), then those tax considerations determine how the firm finances itself.

At the level of the economy, the nature of the corporate veil is important because of its implications for aggregate savings. Currently, in the United States, most private savings is done not by individuals directly but

---

[11] Sanford J. Grossman and Joseph E. Stiglitz, "On the Impossibility of Informationally Efficient Markets," *The American Economic Review* 70, no. 3 (June 1980): 393–408.

[12] See F. Modigliani and M. H. Miller, "The Cost of Capital, Corporation Finance, and the Theory of Investment," *American Economic Review* 48 (1958): 261–297; and J. E. Stiglitz, "On the Irrelevance of Corporation Financial Policy," *American Economic Review* 64 (1974): 851–866.

by corporations.[13] The question is: Does increased corporate savings lead to lower household savings? If individuals see through the corporate veil, they treat the savings done by corporations fully as if it had been done on their own account. As long as there are enough informed investors to bid up the price of shares to reflect the new investment, even uninformed investors act as though they saw through the corporate veil, for they respond to the increase in wealth resulting from the increased share prices by increasing current consumption (decreasing savings). In this view then, the division of savings between household savings and corporate savings is purely an artifact of our current tax laws, which encourage savings within the firm. It is *as if* the corporation distributed all its profits to its shareholders and then they decided how much to save. If the corporate veil is perfectly pierced, then, the corporate tax has less effect on the aggregate level of savings than on its allocation: some firms with high cash flows but lower marginal returns on investments may be induced to retain their funds and invest them

> ## MANAGERIAL FIRMS
>
> - Firms may be run in the interests of managers and not be profit- or value-maximizing.
> - A number of tax paradoxes—behavior that is hard to reconcile with profit or value maximization—are consistent with the managerial theory. These tax paradoxes include the following:
>
>   The dividend paradox—firms pay dividends when there are other ways in which profits could be distributed from the corporate to the household sector which would incur lower taxes.
>
>   Firms' frequent failure to take advantage of accelerated depreciation or other tax preferences.
>
> - Shareholders may not be able to see completely through the corporate veil; accordingly, they may react more to reported current profits (which may go down when a firm takes advantage of a tax preference) than to the long-run impact on the firm.

internally, whereas if the funds had been recycled back to the shareholders, they would have found their way to higher return opportunities. However, in a world with perfect or near-perfect information—a world in which it was easy to pierce the corporate veil—even these distortions would not exist, for the firm with high cash flows would seek out the highest *marginal* return investment opportunities; it would not limit itself to investing in itself.

There is both macro- and micro-evidence (that is, evidence both at the level of the economy and at the level of the firm) that individuals do not see perfectly through the corporate veil. We have already referred to several puzzling aspects of corporate behavior that seem inconsistent with the hypothesis that individuals do see through the corporate veil.

Moreover, stock prices are highly volatile. Many economists believe that they reflect true capital values only imperfectly. More importantly, households, at least in the short run, do not pay much attention to the day-to-day variations in the market value of their securities; they do not act as though they viewed these fluctuations as meaningful changes in their

---

[13] In 2011, household savings ($841.9 billion) accounted for 30 percent of gross private savings ($2849.2 billion), and business savings was more than double that amount ($2007.4 billion). (Source: U.S. Department of Commerce, Bureau of Economic Analysis, *National Income and Product Accounts, 1929–2011,* Table 5.1.)

wealth, requiring an appropriate response in consumption levels. Still, in the long run, any policy that led systematically to increased corporate savings would *eventually* have some impact on household savings, if only because such a policy would have a systematic effect on the value of corporations and, through this, on the shareholders' views of their net worth.

DEPRECIATION

We have seen in this chapter how detailed provisions of the tax code can make a large difference in its effect. The distortionary effect of the corporation income tax may be completely eliminated—for instance, if interest can be deducted, and if firms can borrow at the margin to finance their investment.

An equally important provision of the tax relates to depreciation. We saw in Chapter 21 why it was necessary to make some allowance for the depreciation of an asset; as it wore out or became obsolete, it became less valuable; the reduction in the value needs to be subtracted from income. True economic depreciation represents the true reduction in the value of the asset. However, because it is typically difficult to find accurate estimates of this decrease in value, governments typically allow an estimate that is more generous (that is, the present discounted value of depreciation allowances exceed what they would be under true economic depreciation). The standard method is to estimate the lifetime of the asset, say ten years, and to allow one-tenth of the value of the asset to be subtracted every year. This, as we noted, is called *straight-line depreciation*.

Many governments have used still more generous depreciation allowances to encourage greater investment. With **accelerated depreciation**, firms and individuals are allowed to take more depreciation in the early years of the asset's life. The simplest way this can be done is to use a life span that is shorter than the actual one—say, three years, for an asset that will actually last five years. With a $100 asset, this would mean taking $33 over the first three years of the asset (and nothing thereafter), whereas with straight-line depreciation, using the true lifetime, the depreciation allowance would be $20 each year.

The value of accelerated depreciation can be enormous. This is illustrated in Table 23.2. In this table, we consider an asset that has a five-year lifetime and costs $100 but that the government allows to be depreciated in three years. The asset is assumed to yield a constant amount each year ($24) and to have no salvage value after five years. The interest rate is assumed to be 10 percent. At that interest rate, the present discounted value of the annual return equals its cost. In the first column of the table,

TABLE 23.2 COMPARISON OF TRUE ECONOMIC DEPRECIATION, STRAIGHT-LINE DEPRECIATION, AND ACCELERATED DEPRECIATION

YEAR	DISCOUNT FACTOR (10% INTEREST RATE)	1 TRUE ECONOMIC DEPRECIATION	2 PRESENT VALUE	3 STRAIGHT LINE	4 PRESENT VALUE	5 ACCELERATED DEPRECIATION	6 PRESENT VALUE
1	1	16	16	20	20	33.3	33.3
2	1/(1.10) = 0.909	18	16.36	20	18.18	33.3	30.3
3	1/(1.10)^2 = 0.826	20	16.52	20	16.52	33.3	<u>27.6</u>
4	1/(1.10)^3 = 0.751	22	16.52	20	15.02		
5	1/(1.10)^4 = 0.683	24	<u>16.39</u>	20	<u>13.65</u>		
			81.79		83.37		91.2

we show the true economic depreciation, and in the second column the present discounted value of these depreciation allowances (as viewed at the time of purchase). The third and fourth columns show straight-line depreciation, and the fifth and sixth columns show one form of accelerated depreciation. It is clear that there is a very large subsidy—over 10 percent when compared to true economic depreciation. At higher discount rates, and for longer-lived projects, the benefit is even greater. In this example, the difference between true economic depreciation and straight-line—with the true life of the asset—is relatively small (although straight-line depreciation still has a present discounted value of depreciation allowances that is greater than true economic depreciation).

Far more disturbing, however, is the fact that the magnitude of the subsidy varies greatly from one asset to another. Accelerated depreciation rules typically provide a major subsidy for long-lived assets. As a result, not only are some assets favored over others, but industries that use the favored assets also gain at the expense of those that use the less favored assets.

Using a shorter-than-true life span for an asset is only one of several ways that governments have provided accelerated depreciation. The U.S. government, for instance, offers another alternative, called *double declining balance*. Consider a five-year asset purchased on January 1. Instead of allowing 20 percent each year, an allowance of 40 percent is made. Then, the next year, an allowance is taken of 40 percent of the "declining balance"—which amounts to 60 percent of the original value. Thus, the depreciation allowance the second year is 24 percent. The next year, 40 percent of the balance of 36 percent is taken; that is, 14.4 percent. For the remaining two years, straight-line depreciation is used *over the remaining* 21.6 percent of value—that is, 10.8 percent is taken each year.

The 1981 tax law introduced, for the first time, life spans that deliberately had no relationship with actual life spans. For instance, commercial

buildings could be depreciated over fifteen years. The tax savings gave rise to a boon in commercial building, leading to a glut that took almost two decades to work itself out.

The vastly accelerated depreciation allowances for equipment were intended to revitalize "smokestack America," America's heavy industry. Most economists believed that there was no special reason to give preference to heavy industry over other sectors of the economy. Even though the accelerated depreciation was a major windfall for these industries, in fact there was little evidence that the tax provisions played a significant role in accelerating investment in that sector. Because of the concern that the accelerated depreciation allowances were highly distortionary, they were repealed a short five years after they were introduced.

One's view of the overall impact of the corporation tax needs to integrate the combined effects of both interest deductibility and accelerated depreciation. If firms can borrow to finance investment at the margin, and if there were true economic depreciation, then we have seen that the corporation income tax is nondistortionary. But if, as is the case, there is accelerated depreciation, then there will be overinvestment. On the other hand, for firms that are badly credit constrained, what drives investment is not incentives but resources, and high average tax rates can thus lead to underinvestment. The net impact of the corporation income tax is thus to shift investment away from new entrepreneurial, credit-constrained firms to old, established firms that have good access to credit markets.

COMBINED EFFECTS OF INDIVIDUAL AND CORPORATE INCOME TAX

To assess fully the effects of the corporation income tax, we must see how it interacts with the individual income tax. This, in turn, requires us to look more carefully at the relationship between the corporate and household sector.

DISTRIBUTING FUNDS: THE BASIC PRINCIPLES

In Figure 23.2 we have drawn a schematic picture of the relationship between the corporate and the household sectors. Funds flow from the corporate sector to the household sector in the form of dividends, interest,

and share repurchases. Funds flow from the household sector to the corporate sector in the form of new bonds and new equities. Funds flow within the household sector as individuals purchase shares and bonds from each other. Funds flow within the corporate sector as corporations purchase one another and are merged.

The tax authorities treat interest, dividends, and capital gains differently. Interest and dividends used to be taxed identically at the individual level, but since President Bush's 2003 tax cuts, qualified dividends have been given preferential treatment—rather than being taxed at the ordinary income tax rate, they are instead taxed at the lower long-term capital gains rate. Payments from the corporation to the individual that are labeled "interest" are deductible from corporate income, whereas those labeled "dividends" are not. Therefore, corporate earnings that are transferred to individuals in the form of dividends are taxed twice— once in the form of the corporate tax and again by way of the individual income tax. Earnings transferred in the form of interest are taxed only once. If the firm buys back shares, the individual shareholder is taxed only on the difference between the price at which he or she purchased the share and the price at which the firm buys it back. Although there are legal constraints that prohibit a firm from regularly buying back a pro rata share of its stock from each of its stockholders in lieu of paying dividends, firms can simply buy back shares on the open market.[14]

---

[14] Note that if the firm bought back 5 percent of each individual's shares, the fraction of the firm that each owned would remain the same. This transaction is substantively the same as an equivalent cash dividend.

Note, too, that if the firm buys back the shares on the open market, the advantages of the share repurchase are even greater. Each individual could have sold back 5 percent of his or her shares; the fact that an individual chooses to sell back a different amount means that the individual is better off than if he or she were "forced" to sell back 5 percent of the shares. Individuals who bought the shares at a higher price may be more willing to sell their shares, as in doing so, they encounter a smaller tax liability than those who bought the shares at a lower price. Thus, the transfer of funds from the corporate to the household sector will entail an even smaller tax liability than if the firm repurchased 5 percent of each shareholder's shares.

The fact that our tax code does not tax all transactions at the same rate, and taxes transfers in the form of dividends from the corporate sector to the household sector without allowing any deduction from corporate income, as it does for interest, has two basic implications:

1. Avoid transferring income (after paying interest) from the corporate to the household sector whenever possible.[15]

2. When income must be transferred, do it in a form so that it is eligible for capital gains treatment.

The incentive for firms to retain earnings—not to transfer income from the corporate to the household sector—is of concern for several reasons. Managers of firms must persuade potential investors of the merits of their managerial skills and investment projects; with large amounts of funds retained within the firms—and a large tax wedge associated with redistributing money to households—less discipline is required of managers, and with less discipline required of them, they may not do as effective a job either in managing their funds or in reinvesting their proceeds.

THE DIVIDEND PARADOX

Corporations often seem to engage in financial transactions that are not consistent with the preceding principles. Dividends provide one important example. The puzzle of why firms pay dividends when funds could be distributed from the corporate to the household sector in ways that encountered lower tax liabilities such as share repurchases is called, as we have noted, the **dividend paradox**.[16]

A number of possible explanations have been put forward, most of which are not very convincing. One is that dividends serve as a "signal" concerning the firm's net worth. Even though this may be true, buying back shares should be an equally effective signal.[17]

---

[15] There is an important exception, discussed more fully below: since interest payments are tax deductible, it could pay a firm to pay out money to its shareholders, borrowing the money back. There is an immediate cost of such financial restructuring—the taxes on the dividends paid out—but there is a benefit in lower future corporate income taxes.

[16] The dividend paradox was discussed in J. E. Stiglitz, "Taxation, Corporate Financial Policy, and the Cost of Capital" (note 7). Subsequent studies include J. Poterba and L. H. Summers, "Dividend Taxes, Corporate Investment, and 'Q'," *Journal of Public Economics* (1983): 135–167; and Raj Chetty and Emmanuel Saez, "Dividend and Corporate Taxation in an Agency Model of the Firm," *American Economic Journal: Economic Policy* (August 2010): 1–31.

[17] There is considerable evidence, for instance, that the buyback of shares does serve as a signal; firms that buy back their shares see a marked increase in their market value. See, for instance, P. Asquith and D. Mullins, "Equity Issues and Offering Dilution," *Journal of Financial Economics* 15, no. 1–2 (January/February 1986): 61–89; B. Greenwald, J. E. Stiglitz, and A. Weiss, "Informational Imperfections in the Capital Market and Macroeconomic Fluctuations," *American Economic Review* 74, no. 2 (May 1984): 194–199; S. Myer and N. Majluf, "Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have," *Journal of Financial Economics* 13, no. 2 (June 1984): 187–221.

Although many owners of stock are tax exempt (and thus indifferent to whether the firm issues dividends or buys back shares), individual shareholders who pay taxes should prefer share buybacks.[18]

MERGERS, ACQUISITIONS, AND SHARE REPURCHASES

The tax advantages of distributing funds from the corporate to the household sector through share repurchases (as opposed to dividends) can be obtained in other ways. When one firm buys another for cash, the receipts by the owners of the acquired firm are subject to capital gains taxation.

Although many firms have persisted in policies that appear not to minimize total tax liabilities, in recent decades there has been increasing sensitivity to tax concerns.

During the ten years preceding the Tax Reform Act of 1986, mergers, acquisitions, and share repurchases increased enormously. Whereas, in the early 1970s payments for mergers, acquisitions, and share repurchases amounted to approximately 15 percent of dividends, by 1984 they exceeded dividends, and in 1985 they amounted to almost 50 percent more than total dividends. The cost to the Treasury in forgone tax revenues exceeded $25 billion in 1985.[19] Many economists believe that these activities were tax induced; that corporations had gradually come to recognize the advantages of distributing funds to the household sector in ways that subjected them to capital gains taxation.[20]

The 1986 Tax Reform Act not only reduced the tax advantages of capital gains by taxing them at full rates, but also repealed several provisions that resulted in capital gains taxes' being avoided when a firm was liquidated (either when it was sold to another firm, or when its assets were sold, with the proceeds distributed to the shareholders). The effect of these tax changes appears to have been dramatic: share repurchases, which in 1980 had amounted to only 10 percent of dividends,

---

[18] There are financial transactions that are even more puzzling than just paying dividends. For instance, when a firm simultaneously pays dividends and issues new shares, it unnecessarily increases tax payments. If the funds had been left in the corporate sector, the tax on the dividends could have been avoided. Even if some shareholders wanted the cash that the dividend provided, they would have been better off selling an equivalent amount of their shares (to the individuals who would have bought the new share issues).

[19] From J. Shoven, "The Tax Consequences of Share Repurchases and Other Non-Dividend Cash Payments to Equity Owners," in *Tax Policy and the Economy*, ed. L. Summers (Cambridge and London: MIT Press, 1987): pp. 29–54.

[20] But many economists argue that though there may have been tax benefits, these mergers and acquisitions had other motivations. See M. Jensen, "Takeovers: Their Causes and Consequences," *Journal of Economic Perspectives* 2, no. 1 (Winter 1988): 21–48.

had increased to 57 percent by 1985, but by 1990 had dropped back to 34 percent.[21] With the marked increase in preferential treatment of capital gains from the tax laws of 1993 and 1997, we saw a rise in share repurchases once again. However, firms have now gone the other way and substituted dividends for repurchases in response to the 2003 tax reform, because the reduction in dividend tax rates exceeded the reduction in tax rates for capital gains.[22]

DOES THE CORPORATE TAX BIAS FIRMS TOWARD DEBT FINANCE?

We have seen that debt is tax deductible. In the absence of taxation—and in the absence of imperfections in capital markets and information—firms would be indifferent over whether they financed themselves through debt or through equity. This basic idea, developed by Nobel Prize winners Merton Miller and Franco Modigliani, is sometimes explained by an analogy to milk: the market value of milk consists of the value of the cream and the value of the "skim" milk. One can repackage the product, skimming off some of the cream, to form 2 percent milk, but the overall value remains the same. **Debt** and equity represent different ways of packaging the return to an investment. Equity owners get whatever is left over after the claims of debtors are satisfied. With more debt, the amount left over for equity owners is reduced, and the variability of this residual may be greater, but the total amount that will go from the corporation to the households is the same, and hence the market value will be the same.

Taxes change this, because the amounts sent to households in the form of interest payments are deductible from the corporation tax, but the amounts sent to households in the form of dividends are not. Offsetting this advantage of debt, however, there is an advantage to equity: the amounts retained by firms, which increase the market value of the firm, receive preferential treatment—they are taxed only when the individual sells shares, and then at preferential rates. If the individual holds on to his or her shares until death, they escape taxation completely. Moreover, by financing new investment out of retained earnings, the corporation

---

[21] From U.S. Department of the Treasury, *Report of the Department of the Treasury on Integration of the Individual and the Corporate Tax Systems: Taxing Business Income Once*, January 1992. Shoven and Bagwell report a huge increase in the percent of cash distributions through either acquisitions or share buybacks during the decade beginning in 1977—from slightly more than 20 percent to over 58 percent. See L. S. Bagwell and J. B. Shoven, "Cash Distributions to Shareholders," *Journal of Economic Perspectives* (Summer 1989): 129–140.
[22] See J. Blouin, J. Raedy, and D. Shackelford, "Did Firms Substitute Dividends for Share Repurchases After the 2003 Reductions in Shareholder Tax Rates?" National Bureau of Economic Research Working Paper 13601, November 2007.

avoids the tax on the "round trip" that would be involved in first distributing money to shareholders (even if it manages to do so in a way that receives favorable capital gains treatment). Clearly, whether the firm is better off retaining less and borrowing more depends on the relative tax advantages of the interest deductibility and capital gains. As the rates at which corporate profits, dividends, and capital gains have been taxed have changed, the balance has shifted back and forth; overall, the preferences do not appear to be very strong. There continues to be a slight preference for firms to finance as much of the investment as they can out of retained earnings. This is, in fact, the typical pattern, with investments in excess of retained earnings financed by borrowing, as long as the debt-to-equity ratio does not get too large and so long as the enterprise is not credit rationed. Risky new enterprises often have to resort to issuing new equity to raise the capital they require to grow.

Preferences for equity are sufficiently weak, however, that it does not pay firms to restructure—that is, to borrow money to repurchase shares or issue dividends. Doing so would increase their debt-to-equity ratio—and thus increase the fraction of their gross income that they distribute in a tax-deductible manner. However, there is a tax cost of restructuring: in the process of repurchasing shares or issuing dividends, an individual income tax liability is incurred, which otherwise would not have been; the magnitude of this typically is greater than the (present discounted value of the) savings from the fact that interest payments are tax deductible.[23]

Individuals in different individual income tax brackets might, however, argue for alternative policies. Low-income individuals and tax-exempt organizations would prefer that the firm pursue a high-debt strategy, as they incur no (or few) additional taxes upon restructuring, but the corporation saves money. The fact that different individuals would like the firm to pursue different strategies suggests that different firms should have different clienteles. It is not apparent to what extent they in fact do, perhaps because these tax effects are not as important as other considerations.

If Modigliani and Miller were correct, the distortions in how a firm financed itself would be of little import, but, in fact, for a variety of reasons, how a firm finances itself does make a difference. For instance, firms that are heavily debt financed may be particularly vulnerable to the threat of bankruptcy in the event of an economic downturn. By encouraging heavy debt finance, the tax laws may exacerbate the economy's fluctuations.

---

[23] The result clearly depends on the tax rates. For instance, if the corporate income tax rate is much higher than the tax rate on capital gains, then a restructuring engineered through a stock repurchase may be desirable.

DISTORTIONS IN ORGANIZATIONAL FORM ARISING BECAUSE SOME FIRMS DO NOT HAVE TAXABLE INCOME

We have seen that the desire to distribute earnings in a form that receives preferential treatment may partially account for share repurchases, mergers, and acquisitions. Another aspect of the tax system has perhaps an even larger impact on organizational design: the limitations on loss offsets. Many firms have profits so small that they cannot fully take advantage of depreciation allowances, and still other firms have losses.

Debt financing or economic losses on past investments may lower taxable income to the point at which companies cannot use all their depreciation deductions and tax credits.

Consider a firm with large depreciation allowances, say, because of accelerated depreciation on some new assets, which at the same time is making losses in a variety of other lines. If its taxable income is negative, the depreciation allowances have no value; when the tax is already zero, it cannot be reduced further.

Assume that the firm still will be doing poorly when the returns to the investment occur. Then, because the returns will not be taxed—there will be losses in other parts of the firm to offset the returns on this productive investment—the tax system may cause no distortion. However, most firms that are doing badly and are investing anticipate doing better in the future; this implies that although the firm is not able to take advantage today of the depreciation allowances, it will have to pay taxes in the future on the returns to that investment, just like any other firm. Thus, the tax system causes a strong distortion *against* investment in firms that are currently not doing well, and helps to perpetuate their weak position.

TAX-INDUCED MERGERS   The market always attempts to find ways of dealing with inefficiencies created by the tax system. One way is for a firm with losses to merge with a firm with profits. This is referred to as a **tax-induced merger**.

COMBINED EFFECTS OF INDIVIDUAL AND CORPORATE TAXATION

- The effects of taxation depend on the interaction of the individual and corporate tax systems.
- Whereas debt financing receives preferential treatment under the corporate income tax, capital gains receive preferential treatment under the individual income tax. Whether overall debt or equity is tax preferred depends on the tax bracket of the individual.
- Distributing money from the corporate to the household sector in the form of dividends rather than share repurchases or through mergers and acquisitions (in which case the distributions would receive capital gains treatment) lowers the overall tax.
- Even if it pays a firm to borrow to finance new investment beyond that which it could finance through retained earnings, it may not pay a firm to restructure itself, for example, through borrowing to buy back shares, to increase its debt/equity ratio.
- Whether corporations overall are tax preferred depends on the tax bracket of individuals, and on the extent to which capital gains receive preferential treatment.

Many economists are concerned about the long-run consequences of these mergers. Such mergers may limit competition in the economy. More importantly, the mergers may not be based on underlying economic considerations, such as economies of scope or synergies between the parts of the merged firms. Managerial talents may be stretched, so overall performance may be decreased. Moreover, it is believed that the vitality of a capitalist economy depends on there being a large variety of firms, each with its own strengths and thus in a position to take advantage of different situations. Some have made the analogy between firms and species of animals. Just as it may be advantageous to preserve a rich genetic pool, to be drawn upon in a variety of circumstances, so too may it be desirable to have a diversity of firms in the economy. The provisions of the tax code that encourage mergers (mergers that, apart from taxes, would not be undertaken) should, in this view, be altered.

ARE CORPORATIONS TAX PREFERRED?

Would it pay an individual who owns a business to incorporate? If this individual incorporates, he or she could pay out all "profits" in the form of wages, thus avoiding the corporate income tax, or could choose to retain the earnings. On the retained earnings, the individual would have to pay the corporate income tax—and if the funds are reinvested, he or she would have to pay the corporate income tax on the earnings from those investments. When the individual finally wished to get the funds, he or she would have to pay taxes on the distribution; if the individual managed it well, he or she might succeed in getting those

DISTORTIONS FROM THE CORPORATION TAX

The corporation tax distorts level of investment.

- May result in underinvestment, especially for credit-constrained firms.
- With accelerated depreciation, firms that can finance marginal investment by borrowing may invest too much.

The corporation tax results in distortions in the kinds of investments that are made.

- Encourages long-lived versus short-lived capital goods.
- Encourages investment in assets that can be collateralized (more debt-financed), compared with those (like R&D) that cannot.
- Encourages investment in industries, like real estate, that have higher debt-to-equity ratios, or industries that have more long-lived assets.

The corporation tax may alter the form of financing.

- Net effect depends on combined effects of corporate and individual income taxes.
- Deductibility of interest encourages debt financing.
- Preferential treatment of capital gains encourages financing through retained earnings.

The corporation tax may affect the organization of production.

- May discourage incorporation (with limited liability).
- Under current circumstances, net impact may be small.

Enterprises having easy access to credit may be preferred (face lower cost of capital) to those which are credit constrained.

Enterprises with profits may be preferred to those without profits.

- May give rise to leasing and tax-induced mergers.

Monopolies may be affected differently from firms in competitive markets.

distributions taxed at the favorable capital gains tax rate (and if the firm is left to the children, who then sold the firm, taxes might be avoided altogether). Clearly, whether incorporation pays depends on the relative tax rates on corporations and individuals. For an individual at the maximum individual income tax bracket of 35 percent and the corporation tax at 35 percent, there might be a slight preference for incorporation if funds are retained within the firm long enough, and when they are finally distributed, receive favorable capital gains treatment.

More generally, income earned within a corporation is effectively double taxed—once within the firm, and again when it is distributed to households. There is little evidence, however, that this has a marked effect in discouraging incorporation.

CALCULATING EFFECTIVE TAX RATES

We have seen that the full effect of taxation on investment in a particular asset, in a particular industry, depends on a myriad of features of both the individual and corporate income tax, including the tax treatment of capital gains, losses, dividends, and interest; whether firms are or are not credit constrained; and whether they are profit-maximizing or managers take decisions in their own interests. There are further issues we have not discussed here, such as property and state income taxes, and how those tax payments are treated under the individual and corporate income tax. Needless to say, calculating the marginal tax rate associated with an additional unit of investment is not easy. What is clear, however, is that the effective marginal tax rate typically differs markedly from the average tax rate, as well as from the marginal tax rate stated in legislation.

There have been several attempts at a full calculation of the effective marginal tax rate, taking into account all the marginal taxes—corporate, property, and personal—that are paid as a result of a new investment. Perhaps the most thorough study of overall effective marginal tax rates on investment is that of Don Fullerton and Yolanda Henderson.[24] They asked: If an individual invests a dollar more in an asset that yields a before-tax return of, say, 10 percent, what will the after-tax return be, after paying property taxes, capital gains taxes, corporation taxes, taxes on dividends, interest, and so on? Alternatively, what before-tax rate of return is required if the individual is to obtain, say, an after-tax return of 10 percent? The striking result of their study was that for firms that could finance their *marginal* investment by debt, the effective marginal tax rates

---

[24] D. Fullerton and Y. K. Henderson, "Incentive Effects of Taxes on Income from Capital: Alternative Policies in the 1980s," in *The Legacy of Reaganomics: Prospects for Long-Term Growth,* ed. C. R. Hulten and I. V. Sawhill, (Washington, DC: The Urban Institute Press, 1984), pp. 45–80.

(for equipment, structures, public utilities, inventories, and land) were all very negative, so much so that, for instance, under the 1981 tax law, corporate equipment yielding a before-tax return of 210 percent would yield an after-tax return of 110 percent. If, for some reason, firms could not finance their marginal investment by debt, then they faced positive effective marginal tax rates. Still, the overall effective marginal tax rate on capital appeared to be lower than on most wage income. Another surprising result of their analysis was that owner-occupied housing—which has often been thought to be "tax preferred," as no taxes are paid on the "imputed rent"—actually is taxed more heavily than other residential structures. The Fullerton–Henderson study was based on the highly distorted 1981 tax law; subsequent reforms, especially those enacted in 1986, substantially increased effective marginal tax rates in the corporate sector and on non–owner- occupied residential structures. These reforms, however, have not succeeded in creating a true "level" playing field.

THE CORPORATION TAX AS ECONOMIC POLICY

The distortions discussed in this chapter represent largely unintended consequences of the corporation tax. However, the corporation tax has also been used as a tool of economic policy. One of the purposes for which taxes on capital have been used is economic stabilization: to encourage investment in economic downturns and to slow investment when it appears that the economy is overheating. This was the original motivation for the introduction of the investment tax credit (see case study, "The Proposed Incremental Investment Tax Credit of 1993").

Although changes in tax rates and credits are sometimes used to stabilize the economy, without these adjustments, the corporate income tax would probably exacerbate business fluctuations. A significant part of economic fluctuations is due to variability in investments in small and medium-size enterprises[25] that typically have limited access to capital markets, especially in periods of economic downturn. In a recession, they face a shortage of funds, and are forced to cut back on investments. The corporate profits tax reduces the funds that they have available for reinvestment, and thus exacerbates the decline in investment. Some have suggested that if the government were more evenhanded in its risk

---

[25] S. Fazzari, G. Hubbard, and B. Peterson, "Investment Financing Decisions, and Tax Policy," *American Economic Review* 78, no. 2 (May 1988): 200–205.

THE PROPOSED INCREMENTAL INVESTMENT TAX CREDIT OF 1993: AN IDEA BEFORE ITS TIME?

The investment tax credit was first introduced in 1961 by President Kennedy, in an attempt to stimulate what then appeared to be a lackluster economy (with unemployment standing at 6.7 percent). Until 1981, it had been used mainly as a macroeconomic instrument, to provide a needed stimulus to the economy to help it recover from a recession; but in 1981, it was introduced as a "permanent" feature of the tax code to encourage investment, with a 10 percent credit for long-lived investments and 6 percent credit for shorter-lived assets. Shortly thereafter, the distortionary impact of the tax credit arising from the different treatment of different assets became apparent (although economists recognized it before the credit was enacted), and in 1986 even President Reagan advocated its repeal.

With the economy again in a slowdown in 1993, the newly elected President Clinton searched for low-cost ways to stimulate the economy. One idea was an investment tax credit—based, however, not on total investment, but on *incremental* investment, the additional investment that firms made over and above what they had made in previous years. Economists had long claimed that what matters

for investment decisions (as other decisions) is *marginal* cost, not average cost. By lowering the marginal cost of investing, the incremental investment tax credit might have provided a strong stimulus to the economy, at a fraction of the costs of a full investment tax credit (as the incremental investment would have been but a small fraction of total investment).

If firms had known that such a tax credit would be levied, the credit might have had distortionary effects, as firms reduced their levels of investment before the credit went into effect to make more of their investment eligible for the credit. However, as the credit was explicitly made temporary, and the base from which increments were to be measured was investment in previous years, there was no point for firms to try to "game" their investment strategies.

Unfortunately, most firms saw little payoff in the proposal; they were more concerned with receiving money than with getting incentives, and the whole proposal was designed to minimize the amount of money transferred to the corporate sector while providing strong incentives. Without strong corporate backing, the proposal died in Congress.

sharing—sharing not only profits in good times, but also losses in bad times—it would help stabilize investment. Under current law, firms are allowed to carry forward losses (that is, they can deduct losses incurred in 2011 from profits they earn in 2012), but no interest is paid, and the promise of future tax breaks does little to help them over the cash flow problems that they face today.

Besides stabilization, the most common policy use of the corporation tax is to promote investments in some industries (which means, in a

full-employment economy, at the expense of investments in other industries). For instance, the favorable depreciation allowances enacted in 1981 were designed to help restore America's heavy industries. The special tax preferences afforded to the oil and gas industry are another example. Most economists believe that there is no economic justification for these special provisions; they are the result of political pressures from special interest groups.

Capital taxation, both within incorporated and unincorporated enterprises, is, in fact, subject to a myriad of special provisions, so much so that, by one estimate, 80 percent of capital income receives some kind of preferential treatment.[26] While advocates of these special provisions claim that they help improve the performance of the economy by encouraging particularly worthy activities, in many, if not most cases, they simply reflect the influence of special interests. While a case can be made for encouraging renewable energy, given the externalities associated with global warming, such an argument cannot be made for provisions benefitting investments in fossil fuels—just the opposite, they should be taxed. Yet, they too benefit from special treatment.

TAXATION OF MULTINATIONALS

Today, most major companies operate in many countries. Some of these are American companies, such as Ford and General Motors, which have subsidiaries around the world. Others, such as Toyota, Nestlé, and Philips, are foreign companies that not only ship goods to the United States, but also produce here. Moreover, foreign companies have many American shareholders, just as many American companies have many foreign shareholders. The largest single shareholder in the largest U.S. bank is a Saudi Arabian. Some American companies are more than 50 percent foreign owned. All that it means to be "American" is that the company's official "home" is in the United States; it is incorporated here. It does not mean that the company is owned by Americans; or even that most of its production is here.

Indeed, production is increasingly occurring on a global scale, with parts gathered from all over the world. A label on a computer or car—"Made in USA" or "Made in Korea"—may mean only that it was the place where the product was assembled. The fraction of "value added" or the fraction of total labor costs occurring in that country may be relatively small.

---

[26] See E. Steuerle, "Is Income from Capital Subject to Individual Income Taxation?" *Public Finance Quarterly* (July 1982): 283–303; and J. Gravelle, *The Economic Effects of Taxing Capital Income* (Cambridge and London: MIT Press, 1994).

This presents real problems for tax authorities. What is the "income" on which the corporation tax should be levied? There is a naïve answer: the tax should be levied on that part of the income attributed to economic activity in the country. The problem is how that is to be ascertained. Assume that USAComputer assembles a computer in the United States, using parts made by its factories in Hong Kong, Singapore, Taiwan, Malaysia, and Korea. If USAComputer had bought those parts from an unrelated supplier, calculating U.S. income would be easy. We would simply subtract from its gross sales its wages and other costs of production here as well as the costs of all the parts purchased abroad. However, USAComputer owns its own factories, and does not sell its parts to anyone. There is no market price. To calculate its tax liability, it must make up a price, called a *transfer price*; this is an estimate of what the market price would have been, had it purchased the item in an arm's-length transaction with a third party.

If corporate tax rates in the United States are higher than in the countries where it produces its parts (which is typically the case), then USAComputer has a strong incentive to try to increase its profits there and to decrease its profits here, by claiming a high transfer price. With a high transfer price, U.S. profits are low. The company may claim that the high price is warranted by the high quality, a result of the high skills of the workers in its manufacturing plants. The IRS may claim that these components are little different from those produced by other producers, which sell for pennies in the open market. Alternatively, it might claim that most of the value of the parts is a result of research conducted in USAComputer labs in the United States, and, accordingly, that the value added by the parts manufacturer abroad is small. Perhaps the most egregious example of such tax avoidance involved Apple, which seemed to have been as (or more) inventive in designing tax avoidance schemes as it was in designing products consumers valued. In a scheme that came to be called the "double Irish," Apple shifted large amounts of its profits to Ireland, but then took advantage of a provision in the Irish tax code that allowed it to escape even Ireland's low corporate income taxes. It should be clear that even with relatively intense scrutiny by the IRS, manufacturers have considerable discretion to shift profits from one country to another through transfer pricing, and that preventing more such shifting from occurring requires enormous diligence, and costs, by the IRS.

Moreover, in some cases, the corporation is put into a seemingly impossible situation, with, say, Japanese tax authorities—who would like more of the income attributed to activity in their country—arguing that the firm is using too low a transfer price, whereas the United States claims that the firm is using too high a transfer price.

FOREIGN INCOME AND THE CORPORATION INCOME TAX

The issue of how to treat foreign income of America's corporations is a seemingly arcane subject that has risen to the top of the political agenda. The reason is that there is a concern that the current tax provisions are encouraging jobs moving abroad.

Companies do not have to pay taxes on their income until they bring it back to the United States (called *repatriation*). However, that means if they have earned income abroad, they can defer paying taxes as long as they reinvest it abroad. It is like having an IRA account, the tax is paid only when the money is taken out, or in this case, brought home from abroad. Because of the tax system, firms may be better off putting jobs abroad (especially in a low tax jurisdiction) than creating them in the United States, even though without the tax distortion, it would pay the firm to create the jobs here.

There are three possible responses. The first is to tax corporations on their worldwide income, as they earn it (with full credit, of course, for taxes paid to other jurisdictions). Although this is very difficult to do, if implemented effectively, it would end the incentive to keep money outside the United States.

The second is to impose a "global minimum tax," whereby corporations would have to pay a minimum tax based on their global income. They would be given full credit for income taxes paid to other jurisdictions, and the difference between this and their global minimum tax would be due to the U.S. government. This strategy mitigates many of the administrative difficulties of the first alternative.

The third, favored by most corporations, is to impose taxes only on income earned in the United States. This is what many other countries do. Critics of this perspective, however, argue that American firms expect their government to advance their interests in other countries—protect their property, use diplomacy to open up foreign markets for their goods, and so forth. They want these benefits, but do not want to pay for them.

Moreover, we described in the text how firms can move money around. Even though their profits really originate from, say, research conducted in the United States, they can try to claim that the source of their profits lies elsewhere. Under the current regime, there are already incentives for shifting claimed income to low-taxed jurisdictions; under the third alternative, matters would be much worse.

The huge costs and potential for disputes associated with the transfer price system have led to arguments for a *unitary* tax system, in which taxes are levied on a proportion of a firm's *worldwide* income. The portion is set by a formula, which looks at the fraction of employment, assets, and sales occurring within the country. Similar systems are used by states within the United States for levying state corporation taxes; indeed, given the close intertwining of production across state lines, trying to use a transfer price system within the United States would probably be close

to impossible. As more production occurs within multinational firms, the problems of levying the corporation tax using a transfer price system are likely to become more apparent.

With global corporations, taxes can affect where production occurs and where and how funds are raised. The United States, in setting its tax regime, must worry about how its tax regime interacts with those of other countries. For instance, if it sets tax rates too high, firms may be induced to shift their production abroad. In fact, a special provision of the U.S. tax code has probably encouraged investments abroad. Firms do not have to pay taxes on profits earned abroad until they "repatriate" them home (i.e., bring the funds to the United States). Thus, if they invest in a low tax jurisdiction like Ireland, it pays for them to reinvest their profits there. It is almost as good as a tax-free IRA account.

SHOULD THERE BE A CORPORATION INCOME TAX?

The rationale for the corporation income tax has never been completely clear. Some believe that corporations, like individuals, ought to pay taxes. Most economists find this argument unpersuasive, however, as it is not the corporation that pays the tax, but people: those who work for the corporation, those who supply capital to it, and those who buy the goods produced by it.

Although politicians often justify the corporate tax in terms of its progressive effects, it is possible that the tax has no significant redistributive effect. This is hard to determine because of the difficulties of ascertaining who really bears the corporate tax burden. The tax can be viewed as a tax on the corporate form of organization (on limited liability). Is there any reason why the government should wish to discourage this form of organization, or to penalize those who derive income from it?[27]

INTEGRATION OF THE CORPORATE AND INDIVIDUAL INCOME TAX   In fact, because the advantages of incorporation are so great, the tax may not have a significant effect in discouraging incorporation. There are, nevertheless, concerns about equity, and about the wider range of distortions in the form of finance that result from the interaction of the corporate and individual income tax. This has led to proposals

---

[27] With limited liability partnerships, however, businesses can have the advantage of limited liability without paying the corporate income tax.

for integrating the individual and corporate income tax systems. A simple form of integration would impute the earnings of a corporation to shareholders; corporations would be treated (for tax purposes) as if they were partnerships, in which profits were credited to the "account" of each shareholder. For the most part, such proposals have received, at best, a lukewarm reception, both from corporate executives and from shareholders. Shareholders worry that they would face tax liabilities, even though they have received no checks from the corporation. (They could, of course, simply sell some of their shares.) Corporate executives worry that they would be under pressure to distribute more of the company's profits, thus reducing their degree of discretion. Indeed, corporate managers have been so concerned with these pressures that they did not even support proposals put forward to make dividend distributions tax deductible.[28]

Equally important, most forms of integration would reduce revenues, and in the budget stringency of recent years government officials have preferred to retain the corporate income tax in its current form, with its ambiguous incidence. If tax cuts are to be made, politicians have chosen to grant them in forms that seem to yield higher political payoffs—such as child care or education tax credits.

WHY IS THERE A CORPORATE INCOME TAX AT ALL?

Some critics of the corporate income tax have gone so far as to question why there is any corporate income tax at all. With full **integration of corporate and individual income taxes**, there would, in effect, be no corporate income tax. Short of full integration, though, without a corporate income tax, funds retained within the firm would escape bearing taxation until the funds were distributed—and for corporate holdings passed on to heirs, taxation could be completely avoided. In effect, the return to capital earned within a corporation would escape taxation. A corporation income tax is a necessary part of an individual *income* tax system. As we have seen, its distortionary effects may be limited so long as its rates are similar to those of the individual income tax.

---

[28] There are, however, problems of implementation. Presumably the tax would be levied on the basis of who owns the shares at a particular date. There would be incentives to sell the shares to a tax-exempt or low-taxed individual or institution to be held just over that date.

REVIEW AND PRACTICE

SUMMARY

1. The corporation tax is often viewed as a tax on capital in the corporate sector. The effective tax rate depends on a variety of details, including depreciation allowances and the fraction of debt financing. In the long run, if savings is fairly elastic or if capital is mobile internationally, most of the burden of the tax rests on consumers and workers.

2. If the supply of capital in the economy is fixed and the economy is competitive, the effect of the tax is to shift capital out of the corporate sector into the noncorporate sector, because after-tax returns in both must be the same. After-tax returns to capital may be lowered by even more than the tax.

3. Under our present tax system, interest payments are tax deductible. This means that if marginal investment can be thought of as being financed through debt, a corporation tax with true economic depreciation causes no distortion in the investment of the firm. With accelerated depreciation, investment in the corporate sector is encouraged. The tax is best viewed as a tax on credit-constrained firms, which include many new firms.

4. There is no reason to believe that firms finance new investments in the same way as they financed their previous investments, so taxes may affect the marginal cost of capital differently from how they affect the average cost of capital.

5. If the corporate sector is noncompetitive, the tax is partially a tax on monopoly profits, and, to that extent, it is nondistortionary. However, the tax may also be a tax on the return to corporate capital, and, to that extent, it may increase consumer prices by more than the increase in the costs of production resulting from the tax. There may appear to be more than 100 percent shifting.

6. In assessing the impact of the corporation income tax, one needs to consider the effect of the corporation tax simultaneously with the effect of the individual income tax. The total (corporate plus individual) tax liability associated with a marginal investment depends on how that investment is financed, whether through debt or through equity. The tax structure may affect how firms raise capital.

7. The fact that firms pay dividends when there are other ways of distributing income to shareholders that result in lower total tax payments is called the *dividend paradox*. It is only one example of paradoxical behavior by firms, in which they do not seem to minimize their tax liabilities.

8. The corporation tax falls unevenly on different forms of corporate investment, thereby biasing investment decisions toward certain favored assets or industries and against others that are not so favored.

9. Many economists believe that the corporation and individual income taxes should be integrated.

10. With the growth of multinational firms, there are serious problems in administering a corporate income tax—in particular, in ascertaining how much income (profit) should be attributed to each country. There are two approaches, the arm's-length transfer pricing approach and the unitary approach.

KEY CONCEPTS

Accelerated depreciation

Corporate veil

Debt

Dividend paradox

Integration of corporate and
    individual income taxes

Managerial firms

Tax base

Tax-induced merger

QUESTIONS AND PROBLEMS

1. Discuss some of the controversies concerning who bears the burden of the corporation income tax. To what extent are differences in views accounted for by differences in assumptions concerning the nature of the tax?

2. Is it possible for (a) the price of output of the corporate sector to rise by more than the tax revenues collected (per unit of output)? (b) The after-tax rate of return in the corporate sector to increase, after the imposition of the corporate income tax? Give conditions under which either of these may occur.

3. Discuss the problems that arise when the corporation tax rate exceeds the highest personal income tax rates by a substantial amount.

4. Many firms pay their top executives with stock options, which give them the right to purchase shares in the company at a fixed price. When the firm does well, the value of the stocks increases, and hence the value of the options increases. Moreover, the income they obtain this way receives capital gains treatment. Some critics of stock options claim that similar incentive effects can be obtained by tying executives' pay to the performance of the stock, but that paying executives directly has overall favorable tax consequences, once all taxes—including corporate taxes, the taxes paid by executives, and the taxes of shareholders—are taken into account. Discuss. (When the company pays executives directly, the wages are deductible from the firm's income subject to the corporate income tax; the "costs" of stock options are not deductible.)

5. There have been proposals to allow firms interest on the losses they carry forward on their tax returns from one year to the next. That is, if a firm has a loss this year of $100,000, and the interest rate is 10 percent, it can deduct $110,000 from its income next year (assuming that net income is positive). Why might such a proposal be desirable? Would it completely resolve the problems that it is intended to address?

6. Why do economists place so much emphasis on the difference between average taxes and marginal tax rates? Under what circumstances might these two differ significantly? Are there any circumstances in which you might be particularly concerned about what the average tax rate is?

7. Compare the taxes an individual would pay if he or she had a million dollars to invest in a machine that lasts one period only, yielding a gross return of $1.2 million, if the individual incorporates and if he or she does not. Assume that if the individual incorporates, he or she (a) lends the company the million dollars to buy the machine or (b) invests the money as equity. Also, assume that if the individual incorporates and provides capital to the firm in the form of equity, he or she (a) pays out the net profits as dividends, (b) manages to distribute the funds in a way that gets favorable capital gains treatment, or (c) dies next year, before the profits have been distributed, and leaves the firm—with its cash position of $200,000—to his or her son, who manages to sell the firm for $200,000. Assume that the individual can subtract the full million dollars as depreciation, and that interest is tax deductible.

8. Compare the present discounted value of taxes an individual who is the sole owner of a corporation that has $1 million in profits would pay under the following two scenarios: (a) the indivdual pays out the profits to himself or herself, invests them in a bond yielding 10 percent, which he or she holds for seven years; or (b) the individual retains the profits inside the corporation, and invests in an asset yielding 10 percent. After seven years, the individual sells the asset, which has retained its original value, and distributes the proceeds to himself or herself. (For simplicity, look at two cases: one in which the individual is in the 40 percent marginal tax bracket and pays a 20 percent capital gains tax rate, and the other in which the individual is in the 15 percent marginal tax bracket and pays a 10 percent capital gains tax rate.)

24

A STUDENT'S GUIDE TO TAX AVOIDANCE

There is a widespread belief that the rich are able to avoid paying much of the taxes that they would otherwise owe by taking advantage of loopholes in the tax law. Although tax laws change, there is a constant duel between the government and the tax lawyers, with the tax lawyers developing new loopholes almost as fast as old ones are closed.

From the public policy point of view, it is imperative to understand the nature of tax loopholes, for two reasons. First, the total impact of the tax law depends as much on these special provisions as it does on the law's overall design. Enacting a progressive tax structure may make little difference if the loopholes provide a method by which the rich can avoid paying high tax rates. Second, distortions in the patterns of investment and savings caused by these special provisions may be more significant than distortions in the level of savings and investment caused by uniform capital taxation.

We are concerned here with **tax avoidance**, as opposed to **tax evasion**. Tax evasion is illegal; tax avoidance entails taking full advantage of the provisions of the tax code to reduce one's tax obligations. Tax evasion includes not reporting any or all of one's income—that is, failing to file a tax return or underreporting income when filing a return. Tax avoidance

entails compliance with the tax laws, but recognizing that they tax different forms of income differently. Provisions of the tax code that allow an individual to "escape" paying taxes—or to reduce tax obligations—are called **loopholes**. However, there are often disagreements about what constitutes a loophole. Consider, for example, a provision that encourages expansion of the oil industry. Critics, especially those who view the provision as unwarranted and a result of the influence of a special interest group, will label the provision a "loophole" because it reduces taxes of investors in the oil industry, whereas advocates will describe it as a *tax expenditure*, a reflection of a deliberate government decision to use tax incentives to encourage this vital industry. Like beauty, loopholes often are in the eyes of the beholder.[1]

In 1986, during the administration of Ronald Reagan, the U.S. tax system was substantially reformed; one of the explicit aims of the reform was to make tax avoidance more difficult. Changes in tax laws since then have introduced a variety of new special provisions. Advocates claimed that these changes would encourage education and investments (especially investment in innovative small businesses); critics claimed that they reopened old—and opened some new—opportunities for tax avoidance.

PRINCIPLES OF TAX AVOIDANCE

There are two basic principles involved in income tax avoidance. The first is postponement of taxes. The second is taking advantage of differences in tax rates for different types of income, and between income for different types of taxpayers, by shifting income from high-taxed categories to lower-taxed categories.

POSTPONEMENT OF TAXES

A dollar today is worth more than a dollar next year. Accordingly, if one has a choice, it is always better to postpone one's taxes (assuming, of course, that tax rates do not rise). There are several major methods of postponing taxes.

---

[1] Some loopholes are inadvertently put into the tax law as a result of errors in writing legislation. Some of these are corrected in the "technical corrections acts" that are passed a year or so after the passage of every major tax act. The fact that such errors occur with such regularity is testimony to the complexity of the tax system—the difficulty of making precise legal definitions in a complex economy.

ACCOUNTING TRICKS Accounting devices can be used to postpone the recognition of income. For instance, one way to postpone the capital gains tax on the sale of an asset is to postpone the date at which the transfer of the asset finally occurs. When an individual buys a business (or any other large asset), the seller often lends the buyer part of the purchase price, which the buyer repays over several years. When does the sale of the asset actually occur and, hence, when must the seller pay capital gains tax? Is it when "control" of the asset is transferred, or when the buyer pays off the loan? The answer depends at least in part on how the sale is "designed." If title is not transferred until all funds are received, the later payments may be deemed payment of part of the purchase price, rather than debt repayment. In this case, the seller will be able to postpone the capital gains tax. (Such transactions are called *installment purchases*.)

In construction projects and defense contracts, payments made prior to the completion of the contract are sometimes viewed as "loans" to the contractor, rather than payment for the project. This allows the recipient to defer payment of income taxes until the project is complete and the debt is paid off.[2] Almost half of the projected increase in corporate tax revenues under the 1986 reform act was due to changes in accounting rules, including those related to construction and defense contracts. In fact, many of these accounting gains did not materialize.

CAPITAL GAINS AND THE POSTPONEMENT OF TAXES Capital gains on an asset, as we have observed, are taxed only upon realization— that is, when the asset is sold. If one buys a capital asset and its value goes up, one can postpone paying the tax simply by not selling the asset. If one would like to sell part of the asset in order to buy, say, some consumer goods, it may be better to borrow, using the asset as collateral. This method has a further advantage: if one postpones the sale until death, no capital gains tax is due (even by one's heirs).[3] Standard estimates suggest that the ability to postpone the capital gains tax alone reduces the effective tax rate by 25 percent.

---

[2] In this case, the tax advantages are related primarily to the differences in tax rates for the two sides of the transaction. For example, say one firm hires another to perform a task, if the first firm treats the payment to the second as a loan, it cannot take a business deduction, whereas the second firm does not record the receipt as income. If the first firm is in a lower tax bracket than the second, then the present discounted value of the *total* taxes of the two is decreased by postponing the tax; the two parties can split the gain between themselves. If an individual is purchasing, say, a home, then his or her payments to the contractor are not tax deductible, so there is an unambiguous gain from postponement. These examples illustrate a general principle: in assessing the tax impact of any particular arrangement, one has to look at how all the parties to the transaction are affected.

[3] The relevant provision is called a "step-up of basis." Capital gains taxes are due on the difference between the sale price and the acquisition price (called the *basis*). When an individual dies, his or her heirs take as the basis (the price at which they in effect acquired the asset) the price at the date *they* acquired the asset, not the price at the date their benefactor acquired the asset. Curiously, if a parent gives (rather than wills) an asset to his or her child, then the "basis" is not "stepped up," but remains the value at which the asset was originally acquired.

SHIFTING AND TAX ARBITRAGE

The second major strategy for avoiding taxes is based on the fact that income that accrues to different individuals is taxed at different rates, or that different kinds of income are taxed at different rates.

INCOME SHIFTING   Under a tax structure with increasing marginal rates, a taxpayer at a high marginal rate will always want to "shift" income to a taxpayer with a low marginal rate. In particular, it pays for parents to shift income-producing assets to their children. The 1986 Tax Reform Act tried to limit such shifting by taxing those under age 14 at the marginal tax rate of the parent (if it is higher).

There are several important points to note about **income shifting**. First, *typically it requires the transfer of an asset, such as stocks, bonds, real estate, or a share in the parents' business.* Working parents cannot simply ask their employer to make out their paychecks in their children's names.

Second, *income shifting works simply because of the fact that marginal tax rates increase with income.* With a flat-rate tax structure, in which the marginal rate is constant (the individual is taxed, at a fixed rate, on the excess of his or her income over some exemption level), there is no incentive for income shifting, provided that the exemption level for a family is proportional to the number of individuals in the family. Thus, the 2001 tax law (which lowered top rates) reduced the incentive to shift income, whereas the 2013 law (which raised top rates) increased it. With a current 39.6 percent top marginal tax rate, shifting $1000 from a parent to a 15-year-old child saves $396.[4]

Third, *there is a limit to the tax savings an individual can achieve through income shifting.* Consider a self-employed family of two adults and two children, and total family taxable income of $220,000. By shifting $81,000 to the 15-year-old son, the parents can reduce total family taxes by $6679.50,[5] a substantial amount, but still a relatively small percentage of their total tax liability.[6]

Fourth, *the government has attempted to limit income shifting, with only partial success.* Consider the problems posed by divorce: How should income from an ex-husband to his ex-wife and their children be treated?

---

[4] One quirk in the tax law that encourages tax shifting for those who own businesses is that a parent who hires a child does not have to pay Social Security taxes on the child's behalf. Although the provision was motivated by a concern for small businesses, the tax shifting advantages are significant.

[5] The total tax is minimized by transferring income until the marginal tax rates of the parents and son are equal. (Once marginal tax rates are equal, there are no further benefits from shifting income from one party to the other; this limits the amount of tax savings.) For couples filing jointly, the marginal tax rate on income over $212,300 (in 2011) is 33 percent. Of the $81,000 shifted to the son, $8500 is taxed at the 10 percent rate, saving $1915; $26,000 is taxed at 15 percent, saving $3380; and $46,500 is taxed at 25 percent, saving $1384.50.

[6] With income of $220,000, the tax would be $50,054.50, and hence the reduction in taxes would be just 13 percent.

Under current law, alimony is excluded from the income of the payer and included in the income of the recipient. Therefore, characterizing payments from an ex-spouse as alimony (rather than as a property settlement or child support) may have significant tax advantages, if the payer has a higher income than the recipient.

Although income shifting among members of a family represents a highly visible way in which upper-income families reduce their total tax liability, more important in terms of lost tax revenue are two other forms of shifting: (a) shifting of income among corporations; and (b) shifting income into a form that takes advantage of the preferential treatment afforded to capital gains.

CORPORATE SHIFTING   Corporations are allowed to deduct depreciation allowances, and, at times, there have been large investment tax credits. If a firm has no income, however, a depreciation allowance is of no value. Consider a firm, which has been making losses, that needs some cars or trucks. If it purchased the vehicles itself, the depreciation allowance would have no immediate value, as the firm is not paying any taxes. However, if another firm buys the vehicle and then leases it to the firm, the firm buying the vehicle gets to deduct the depreciation allowances. With strong competition among lessors (those buying the vehicles), most of the benefits will be received by the lessee (the one using the vehicle); that is, the price at which the lessor rents the vehicle will reflect the tax breaks. This example serves to illustrate yet again an important general principle of taxation: it is often difficult to tell who really receives the benefit from a tax break. It may not be the person who enjoys the deduction.

CAPITAL GAINS   The 1997, 2001, and 2003 tax laws substantially lowered the tax rate on capital gains: for upper-income individuals, from 28 percent before 1997 to 15 percent in 2011, compared to a top tax rate on ordinary income now at 39.6 percent. In 2013, the top capital gains rate was increased to 20 percent. Consider a $1 million asset that is expected to rise in value by 10 percent (and pays no dividend or other returns). If the interest rate is 10 percent without taxation, the individual would be just indifferent to buying the asset. If the individual can borrow to buy the asset, however, the interest may be tax deductible; at a 39.6 percent marginal tax rate, the after-tax interest cost is only about 6 percent (or just over $60,000). But the capital gain is taxed at only 20 percent, so the after-tax value of the capital gain is $0.80 \times \$100,000$, or $80,000. The individual makes a pure gain of $20,000 after tax. The same reasoning makes clear that it will now pay for the individual to buy assets that, even in the absence of taxation, would entail losses. The entire profit is due to "tax arbitrage," discussed next.

SHORTING AGAINST THE BOX

Most stock brokerages permit clients to borrow shares of stock. An investor may sell borrowed shares today, and later buy the same shares on the market so that they may be returned to the lender. Selling borrowed shares is called *short selling*. If the value of the borrowed shares falls between the time they are sold and the time they are purchased again and returned, the investor may realize a profit. (Of course, if the share value rises, big losses may be racked up.) A related tax avoidance scheme, called *shorting against the box*, emerged from the favorable capital gains treatment that results when individuals leave assets to their heirs.

Shorting against the box worked like this: Suppose a wealthy individual owned 2000 shares of Microsoft that were trading at $100 per share.

The individual could raise $50,000 in cash by selling 500 shares, but if the shares were purchased at less than their current price, he or she would have to pay a capital gains tax. However, if the individual sold short 500 shares by borrowing 500 shares from a brokerage firm and selling them on the market, while at the same time promising to return 500 shares to the brokerage firm by his or her heirs after death—putting them in the brokerage's "box," as they say on Wall Street—the individual could raise the same $50,000, and avoid the capital gains tax. This way, the wealthy codger could have his or her cake and eat it too: the individual gets the cash, bears no risk, but avoids taxes as if he or she postponed selling the shares until after death.

This particular loophole was closed in 1997.

TAX ARBITRAGE   Arbitrage involves taking advantage of price differences for the same commodity. If gold is selling for $350 an ounce in New York and $375 in Zurich, and the cost of shipping gold between the two is $20, someone can buy gold in New York and ship it to Zurich for a sure profit. **Tax arbitrage** entails taking advantage of the different rates at which different kinds of income or different individuals are taxed.

Strictly speaking, the term *arbitrage* refers to situations in which there is a sure gain—that is, there is no risk assumed. Although in theory the tax code provides many opportunities for riskless tax arbitrage, in practice most tax avoidance activities involve the assumption of some risk. This is partly because of the general provision in the tax code that a set of transactions undertaken solely to avoid taxes will not be granted the favorable tax treatment. In many situations, individuals must show that they are "at risk" to obtain the favorable tax treatment, but the risks that have to be borne are minimal.

The term *arbitrage* is also applied to situations in which different individuals face different tax rates, and a set of riskless transactions are designed so both are better off as a result of the reduction in their joint tax liabilities.

Still another form of arbitrage occurs when individuals borrow to put money into a tax-exempt bond or an IRA account. If an individual borrows, ostensibly to buy a house, then the interest payments are tax deductible. The government does not ask whether the individual *needs* to borrow the funds; it only traces where the dollars actually went. It might pay someone with $100,000 in a bank account to borrow $100,000 more to buy a house, and put the extra money into tax-exempt bonds. If this individual receives, say, 4 percent interest on the tax-exempt bond, even if he or she has to pay 6 percent interest to the mortgage company (if the combined federal and state marginal tax rate is 50 percent), the individual will have an extra $1000 to spend every year: the net cost, after tax, of the mortgage interest is only 3 percent, or $3000, and the individual receives $4000 on the tax-exempt bonds.

Tax authorities have tried to limit tax arbitrages by restricting the ability to borrow to buy tax-exempt bonds or to put money into an IRA. Such restriction is difficult in practice, however, since individuals do not borrow to buy a municipal bond, but take out a larger mortgage on their house, so that they have more cash on hand to purchase a tax-exempt bond a few months later on.

TAX SHELTERS

Investment schemes that reduce one's tax liabilities are called **tax shelters**. A tax shelter exists when deductions from one income source (e.g., oil and gas or real estate) can be offset against income from another source (e.g., salaries and wages).

There are a wide variety of tax shelters, but exploration of gas and oil is perhaps the most notorious. This tax shelter is based on a number of special, favorable tax provisions for the gas and oil industries. In Chapter 21, we discussed depreciation allowances. These are provided to take account of the fact that as a machine is used, it becomes less valuable (it wears out and becomes obsolete). Similarly, as oil is extracted from a well, the well becomes less valuable. To compensate for this, the government provides *depletion allowances*. These are related not directly to the change in the value of the asset, but rather to the value of the oil extracted. The level of depletion allowances has varied over time, at one time reaching 27.5 percent of the value of the oil sold. The correspondence between

THE ECONOMICS OF TAX AVOIDANCE

For individuals with complicated tax situations, there are often gray areas, such as issues for which there is ambiguity about the proper tax treatment. A wall, for example, is part of a structure of a building, and, as such, needs to be depreciated over the life of a building (a commercial building has an assumed life of thirty-nine years). However, a *portable* wall is like equipment, and can typically be depreciated as if it were "equipment," not a building, and thus can be depreciated over ten years. How portable does a wall have to be to be "equipment" rather than "structure"?

Individuals, corporations, and their accountants engage in a risk analysis: they balance the risk of being caught taking an "aggressive" stance (that is, interpreting the tax law in the most favorable way, from their own perspective) against the consequences. Typically, as long as there are *reasonable* grounds for a taxpayer's position, no penalty is levied, even if the IRS rules against the taxpayer. Hence, apart from the costs of hassle, it pays firms to take an aggressive stance. Corporations typically have large legal staffs to handle tax issues; the "battle" with the IRS is not viewed as hassle, but simply as a part of doing business.

On the other hand, the IRS can, and does, impose penalties for actions it views as unreasonable. Accountants often face a delicate balance in trying to decide how hard to push their clients' tax interest. The boundary between the reasonable and unreasonable is often blurry.

---

the depletion allowance and the change in the value of the well is even weaker than that between the depreciation allowance for a machine and "true economic depreciation." For instance, over the life of the well, the depletion allowance may exceed the purchase price of the asset. Moreover, when an oil well (or lease) is sold, a capital loss can be taken against the original purchase price without accounting for the depletion allowances taken in the interim. It is as if the government allows two tax deductions for the decrease in the value of the asset.

The taxpayer can use these deductions to shelter other income from taxation.

WHO GAINS FROM TAX SHELTERS

Consider an oil industry tax shelter. There are five possible beneficiaries: the benefit could accrue to the "intended" beneficiary, the oil industry; it could be shifted forward to consumers, in the form of cheaper oil; shifted back to the owners of the land under which there is oil; it could accrue to the Wall Street firms that put together tax shelter deals; or completely dissipated in excessive transactions costs.

If there were no transactions costs, the theory of incidence presented in Chapter 18 would tell us that who benefits depends on elasticity of demand and supply curves. Figure 24.1A shows the case of a highly elastic supply curve. This might be more appropriate for subsidies to cattle. The special tax treatment can be thought of as lowering the cost of production, shifting the supply curve down (that is, the price required for the market to be willing to supply a given level of output is lowered). Because of the highly elastic supply curve, the market price is lowered by almost the full amount of the subsidy: consumers, not producers, benefit from the subsidy in this case. The fact that consumers are better off, however, is not an unmitigated blessing. They benefit by less than the amount of the subsidy; there is a deadweight loss. Because the supply curve is elastic, there is a large distortionary effect, and a large deadweight loss.

Figure 24.1B shows the case of a highly inelastic supply curve. This might be more appropriate to special tax provisions for oil and gas, for which the supply (in the long run) is relatively inelastic. Then, prices paid by consumers remain relatively unchanged. At first blush, it looks as if the industry is better off—by the full magnitude of the tax benefits. However, because producing oil is more attractive, oil producers compete more actively for leases of land under which there is oil. It is these owners of land—of the inelastic factors required to produce oil—who get the benefit of the subsidy. (Of course, many oil producers also are large owners of oil-producing land, and, to that extent, they benefit directly.) With an inelastic supply curve, the distortionary effect is small, but from a social point of view, the subsidy is "wasted." The usual argument for a producer subsidy is to maintain or augment the size of an industry; with inelastic supply, there is no supply effect, just a redistribution effect.

In neither case would we expect to see the returns to investment in the subsidized industry (in the long run) higher than normal. Capital flows in until its rate of return (taking into account the subsidy) is the same as elsewhere. Thus, capital in the affected industry does not receive much of the benefit. Evidence supports these theoretical predictions: the returns in heavily subsidized industries, like gas and oil, are no higher than returns elsewhere, adjusted for risk.

This analysis ignores transactions costs. Putting together and marketing tax deals costs money. If some firms have specialized talents in doing this, then they receive much of the benefits, in the form of payments for their specialized skills, which the tax law has made more valuable. In the short run, much of the benefit of tax shelters accrues to such firms, because there are always some firms that are quicker to recognize the full opportunities afforded by tax laws. In the long run, however, the "tricks" of the trade disseminate, more firms learn how to put together the tax shelters, and even the profits of the Wall Street promoters get competed away. Effectively, the value of the tax shelter is dissipated in transactions costs, including the costs of marketing the tax shelter.

MIDDLE-CLASS TAX SHELTERS

Although loopholes and tax shelters are typically thought of as provisions of the tax code that reduce tax liabilities for the rich, in fact, there are a variety of provisions in the tax code that reduce taxes for middle-income taxpayers. The most important are employer-provided health benefits (worth an estimated $163 billion in 2011), mortgage interest deductions (worth an estimated $72 billion in 2011), and the deductibility of state

and local income and property taxes (worth an estimated $64 billion in 2011). Each of these provisions results in economic distortions: for instance, mortgage deductibility results in higher expenditure on housing. It also biases individuals to buy their homes, rather than rent, and to incur debt for home purchases and home equity loans. As always, there is some ambiguity about the motivations of these provisions: Is the provision mainly a response to the efforts of the real estate lobby, to provide a subsidy that benefits them at the expense of other sectors of the economy? Is it a response to pressures from financial institutions, so they can profit from selling subprime mortgages and mortgage-backed securities despite the systemic risks this might create? Or is the provision part of a social policy that encourages home ownership, in the belief that home ownership contributes to the stability of society?

TAX REFORM AND TAX AVOIDANCE

There have been six major tax reforms since the 1980s. The first, in 1981, opened up a wide variety of tax loopholes. The market responded with enthusiasm and the tax shelter industry boomed, giving rise to a demand for tax reform. Although the 1986 tax reform represented the most serious attempt at reducing tax avoidance, the two subsequent tax reforms of 1993 and 1997 opened up new opportunities for tax avoidance at the same time as they markedly increased incentives for tax avoidance. The tax reforms of 2001 and 2003 continued the trends of the 1990s by creating even more opportunities for tax avoidance, although these reforms also decreased the incentives for tax avoidance because of significant reductions in tax rates for both ordinary and investment income.

THE 1986 TAX REFORM

One of the major objectives of the 1986 tax reform was to design a tax system that was, and appeared to be, more fair. This meant that something had to be done about tax shelters. Three approaches were considered by the Treasury Department. The first was eliminating the provisions, such as the favorable treatment of capital gains and the special treatment of the gas and oil industry, that gave rise to tax shelters. The second was limiting the extent to which losses on one category of income could be used to offset income in other categories. The third was imposing a more effective

minimum tax. After careful consideration, the Treasury, in formulating its original proposals, decided on the first approach—to go after the basic source of the problem—and clearly rejected the second. Unfortunately, when Congress took up the matter, members found it politically difficult to attack many of the shelters directly, although one change, the full taxation of capital gains, reduced the value of many tax shelters. Because Congress left many of the loopholes in place, it had to turn to the second and third methods of attacking tax shelters.

The most important way Congress did this was to divide income into three categories: ordinary (earned) income, investment income, and passive income. Income generated by tax shelters, in which the individual did not take an active role, was categorized as passive income; so was most real estate income. Losses in one category could not be used to offset income in another. Thus, interest expenses on one investment could be used to offset income from another investment, but net losses on investments as a whole could not be used to offset ordinary income. Nor could losses on real estate be used to offset ordinary income.

The Treasury had rejected this approach to controlling tax shelters for two reasons. First, one of the original objectives of tax reform had been to *simplify* the tax code. Distinguishing between passive, investment, and ordinary income requires a host of definitions, regulations, and court cases that inevitably make the tax code even more complicated. Second, there was, and is, concern that these provisions are of only limited effectiveness. Although they limit the extent to which tax loopholes can be used to avoid taxation of wage income, they do not effectively limit the extent to which individuals (particularly the rich) can avoid taxation of capital income; they increase the transactions costs of tax avoidance. Real estate projects that generate taxable income are bundled together with real estate projects that generate taxable losses. Taxes on real estate can thus continue to be avoided.

WHO GAINS AND WHO LOSES FROM TAX SHELTERS?

When producers receive favorable tax treatment, the extent to which output increases or price falls depends on elasticity of demand and supply curves. Elasticity of supply is likely to be larger in the long run than in the short run.

Short-run beneficiaries:

- Owners of factors which are specific to industry at time preferential treatment is introduced (announced)
- Those who market tax shelters

Long-run beneficiaries:

- Consumers (lower prices as industry expands)

Losers from tax shelters:

- Taxpayers
- Economy—misallocation of resources

MINIMUM TAX ON INDIVIDUALS

The 1986 tax act also attempted to reduce tax avoidance activities by imposing a somewhat stiffer minimum tax. The minimum tax on individuals is levied on a much broader definition of income; for instance, state

and local income and property taxes are not deductible, and depreciation allowances are far less generous. The rules allowing individuals to take tax shelter losses are even more stringent than under the ordinary income tax. The alternative minimum tax (AMT) rate was increased both in 1986 and 1993. Today, it stands at 26 percent for "AMT taxable income" (the taxable income defined under the alternative minimum tax rules) of $175,000 or less, for a married couple filing jointly, and a 28 percent rate on income in excess of that threshold. However, since the AMT was not indexed to inflation, many individuals who simply had been taking advantage of the ordinary deductions were caught within the AMT net. The American Taxpayer Relief Act of 2012 indexes to inflation the AMT income thresholds.

The 1986 tax law substantially reduced the demand for tax shelters and the opportunities for tax avoidance by narrowing the gap between regular rates and the minimum tax rates, by lowering the top marginal tax rates, by restricting the ability to use losses on one type of income to offset gains on another, and by taxing capital gains at full rates.

SUBSEQUENT TAX ACTS

The effects of the 1993 law were more ambiguous. The alternative minimum tax rates were increased, but so were the regular rates, with an increase in the absolute gap for upper-income individuals; and new ways of tax avoidance were introduced. The 1997 act unambiguously made matters worse: still further avenues of tax avoidance were introduced; a gap between the rates at which capital gains and ordinary income were taxed was introduced for all taxpayers, not just upper-income individuals; and the gap for upper-income individuals was increased. The net impact of the 2001 and 2003 tax acts is unclear. Although they introduced yet more opportunities for tax avoidance, the gap between capital gains and ordinary income tax rates remained the same as the top rates for both were reduced by 5 percent. The 2013 tax act increased the tax (at the top) on capital gains and ordinary income in tandem, to 20 and 39.6 percent, respectively.

EQUITY, EFFICIENCY, AND TAX REFORM

This chapter has not attempted to provide an exhaustive list of loopholes, tax avoidance devices, and tax shelters. These change rapidly; at a given moment, some of the loopholes will have been closed and others opened up. The principles involved, however, remain the same.

Different industries have very strong incentives for attempting to garner special treatment for themselves. Often there is some small justification for the special treatment. This special treatment opens up a loophole, which can usually be put into one of the categories that we have described in this chapter. It is important to remember that the benefits of these tax shelters usually do not accrue to the investor attempting to take advantage of them. In a competitive market, investors compete vigorously to take advantage of the special tax preferences that the after-tax return—which, after all, is what the individual is really concerned with—is driven down to the after-tax return on other, less advantaged investments.

The major beneficiaries are the owners of the assets in the industry at the time that the loophole is opened up. The tax advantages are capitalized in the value of their assets; that is, if they sell their assets, they will receive a higher price for them; the buyer of the asset will pay a sufficiently high price that his or her after-tax return will be the same as it would be on any other asset.

Just as the imposition of such a tax benefit causes an inequity, a windfall capital gain for the current owners, the removal of the tax benefit causes an inequity, a windfall capital loss for the current owners. If the assets in the industry are owned by the same individuals when the benefit is granted as when it is withdrawn, the two cancel each other. Frequently, however, the removal of the special treatment occurs several years later, and it is often different individuals who will be affected by the removal of the special treatment. Closing the loophole is likely to be inequitable. This makes reforms—eliminating the distorting tax preferences—all the more difficult.

However great the magnitude of the inefficiencies and inequities introduced by the various tax loopholes and tax shelters, and regardless of whether such opportunities for tax avoidance can be justified as advancing (even successfully) some important social objective, loopholes and tax shelters have one very negative consequence: they erode confidence in the tax system, because they give rise to the impression that the system is unfair, with some individuals able to avoid bearing their proper share of the tax burden. These concerns have been one of the major motivations for the tax reforms—the subject of the next chapter.

REVIEW AND PRACTICE

SUMMARY

1. Two major principles underlie most of the devices by which individuals can legally attempt to reduce their tax liabilities: tax deferral and income shifting, from high-taxed individuals and categories to lower-taxed individuals and categories (tax arbitrage).

2. Income shifting occurs under progressive taxes, under which a family, by transferring assets to children, reduces its total family tax liability.

3. Tax deferral is based on the concept that a dollar today is worth more than a dollar tomorrow, so taxes paid in the future are less costly than taxes paid today.

4. Tax loopholes have distortionary effects, and the benefits often do not accrue to those that they seem to be benefiting. The tax benefits of industry-specific loopholes (such as those relating to oil and gas) accrue to the owners of the inelastic factors in the industry (the land under which the hydrocarbon deposits lie), not to elastic factors (labor and capital).

5. The Tax Reform Act of 1986 attempted to restrict tax loopholes, not by eliminating them, but by imposing a more effective minimum tax; by dividing income into three categories (ordinary income, investment income, and passive income) and stipulating that losses attributable to one category cannot be used to offset income in another; and by taxing capital gains at the same rate as ordinary income. The 1993 and 1997 tax acts increased opportunities and demand for tax avoidance by introducing new tax shelters; by increasing the top marginal tax rate; and by opening up a gap between the rates at which ordinary income and capital gains are taxed. These effects were partially offset by an increase in the alternative minimum tax rate. The net impact of the 2001 and 2003 tax acts is unclear. Although they introduced yet more opportunties for tax avoidance, the gap between capital gains and ordinary income tax rates remained the same for high-income individuals. The 2013 act increased the tax (at the top) on capital gains and ordinary income in tandem.

6. The alternative minimum tax, which was introduced to limit the extent of tax avoidance, was not indexed to inflation. The result was that many individuals who simply had been taking advantage of the ordinary deductions were caught within its net. The American Taxpayer Relief Act of 2012 indexes to inflation the AMT income thresholds.

KEY CONCEPTS

Income shifting

Loopholes

Tax arbitrage

Tax avoidance

Tax evasion

Tax shelters

QUESTIONS AND PROBLEMS

1. A tax avoidance device that became popular in the late 1970s and early 1980s was the zero coupon bond. This was a bond that paid no interest. When the interest rate was 7 percent, a ten-year bond promising to pay $100 in 1990 would sell for $50 when issued. The government required the individual to *impute* the receipt of interest—to assume that one-tenth of the $50 gain that occurred between 1980 and 1990 occurred in each year; at the same time, the issuer of the bond could impute the payment of interest. If the two (the issuer of the bond and the purchaser) were in the same tax bracket, what would be the consequences of these imputed interest payments and receipts? If they were in different tax brackets?

2. Another popularly used tax avoidance device before 1981 was a *straddle*, in which an individual would sign one contract to buy a commodity (like wheat) at some future date and another contract to sell the same commodity at a date shortly earlier or later at the same time. Thus, when the individual had a gain on the first contract, he or she generally would have a loss on the second. What the individual gained on one, he or she lost on the other. Can you think how you could use a straddle to postpone taxes? (Hint: Consider the consequences of selling one of the securities on December 31, and the other on January 1.) Prior to 1986, long-term capital gains were taxed much more lightly than short-term capital gains. Can you think how you could use straddles to take advantage of this difference?

3. Describe the tax savings for an individual in the 28 percent marginal tax bracket who owns a business with $10,000 in "profits" if the individual incorporates, giving his or her child (over age 14) a 50 percent interest in the business. Assume the child has no other income and, as a dependent, cannot claim a personal exemption and can take only a standard deduction.

4. Citizen groups that monitor taxes paid by different corporations often complain about the low average tax rates that some corporations, especially those engaged in extensive leasing, pay. Is this allegation "fair"? Who may really benefit from such leasing?

5. Some have argued that leasing agreements among firms with different tax situations may enhance economic efficiency. Explain.

6. Who benefits from the fact that state and municipal bonds are tax-exempt—the buyer of the bond, or the municipality that issues them?

    a. Assume that there are so many individuals in the top income bracket that all the bonds are purchased by them. What is your answer then?

    b. Assume that the government allows wealthy individuals to borrow to buy tax-exempt bonds. How does that affect your answer as to who benefits?

7. Explain how an individual can engage in tax arbitrage by borrowing: (a) to put money into an IRA; or (b) to buy a tax-exempt bond.

8. Insurance policies often are a combination of a savings program and life insurance. The individual pays the company, say, $1000 a year; $100 of that goes to cover the risk of his or her dying during the year, and the remainder goes into a savings program. The return on the amount in the savings program accumulates free of tax—just like an IRA. Explain how insurance can be used as a tax avoidance device.

25

REFORM OF THE TAX SYSTEM

The past three decades have witnessed a succession of tax reforms. Each reform, promising a new era, in part undid the excesses of the previous reform. One reform reduced progressivity; the next increased it. One reform provided more investment tax incentives; the next tried to "level the playing field"; the next tried to tilt it again in a slightly different direction. Each reform was introduced with grand rhetoric. For instance, President Reagan, in transmitting his tax reform to Congress on May 29, 1985, wrote:

> We face an historic challenge: to change our present tax system into a model of fairness, simplicity, efficiency, and compassion, to remove the obstacle to growth and unlock the door to a future of unparalleled innovation and achievement.
>
> For too long our tax code has been a source of ridicule and resentment, violating our Nation's most fundamental principles of justice and fair play. While most Americans labor under excessively high tax rates that discourage work and cut drastically into savings, many are able to exploit the tangled mess of loopholes that has grown up around our tax code to avoid paying their fair share—and sometimes paying any taxes at all.

But by the time each reform had wended its way through Congress, it was but a pale shadow of the original.

Tax reforms have undergone enormous vicissitudes. In 1981, a variety of special provisions were introduced to encourage particular investments; the reforms of 1986 stripped these away, as well as a number of other special provisions that had accumulated over the years. Then, in 1993, and even more so in 1997, a number of new special provisions were introduced. This trend continued with the tax reforms of 2001 and 2003, accompanied by a significant reduction in tax rates for both ordinary and investment income especially. In between—for instance, 1982 and 1990—there were other tax changes. Although the tax acts each had more or less grandiose names—the Economic Recovery Tax Act of 1981 (ERTA), the Tax Reform Act of 1986 (TRA86), the Omnibus Budget Reconciliation Act of 1993 (OBRA93), the Taxpayer Relief Act of 1997 (TRA97), the Economic Growth and Tax Relief Reconciliation Act of 2001 (EGTRRA), and the Jobs and Growth Tax Relief Reconciliation Act of 2003 (JGTRRA), and the American Taxpayer Relief Act of 2012 (ATRA)—we shall refer to these simply as the 1981, 1986, 1993, 1997, 2001, 2003, and 2012 tax acts or reforms.

The rhetoric behind each tax reform—even those that went in opposite directions—was much the same: each promised a fairer, simpler tax code that would promote economic growth and efficiency. However, there was debate about what "fairer" meant, and what changes would most effectively promote economic growth or increase equity. The issues are intertwined: the complexities of the tax code provide scope for tax avoidance, which is viewed to be unfair and which introduces distortions into the economy.

Inevitably, debates about tax reform reflect broader concerns in society: given the large increase in inequality in the United States, especially since 1980 (with the share of the top 1 percent doubling, the share of the top 0.1 percent increasing three- to fourfold), it is not surprising that there is an increased focus on equity and on the provisions of the tax code that have facilitated that; given the threat of climate change, it is not surprising that there is an increased focus on environmental taxes, designed to curb greenhouse gas emissions; and given the role that the financial sector played in causing the Great Recession of 2008—the worst economic downturn since the Great Depression—it is not surprising that there is an increased focus on taxes to that sector.

Advances in technology have provided another motivation for tax reform: with these advances, it is possible to lower costs of administration and compliance. Taxes that might have been difficult to implement in the past may now be more feasible.

In this chapter, we first review the major themes of tax reforms over the past three decades, and assess each of the major tax reforms in terms of those themes. The second part of the chapter discusses likely themes of tax reforms in the twenty-first century.

FOCUS QUESTIONS

1. What have been the major impetuses for tax reforms during the past three decades?

2. How successful was the 1986 Tax Reform Act in lowering marginal tax rates, providing a level investment playing field, closing loopholes, and simplifying the tax code?

3. How did tax changes from 1981 to 2012 alter the degree of progressivity of the tax code?

4. What are some of the basic trade-offs in the design of tax reform?

5. What are likely to be the major directions of tax reform in coming decades?

FAIRNESS

In Chapter 17, we noted that fairness, like beauty, is in the eye of the beholder: most taxpayers feel that they pay more than their fair share of taxes; few taxpayers feel that they pay too little.

The fairness debate has focused around both issues highlighted in Chapter 17: Do individuals in similar economic circumstances (similar ability to pay) pay markedly different taxes—the issue of horizontal equity? And do richer individuals pay their "fair share" or do they manage to escape taxation by taking advantage of loopholes—the issue of vertical equity?

HORIZONTAL EQUITY ISSUES

A fair tax system imposes similar taxes on those in similar economic circumstances. Critics of the current tax system argue that it is unfair, because it both makes distinctions that it should not make, and occasionally does not make distinctions that it should.

In Chapter 22, we noted the variety of deductions and credits—for instance, for mortgage interest, state and local income taxes, and college tuition. As a result, homeowners may be favored over renters[1]; residents of states with income taxes are favored over residents of states that use sales taxes as the main instrument for raising revenues; and taxpayers with children going to college are favored over those without children going to college. *Base-broadening reforms seek to eliminate these special provisions, and, in doing so, to increase the fairness (and efficiency) of the tax system.*

However, there are some deep philosophical questions: Which distinctions *should* be made? An individual facing large medical expenses— say, for heart surgery—may not have the same capacity to pay as someone with a similar income not facing such costs. Many of the debates about tax policy center around these questions.

HOUSEHOLD STRUCTURE   There is, for instance, an active debate about whether the tax system is fair in its treatment of those with nontraditional household structures. These concerns have grown as the diversity in household structures in the United States has increased. Under the current tax system, as we noted in Chapter 17, when two people with similar

---

[1] As we noted in Chapter 21, the magnitude of the advantage may not be as great as appears at first blush, because the tax advantages of accelerated depreciation lead to lower rents in competitive market.

incomes marry, they no longer face a "marriage penalty"—that is, their taxes do not rise simply as a result of their getting married. However, when two people with dissimilar incomes marry, they may still derive a "marriage bonus"—that is, their taxes fall simply as a result of their getting married. Also, there are deductions and credits for children. Is it fair, some ask, to give favorable treatment to a household who chooses to have, or is blessed with, children, while imposing higher tax rates on those who choose not to, or cannot, have children? Critics of these deductions and credits view a decision to have a child as a consumption decision, only slightly different from a decision to buy a car or a home, and argue that the government should not favor one form of consumption over another. Others, however, argue that the cost of an additional child is far greater than the tax deductions and credits, so those with more children are less able to pay, and the tax code does not adequately reflect this difference in ability to pay.

There is also a debate about whether the tax code is fair to families in which both parents work. Although the child care credit recognizes the additional expenses arising when both parents work, the credit goes only part of the way in reflecting the additional costs.

These issues—particularly the marriage tax or marriage subsidy—have drawn increasing attention from tax reformers, especially to the extent that they suggest conflicts between the tax code and basic American values. (Is the tax code anti-family? Anti-children? Against women taking an active role in the labor market?)

FAIRNESS IN CAPITAL TAXATION   Some view any capital taxation as unfair. Two individuals with similar lifetime budget constraints will face different lifetime taxes, with the individual who chooses to consume more of his or her income later in life paying higher taxes than the one who chooses a more profligate lifestyle in his or her younger years.

Moreover, much of the return to capital simply reflects inflation; it is not a real return. Why should someone who sees the real value of his or her wealth decreasing pay a tax, simply because the nominal value has increased?

In addition, the failure to integrate the corporate and personal income taxes means that income in the corporate sector may be "double taxed," once within the corporation and a second time when the profits are distributed to the household sector.

On the other hand, the fact that a number of special provisions allow much of capital income to escape taxation is viewed by some as unfair. Critics claim that the tax system is unfair both in the way it discriminates between owners of different kinds of assets and in the way it discriminates between capital income and wage income. For instance, capital gains are taxed only upon realization, and, as we have noted, the value of this postponement may exceed the value of the taxes on the illusory increase in

values as a result of inflation. Then, too, because of the "step-up in basis at death" (see Chapter 21), a considerable fraction of capital gains totally escapes taxation.

Not surprisingly, given the disparity in views concerning the fairness of the current system of taxation of capital income, there is a disparity of views concerning desirable reforms. Some argue for eliminating all taxes on capital income (those who see all taxation of capital income as unfair); others argue for closing "capital income loopholes," especially the favorable treatment of capital gains, including the step-up of basis.

VERTICAL EQUITY

Even if it is accepted that the rich should pay a larger fraction of their income in taxes, the question remains, how much larger? As Figure 25.1 shows, historically, the rich have indeed paid a much higher percentage

FIGURE 25.1

COMPARISON OF FEDERAL INCOME TAXES PAID

Historically, families in the upper range of the income distribution have paid a substantially larger percentage of income taxes compared to all families, whether in terms of their effective tax rate or their share of tax liabilities.

SOURCES: Congressional Budget Office, *Historical Effective Federal Tax Rates: 1979 to 2007* and *Shares of Federal Tax Liabilities for All Households, by Comprehensive Household Income Quintile, 1979–2007*; www.cbo.gov.
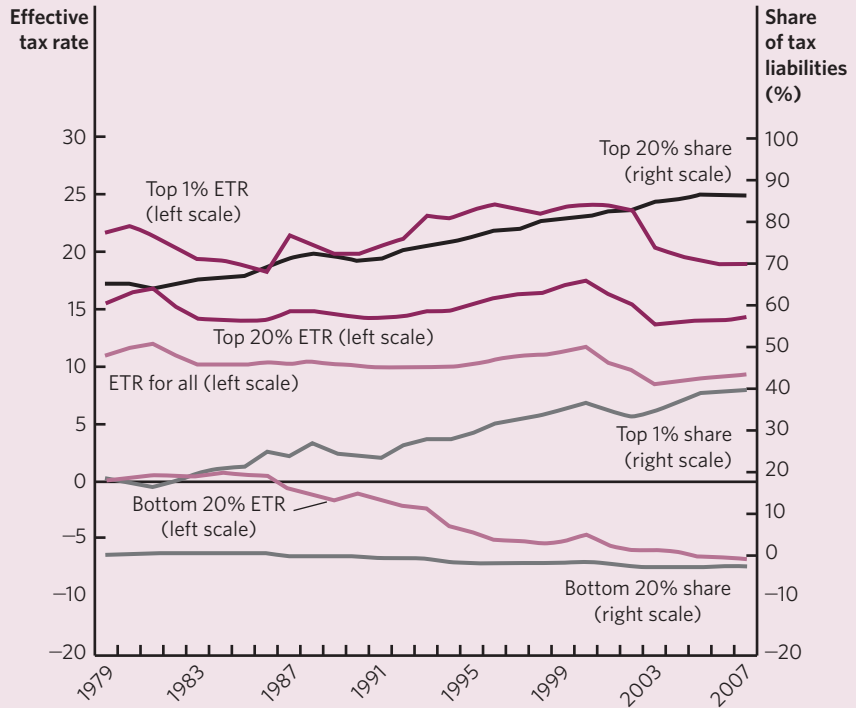
Note: Quintiles based on comprehensive household income.

Share of gross income (%)

TRENDS IN INCOME DISTRIBUTION
The share of gross income going to the richest Americans has risen dramatically since 1980.

Top 10%

Top 5%

Top 1%
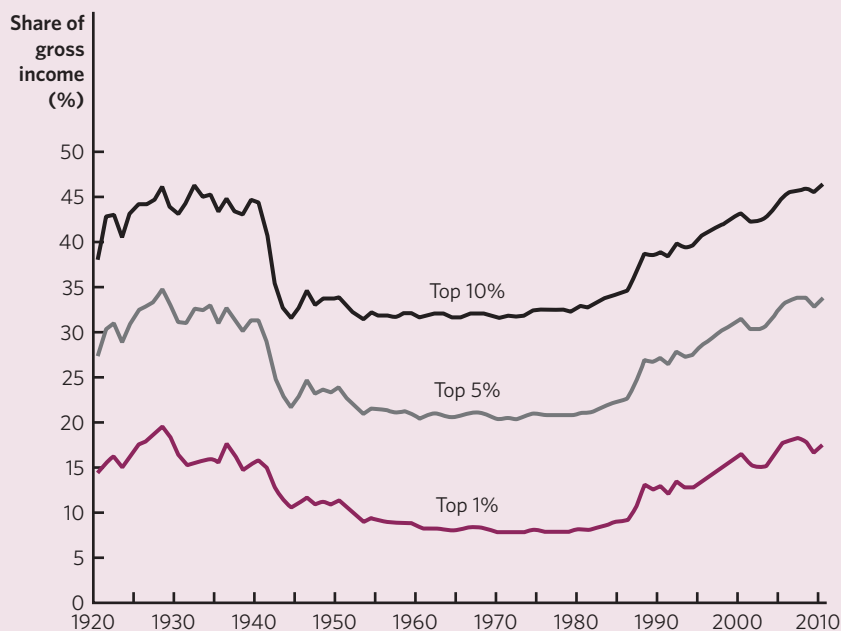
Note: Taxpayers ranked by gross income (excluding capital gains and government transfers).

than those with average income. However, some feel the rich still do not pay their fair share of federal income taxes when compared with their share of total income, as depicted by trends in gross income distribution over the past century (see Figure 25.2).

The past two decades have seen marked increases in inequality, with median income essentially stagnating, while those at the top have done particularly well. From 1993 to 2011, the average income of the top 1 percent (families with income above $367,000 in 2011) increased 57.5 percent, while the average income of the bottom 99 percent grew by just 5.8 percent (both figures are adjusted for inflation). In other words, the top 1 percent captured 62 percent of the overall economic growth of real incomes per family over the period 1993–2011. The share of total (reported) income, including capital gains, of the top 1 percent rose from 14.2 percent in 1993 to 23.5 percent in 2007, and fell to "only" 20 percent in the aftermath of the Great Recession.[2] These growing inequities have led many to argue that those at the top have an increased ability to pay

[2] E. Saez, *Striking It Richer: The Evolution of Top Incomes in the United States (Updated with 2011 Estimates)*, January 23, 2013; monograph update of an article with the same title that originally appeared in Stanford Center for the Study of Poverty and Inequality, *Pathways Magazine* (Winter 2008): 6–7.

relative to those in the middle, and that, accordingly, there should be more progressivity.

Perhaps more disturbing to the legitimacy of the tax system are the large numbers of those at the very top who manage to avoid paying taxes by taking advantage of certain provisions of the tax code, such as moving money offshore to tax shelters like the Cayman Islands or figuring out how to claim that their income should be subjected to the much lower tax on capital gains. These concerns were crystallized in the 2012 election, in which one of the candidates had a very high income (in excess of $20 million per year), but managed to keep his tax bill down to 14 percent of reported income.

Although the 1986 tax act was intended to close loopholes and lower rates, it was not intended to change the overall degree of progressivity, which had been markedly reduced in 1981. A major objective of the 1993 tax act was to increase progressivity in a limited and partial way, by increasing the tax rates on the upper 2 percent of families. Some of the increased progressivity of 1993 was reversed in 1997 with the reduction in capital gains tax rates, most of the benefits of which accrue to the wealthy. Progressivity was further decreased in 2001 and 2003, not only by a reduction in the top marginal tax rates (from 39.6, 36, 31, and 28 percent to 35, 33, 28, and 25 percent, respectively), but also by a further reduction in capital gains tax rates and the extension of these preferential rates to dividends. This trend was reversed and progressivity enhanced in 2012 when the top marginal rates were increased to 39.6 and 20 percent for ordinary income and capital gains/dividends, respectively.

Similarly, whereas there is consensus that the poor should pay a smaller fraction of their income in taxes than the average family, the question is, how much smaller? One view is that those with incomes below the poverty line—the minimum income required to attain a basic standard of living—should pay no taxes. This was perhaps the original intent of the personal exemptions, but because they were not indexed until 1985, for long periods of our nation's history, we have, in fact, imposed taxes on those below the poverty threshold. The 1986 tax reform removed most of those below the poverty line from the tax rolls and, in doing so, saved greatly on administrative costs. Today, the tax threshold is more than double the poverty threshold (see Figure 25.3).

The 1993 tax act went one step further. It not only reduced taxes on the poor, but it also actually increased subsidies available to the working poor with children who, often in spite of full-time work, remained below the poverty level. As a result, the goal of ensuring that all families with one full-time earner would be out of poverty was almost attained.

2010$

60,000

50,000

40,000

Tax threshold, 2010$

30,000

Poverty threshold, 2010$

20,000

10,000

0

1980 1990 2000 2005 2010
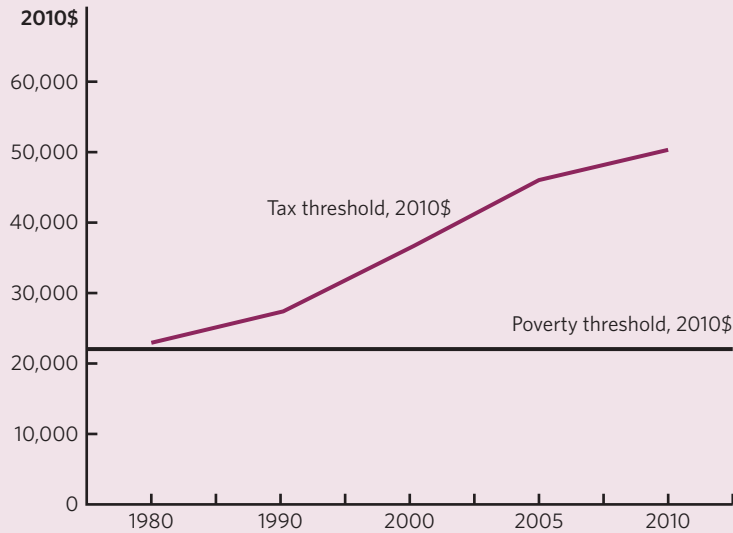
Note: Assumes all income is wages, children qualify for EITC and CTC, and there are no child care expenses.

EFFICIENCY

Issues of equity are always contentious: as Senator Russell Long was fond of saying, "A tax loophole is something that benefits the other guy. If it benefits you, it is tax reform." There is, however, more consensus on what is meant by efficiency. Thus, the Reagan administration did not argue for lowering marginal tax rates on the rich on grounds of equity, that they were paying too much. Such an argument probably would have fallen on deaf ears. Rather, it based its contention on grounds of *efficiency:* that the high marginal tax rates were causing large economic distortions, and that the country as a whole was suffering from the reduced labor supply and savings. Indeed, advocates of the 1981 and 1986 reductions in top marginal tax rates went so far as to argue that the supply effects resulting from the lowering of tax rates, combined

with reduced incentives to avoid and evade taxes, would be so large as to *increase* tax revenues.

The point that tax rates can be so high that reductions in tax rates can increase tax revenues was popularized by economist Arthur Laffer. The so-called Laffer curve is depicted in Figure 25.4. If tax rates were set at 100 percent, clearly individuals would have no incentive to work, so tax revenues would be zero. The question of how high tax rates have to be before tax revenues start to decline is an empirical one, about which economists disagree. There is a consensus that, overall, the lowering of tax rates in 1981 did what most economists outside the Reagan administration predicted it would—it lowered tax revenues; and the lowering of marginal tax rates in 1986 had at most a negligible effect on savings (which remained abysmally low) and on labor supply. Lowering of taxes on the return to capital (dividends or capital gains) in 2001 and 2003 were followed by still further decreases in savings.

As the United States entered a period of high unemployment after the financial crisis of 2008, many conservatives argued that it would be disastrous to raise taxes, even on the rich—although eventually a political compromise was reached in which the Bush tax cuts for the very rich (families with an income over $450,000) were repealed, raising top margin income tax rates back to 39.6 percent (from 35 percent), with top rates for dividends and capital gains raised to 20 percent (from 15 percent).

The debate has been plagued by a number of confusions. First, although repealing the tax cuts would dampen aggregate demand (and insufficient aggregate demand was the cause of the high unemployment), if the revenues raised were then spent, total output and employment would have

FIGURE 25.4

THE LAFFER CURVE
Raising tax rates beyond some level may reduce incentives enough to reduce output and tax revenues. There is, then, a tax rate at which tax revenues are maximized.
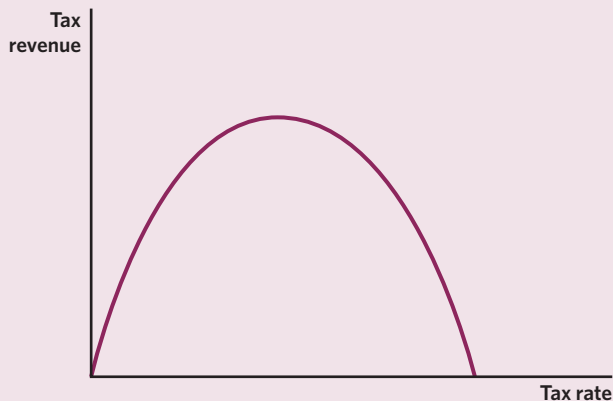
increased. (This is called the *balanced budget multiplier*. Because savings rates of those at the top are very high, the multiplier may be quite large—well above 1.[3]) Moreover, if the additional spending was devoted to high-productivity investments, then future growth would be enhanced, and realization of this might induce individuals *today* to consume more, further stimulating current aggregate demand.

The advocates of keeping taxes on the top income earners low focused on the effects on their labor supply or savings—*supply* side effects. Whether these supply side effects are significant is a matter of some contention, but even if they were important, increasing supply when the constraint is inadequate demand will not lead to more employment.

Even in more normal times, however, the link between employment and tax policy is often confused in the political debate. Most economists believe that, in the long run, monetary policy, if appropriately managed, can ensure full employment. Looser monetary policy generates more investment and more jobs; the Federal Reserve Board, which is in charge of monetary policy, seeks to keep unemployment as low as it can, without tightness in the labor market giving rise to inflation pressures. Thus, in normal times, tax policy has little to do with jobs and job creation.

Tax policy can affect the rate of growth and overall level of national income in several ways. It can encourage greater labor force participation, and can encourage individuals to work longer hours. Doing this will increase the level of national income, and, in the short run, the rate of growth, while participation rates and hours increase. It can encourage greater savings and investment and, through this, greater productivity. Again, in the long run, this will increase the level of national income, and in the short run, while productivity is increasing, it will increase the rate of growth. Finally, tax policy may be able to affect investments in research and development, and the *rate* of productivity increase, and in doing so, it can affect the rate of growth over the long run. While all off these effects may occur, it has been hard to establish their quantitative significance.

There is evidence that tax revenues from the very rich did increase after the 1981 and 1986 tax reforms, but there is disagreement about the reason.[4] Some believe that the rich recognized—correctly, as it turned out—that these low tax rates were too good to be permanent, and took

---

[3] Some economists have expressed skepticism about the size of the balanced budget multiplier—the extent to which total GDP increases when the government simultaneously raises taxes and expenditures. Much of the empirical results are based on periods when the economy is close to full employment, when, by definition, there is little scope to increase output. In the period following the financial crisis, with vast underutilized resources, there was considerable scope for increased output.

[4] See, for instance, B. Bosworth and G. Burtless, "Effects of Tax Reform on Labor Supply, Investment, and Saving," *Journal of Economic Perspectives* (Winter 1992): 3–25; L. Lindsey, *The Growth Experiment: How the New Tax Policy Is Transforming the U.S. Economy* (New York: Basic Books, 1990); and "Individual Taxpayer Response to Tax Cuts: 1982–1984, with Implications for Revenue Maximizing Tax Rate," *Journal of Public Economics* 33 (1987): 173–206.

MARGINAL TAX RATES AND THE 1986 TAX REFORM

The 1986 tax reform was heralded as a major structural reform in our tax system—a "second American Revolution," to use President Reagan's rhetoric when he introduced it. In fact, the changes in marginal tax rates were far more modest, and some taxpayers (4 percent) actually experienced a marginal tax rate increase of more than 10 percent.* Only 11.3 percent experienced a marginal tax rate decrease of more than 10 percent (see figure on the right). Much of the reduction in marginal tax rates simply reversed the bracket creep resulting from inflation that had occurred during the 1970s. Indeed, for a married couple with two children with an income of $40,000 in 1985 real dollars, the marginal tax rate in 1988 of 28 percent was the same as in 1985, and still considerably above that in 1970 and 1960.

*For instance, the 1986 tax reform abolished the provision by which 10 percent of earnings of working spouses was deductible (up to $3000).
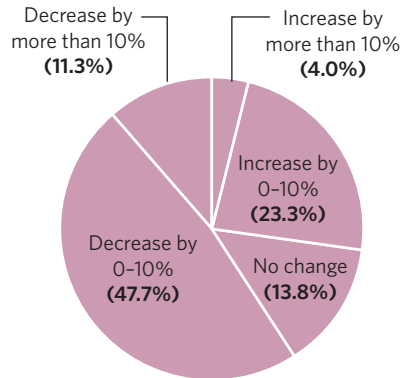
MARGINAL TAX RATE CHANGES FROM THE 1986 TAX REFORM Only a quarter of the population had a marginal tax rate change of more than 10 percent.
SOURCE: J. Hausman and J. Poterba, "Household Behavior and the Tax Reform Act of 1986," *Journal of Economic Perspectives* (Summer 1987): 101–120.

advantage of the lower rates to realize, for instance, capital gains. Thus, some of the increased tax revenues were at the expense of tax revenues that would have been received later. Some believe that the increased reported income of the rich was the result not of increased incentives provided by the tax bill, but of reduced opportunities for tax avoidance, the closing of loopholes combined with the increased alternative minimum tax. Some argue that it was not just the opportunities for tax avoidance that were decreased in 1986, but the incentives as well (see Chapter 24).

The most important reason for the increased income of the rich, though, had to do with underlying social and economic forces: a continuation of the movement toward greater inequality that has characterized

the economy since 1973, and especially since 1980 (see Figure 25.2). The rich were getting richer. Loopholes in the tax code that enabled many of the rich to avoid much of the taxes that they would otherwise have paid may, of course, have reinforced the country's growing inequality.

The fact that tax revenues failed to increase in 1981 reinforced skepticism about the relevance of the Laffer curve. In recent discussions of tax reforms, few claim that lowering tax rates will lead to increased taxed revenues. The effects of the 1993 tax act provide some insight into these issues. Taxes on the very rich increased substantially, yet their incomes and their tax payments continued to rise, at a pace even faster than that of the economy as a whole. Whereas in the 1970s, the personal savings rate fluctuated between 7 and 9 percent, and it decreased through the 1980s, to 4.8 percent in 1989. It hit a low of 3.8 percent in 1994, but later rebounded to 4.5 to 5 percent, only to fall again, to near zero, in the years before the global financial crisis. Bush's tax cuts—not just on marginal tax rates, but also on dividends and capital gains—which was supposed to induce more savings, clearly failed to do so. The tax changes seem to have had little effect on these trends.[5]

As inequality in the United States has grown and its economic performance (reflected in rates of growth) has deteriorated, some have suggested that low tax rates at the top (half of what they were before Reagan's tax reform) may actually be part of both problems. Much of the income at the top today is considered by some as rent seeking—activities directed not at increasing the size of the nation's income, but at getting a larger share of that income. Financial sector activities, such as predatory lending, seek to take advantage of the poor. Corporate CEOs are seen as taking an outsized share of corporate revenues, with little link to the performance of their companies. With lower taxes at the top, the returns to such rent-seeking activities has increased. Studies looking across countries at the effect of increases at top marginal tax rates have shown that such increases do not lead to reduced growth—a result consistent with the hypothesis that any adverse effects associated with reduced incentives to work and save are offset by positive effects associated with reduced rent seeking.

BASE BROADENING

The principle that lower marginal tax rates reduce deadweight losses and incentives for tax avoidance is one of the fundamentals of tax reforms aimed at increasing efficiency.

---

[5] The 1986 tax reform has been extensively studied to ascertain its effects. See, for instance, H. J. Aaron, "Lessons for Tax Reform," in *Do Taxes Matter: The Impact of the Tax Reform Act of 1986*, ed. J. Slemrod (Cambridge, MA: MIT Press, 1990), pp. 321–331.

There are only two ways to lower marginal tax rates without revenue losses: by reducing the degree of progressivity, or by broadening the tax base. The former entails a trade-off: one can obtain greater efficiency only at the expense of less (vertical) equity.

Base broadening holds out the prospect of increasing both efficiency and, especially, horizontal equity. Special tax provisions that lower or exempt taxes on particular categories of income distort the economy and benefit some groups (those that have the favored forms of income) over others. The larger the tax base, the smaller the tax rate required to raise a given revenue. Thus, base broadening has been another principal tenet of tax reform.

Unfortunately, the principle of base broadening often conflicts with other objectives of tax policy—in particular, with using tax policy to promote other desirable economic and social goals, including correcting market failures and helping particular groups that may be in need. For example, the objective of promoting savings in general, and of encouraging individuals to put aside money for their retirement in particular, has led to special tax treatment of retirement savings. The objective of encouraging home ownership has led to favorable treatment of owner-occupied housing. However, critics of these provisions rightly point out that there are often less expensive ways of obtaining these social and economic objectives than through the tax code.

Of the tax reforms of recent decades, only the reform of 1986 made a serious attempt at base broadening. Other tax reforms actually narrowed the tax base as they attempted to pursue other objectives. This was most notable in the 1981, 1993, 1997, and 2001 tax acts. The 1981 act granted very favorable depreciation rates and investment tax credits to plant and equipment, in an unsuccessful attempt to encourage "smokestack America." The 1993 tax bill similarly provided special treatment for investments in new small businesses. The 1997 tax act introduced still further special provisions to encourage education and savings. The 2001 tax act raised pretax contribution limits for retirement plans substantially and phased in repeal of the estate tax. Most significantly, the 1997, 2001, and 2003 bills lowered taxes on capital gains and dividends.

Whether these special provisions actually increase overall economic efficiency, or whether they increase distortions, remains a subject of debate. If markets work well, providing preferential treatment simply decreases economic efficiency. The preferential treatments granted in 1981 were not based on any theory of market failures. The excess investments that they helped lead to in commercial real estate certainly corroborate the view that they distorted investment patterns. The 1986 tax reform took up as one of its mottoes "leveling the playing field," undoing the distortions introduced into the tax code a scant five years earlier. The 1993 act attempted to focus attention on what the economists in the Clinton administration saw as a market failure, the difficulty small

businesses have obtaining capital. The 1997 act was based on the notion that in today's world, people need a higher level of education: fourteen years of schooling rather than twelve, needs to become the norm. Even if individuals could borrow, the standard loan contract imposed enormous risks; imperfections in risk markets meant that individuals could not divest themselves of these risks. However, critics argued that the tax proposals were possibly counterproductive and reforms should have focused on improving the loan program and extending finance to the poor. Australia, for example, has a very successful student loan program called *income-contingent loans*, in which the amounts repaid depend on the income the individual earns later in life. The group in the population for whom financial resources are the largest barrier to access to higher education are the very poor, but tax reforms—aimed at middle-class parents—do nothing for them; critics claimed that the money could have been better spent on expanding Pell grants (the government program for this group). Finally, some worried that the financial assistance provided through the tax system to middle-class students would only encourage colleges to increase tuition; the higher tuition would actually reduce affordability to poor students, for whom the tax breaks meant little.

The 1986 tax reform, for all its rhetoric, did not get very far in base broadening. Table 25.1 looks at three categories of potential base-broadening. Even though the deductibility of state and local sales

TABLE 25.1 TAX EXPENDITURES RETAINED AND DROPPED BY THE 1986 TAX REFORM ACT (BILLIONS 2005 $)

TAX EXPENDITURE	1986 ACT	1986 BUDGET	2011 BUDGET
State and local tax deductions			
Sales tax	Dropped	−9.1	na
Income and property tax	Retained	46.2	na
Sales and income tax			35.7
Property tax			20.1
Interest deductions			
For consumer credit	Dropped	−29.2	na
For mortgage interest	Retained	44.7	62.6
Health care provisions			
Medical expense deduction	Modified	6.3	7.2
Exclusion of employer contibutions for medical insurance premiums and medical care	Retained	39.0	141.4

SOURCE (for update and conversions): OMB Budget FY 2013, *Analytical Perspectives*, Table 17-2, and *Historical Tables*, Table 10-1.

taxes was removed, the far more important provision of deductibility of income and property taxes was retained; the reform simply encouraged states to shift to these forms of taxation. Deductibility of interest on consumer credit was eliminated, but the more important deductibility of mortgage payments was kept; individuals were simply encouraged to take home equity loans (loans in which the equity in their house is used as collateral). The medical expense deduction was modified (only expenses in excess of 7.5 percent of income were deductible), but the far more important exclusion of employer contributions for medical insurance premiums and medical care was retained.

Two sets of distortions are associated with the current income tax. The first is that the special provisions encourage certain kinds of activities at the expense of others, and although *some* of the special provisions can be justified in terms of social or economic objectives, many are simply the result of special interests. Although there may be a social objective in home ownership, there is no rationale for encouraging wealthy Americans to buy bigger homes, or to encourage debt finance (as a result of mortgage deductability). Although it may be desirable to encourage some forms of investment, it makes little sense to encourage financial or land speculation (as the special treatment of capital gains does), and there is no good rationale for earnings of those who work in certain Wall Street firms (called private equity firms) to be taxed at lower rates than those who work in other sectors.

The more general distortions associated with the income tax are largely related to marginal tax rates (see Table 25.2), but the magnitude of these effects (given current tax rates) remains a subject of controversy. We have noted that lowering the tax on capital gains and dividends or the top marginal tax rate more generally has *not* resulted in more savings, and there is similarly little evidence of large labor supply elasticities for most workers.

One area of concern is that increased progressivity (combined with other tax changes) has increased the marginal tax rates facing secondary-worker wage earners (especially in upper–middle-income families). Again, with secondary-worker labor force participation rates reaching an all-time high

TABLE 25.2 MARGINAL TAX RATES OVER FIVE DECADES

FILING STATUS	ADJUSTED GROSS INCOME (2011$)	MARGINAL TAX RATE (PERCENT)								
		1960	1970	1980	1985	1988	1994	1998	2004	2011
Single	$30,000	22	22	24	20	15	15	15	15	15
Married filing jointly	$50,000	22	22	24	22	15	15	15	15	15
Married filing jointly	$90,000	26	25	37	33	33	28	28	25	25
Married filing jointly	$385,000	59	55	64	50	28	39.6	39.6	35	35

SOURCE: Tax Foundation, *U.S. Federal Individual Income Tax Rates History, 1913–2011.*

before the crisis, any adverse effects appear to be limited. Critics say that participation rates increased because of declining real wages, especially at the bottom of the income distribution (a backward-bending supply curve), but participation rates have increased throughout the economy (or did until the onset of the Great Recession), even in upper-income households. These increased participation rates clearly reflect social trends (attitudes toward women working and more households headed by single women). The debate is over whether participation rates would have been even higher with lower tax rates. The marked decline in labor force participation since 2008 seems largely a demand-side phenomenon. With fewer jobs available, many have dropped out of the labor force.

INTERACTION OF FAIRNESS AND EFFICIENCY CONCERNS

Many tax reform discussions entail trade-offs between equity and efficiency, different perceptions of fairness, and different judgments about the magnitude of the distortions generated by taxes. Not surprisingly, economists who focus on closing loopholes, increasing progressivity, and eliminating the preferential treatment of capital gains, worry more about inequality, the distortions associated with *differential* taxation, and the ability of those at the top to escape paying their "fair share." More conservative economists believe that there are greater distortions associated with marginal tax rates, and so are less concerned about equity or the distortions that arise from differential taxation (of, say, capital gains and ordinary income). They are also more likely to believe that if the economy grows faster, all will benefit—an idea that is called *trickle-down economics*, for which there is scant evidence. Indeed, in recent decades, median incomes have stagnated even though incomes at the top have done very well.

Consider the discussion of lowering taxes on dividends in the early years of the Bush administration. Advocates argued that there was "double taxation"—with money being taxed when it was earned by the corporation and again as it was distributed—and thus unfair. Moreover, they argued that the tax discouraged savings and investment, so lowering the tax rate would promote growth. Critics pointed out that because so many corporations had figured out how to avoid the corporate income tax, the problem of double taxation was not that serious. The real problem was "zero taxation"—that those with capital income had figured out how to avoid paying their fair share of taxes. There was an obvious way of at least partially reconciling these views: only allowing preferential treatment of

dividends for those corporations who had in fact paid taxes. Though there was a bipartisan agreement that this should be done, at the last minute, this particular provision was dropped; so now, dividends, even from corporations that totally avoid taxes, still receive preferential treatment. The efficiency effects, too, did not play out as the advocates had hoped: savings did not increase, but fell.

SIMPLIFYING THE TAX CODE

A third major impetus for tax reform has simply been that the tax code has grown too complicated. Complexity increases administrative costs for the government and compliance costs for the taxpayer; it creates uncertainty in calculating tax liabilities, even with the help of tax professionals; it reduces revenue by creating opportunities for tax avoidance, particularly for wealthy taxpayers; and it decreases transparency and increases perceptions of inequities, undermining incentives for voluntary compliance.

ASSESSING COMPLEXITY

There are many indicators of the complexity of the tax system. The IRS has more than 600 different forms. Of the 231 volumes comprising the Code of Federal Regulations (describing all general and permanent laws in force in the United States), the Internal Revenue Code is responsible for twenty. Of these, fifteen volumes, filling 11,586 pages, and weighing more than twenty-eight pounds are devoted to the individual and corporate income and payroll taxes. Not only is the tax code exceedingly long, running to almost 4 million words, but it is also in eternal flux: the tax code has been changed at a rate of more than once a day over the past decade. The tax system is so complex that 60 percent of individual taxpayers have resorted to using tax preparers, and even these tax professionals often cannot agree on interpretations of tax laws and regulations.[6]

The complexity is reflected in the difficulties that taxpayers, tax preparers, and even the IRS have in accurately assessing tax liabilities, undermining the foundation of voluntary compliance on which the tax system is built. Taxpayers have improved their accuracy by taking advantage of the now widely available low-cost, off-the-shelf software to assist them in

---

[6] The National Taxpayer Advocate, "The Most Serious Problems Encountered by Taxpayers," *2008 Annual Report to Congress.*

filing their returns—more than half the returns filed in 2011 were filed electronically. Although tax software reduces the difficulty of filling out forms and making calculations, it does not reduce the burden of record keeping. Given taxpayer reliance on tax preparers, in 2010, the IRS launched a comprehensive, multiyear initiative that will ultimately result in registering all paid return preparers with the IRS, and requiring these paid return preparers to be identified on the returns they prepare, pass a compentency exam, and complete annual continuing education on tax law and professional conduct. The IRS has tried to address its own inaccuracy and inconsistency in responding to taxpayer queries by centralizing its call center operations to facilitate greater specialization in fielding taxpayer queries.[7]

INCREASING COMPLIANCE

To collect its taxes, the federal government relies on a combination of voluntary compliance coupled with the threat of stiff fines and prosecution for outright fraud. To assist individuals whose sense of moral responsibility might be too weak to induce them to report all their income, the government requires employers to report the wages they pay to their workers and firms to report the dividends and interest they pay to shareholders and bondholders. The government has only limited facilities for checking on cash transactions, however, and the ability to avoid taxes by using cash has encouraged the growth of unreported transactions, referred to as the **underground economy**. Although precise estimates of the size of the underground economy are hard to come by, some observers believe that it may involve up to one-quarter of the work force and 15 percent of GNP. It includes not only criminals, but also some waiters, babysitters, domestic help, carpenters, gardeners, and others involved primarily in the service sector.[8] The IRS estimates that the net misreporting percentage (NMP) for income subject to substantial information reporting and withholding is only 1 percent, whereas the NMP for income with little or no information reporting is 56 percent.

The estimated individual income tax owed on unreported income, together with underreported and underpaid income resulting from overstated adjustments, deductions, and exemptions or unjustified tax credits, is now approximately $300 billion a year (see Table 25.3). In addition to wage income, this includes other major sources of noncompliance, such as small business and small farm income and income from rents and

---

[7] *IRS Data Book 2011*, Tables 2 and 4, IRS FS-2012-6, January 2012.
[8] By its nature, it is difficult to obtain accurate measures of the underground economy. For a review of the literature and alternative methodologies, together with a range of estimates, see E. Feige and R. Cebula, *America's Unreported Economy: Measuring the Size, Growth and Determinants of Income Tax Evasion in the U.S.*, MPRA Paper No. 34781, September 2011.

TABLE 25.3 TAX GAP FROM INDIVIDUAL AND CORPORATE INCOME THAT ESCAPES THE TAX NET ($ BILLIONS)

TAX GAP COMPONENT	TY2001	TY2006
Estimated total tax liability	**2112**	**2660**
Tax paid voluntarily and timely	**1767**	**2210**
Gross tax gap	**345**	**450**
Voluntary compliance rate	**83.7%**	**83.1%**
Enforced and other late tax payments	**55**	**65**
Net tax gap	**290**	**385**
Net compliance rate	**86.3%**	**85.5%**
Nonfiling gap	**27**	**28**
Individual income tax	*25*	*25*
Estate tax	*2*	*3*
Underreporting gap	**285**	**376**
Individual income tax	*197*	*235*
Non-business income	56	68
Business income	109	122
Adjustments, deductions, exemptions	15	17
Credits	17	28
Corporation income tax	*30*	*67*
Small corporations (assets < $10m)	5	19
Large corporations (assets ≥ $10m)	25	48
Employment tax	*54*	*72*
Self-employment tax	39	57
FICA and unemployment tax	15	15
Estate tax	*4*	*2*
Underpayment gap	**33**	**46**
Individual income tax	*23*	*36*
Corporation income tax	*2*	*4*
Employment tax	*5*	*4*
Estate tax	*2*	*2*
Excise tax	*0.5*	*0.1*

SOURCE: Internal Revenue Service, *Tax Gap for Tax Year 2006: Overview*, January 2012, Table 1; and *Tax Gap "Map": Tax Year 2006*, December 2011.

royalties. Currently, around half of small business income, three-fourths of small farm income, and half of rents and royalties go unreported.[9]

Unreported illegal income, though not so important as some of the sources of tax evasion given in Table 25.3, is still an important source of tax evasion.[10] (Al Capone, the 1920s mobster, was convicted for failing to report his illegal income when other charges would not stick.) The IRS estimates that the total net "tax gap" (tax liabilities that are not collected after enforcement) for all taxes it administers is $385 billion, indicating a net taxpayer compliance rate of 86 percent.

In recent years, the government has been auditing less, but increasing the efficiency of its audits; a higher fraction of those audited have had to pay more in taxes.

REDUCING TAX AVOIDANCE

Tax avoidance—taking advantage of all the loopholes in the tax structure—results in significant erosion of the tax base. In the previous chapter, we discussed the principles of tax avoidance, as well as some of the more important tax shelters. In the early 1980s, tax avoidance had become rampant, fostering a general public impression that the tax system was not working. This provided the impetus for many of the reforms introduced in 1986. Although these reforms managed to reduce the degree of tax avoidance through tax shelters, they came at the price of increasing the complexity of the tax code.

REDUCING ADMINISTRATIVE AND COMPLIANCE COSTS

We have seen that the complexity of the tax system results in problems of compliance and in perceptions of inequities—but it also contributes

---

[9] It is obviously easy for those who receive payments in cash simply not to report cash income. Other noncompliance problems arise from reporting as expenses items that are not really business expenses. Today, sales of stocks and bonds are reported to the IRS, but other asset sales typically are not, and until recently, the IRS did not receive independent information about the basis (the purchase price). However, stock brokers and mutual fund companies became subject to basis reporting for most stock purchases in 2011 and for all stock purchased in 2012 and subsequent years.

Estimates of noncompliance are obtained through detailed and thorough audits of a relatively small number of taxpayers under the IRS Taxpayer Compliance Measurement Program. The costs of such audits (both to the government and the taxpayer) are great, and do not justify the extra revenue raised. The purpose of these audits is to provide the IRS general information concerning compliance. The figures cited here are taken from the most recent "Tax Gap" study, undertaken in 2006.

[10] It is estimated that Americans spend more than $100 billion on illegal drugs. (Source: Executive Office of the President, Office of National Drug Control Policy, *What American Users Spend on Illegal Drugs, 2012*, Table ES.3.)

greatly to the costs of collecting taxes. Because the United States has a self-assessment system of individual and corporate taxation, the burden of coping with this complexity falls mainly on the taxpayer, creating very high compliance costs. This burden consists not only of the direct costs of filling out the tax returns, but also the indirect costs of record keeping required to comply with the tax laws, accounting for almost two-thirds of taxpayer compliance costs.

In 2008, the IRS estimated that households and businesses spent 7.6 billion hours a year complying with tax filing requirements. This was equivalent to one year of work for 3.8 million full-time employees. The total value in monetary terms was $193 billion, equal to 14 percent of individual and corporate income tax receipts (net of refunds) in 2008. This contrasts sharply with the IRS collection cost in 2008 of 0.41 percent.[11]

The IRS's own resources devoted to tax collection have grown in tandem with the taxes collected. The cost of collecting $100 in taxes was 45 cents in 1967, 60 cents in 1993, and 53 cents in 2010.[12]

SOURCES OF COMPLEXITY

Whenever government taxes some income at different rates than other income (or does not tax some income, or provides a tax deduction or credit for some item), taxpayers have an incentive to ensure that income is received in a form that receives favorable treatment. The greater the tax consequences—the higher the tax rates and the greater the differences in tax treatments—the greater the incentives. Making clear distinctions and definitions in tax law is much harder in practice than in theory. Questions that might seem to be merely philosophical—What is a family? When is a wall a "structure" and when is it "equipment"? What is interest? How do we distinguish between repayment of principal and interest?—take on real importance when money is at stake.

Just as we saw earlier a trade-off between base broadening and the pursuit of other social and economic objectives, so too is there a trade-off between these other objectives and complexity.

Much of the complexity arises out of an attempt to reduce tax avoidance—particularly in the area of taxation of returns to capital, but in other areas as well. When is an individual really a farmer, so that losses should be tax deductible? When is farming simply a tax avoidance

---

[11] The National Taxpayer Advocate, "The Most Serious Problems Encountered by Taxpayers," *2008 Annual Report to Congress*; *IRS Data Book 2008*, Tables 1 and 29; and J. Slemrod, "Old George Orwell Got It Backward: Some Thoughts on Behavioral Tax Economics," CESifo Working Paper No. 2777, September 2009.

[12] *IRS Data Book 1997*, Table 28; *IRS Data Book 2011*, Table 29; and Office of Management and Budget, *Budget of the United States Government, Fiscal Year 2013*, Historical Table 10-1.

scheme? When is it a "consumption" activity, in which case losses should no more be deductible than expenditures on food or clothing? Even if the taxpayer knew his or her own motives clearly, the tax authorities cannot tell what motivated someone to buy a farm. Rules designed to distinguish real farmers from those who claim to be farmers for reasons of tax avoidance add to the complexity for everyone involved.

The complexity of the tax system has led to increasing dissatisfaction with it and with the IRS, which administers it. Many are convinced that there are a myriad of loopholes of which others can take advantage, but they cannot—the tax system is fundamentally unfair. The seeming complexity induces anxiety, compounded in some cases by a fear of harsh enforcement techniques of the IRS. Horror stories fill the newspapers— small underpayments cascading through penalties and fines to massive obligations, with the IRS attaching bank accounts and putting liens on houses. In some cases, the IRS appears to have been misguided, and given the millions of tax forms that have to be reviewed every year, such mistakes are inevitable. The crescendo of complaints against the IRS reached such an intensity that the Clinton administration undertook major administrative reforms in 1997, which gave taxpayers more rights in their struggle against the IRS. More of the burden of proof—that the taxpayer owed the government money—was placed on the IRS. In addition, to assist taxpayers in resolving their disputes with the IRS, the Taxpayer Advocate Service was strengthened.

The overall objective of these reforms was to make the IRS more service oriented so it treated taxpayers as clients rather than criminals. In the aftermath of these reforms, some critics claim that IRS enforcement capabilities were severely compromised as reformers pushed the comparison between taxpayers and clients too far. Unlike clients of private establishments, such as department store or hair salon customers, taxpayers are involuntary customers—they cannot refuse to patronize the IRS. Thus, although the emphasis should indeed be on facilitating voluntary taxpayer compliance (payment before enforcement), now at 83 percent (see Table 25.3), an understanding of taxpayer composition and behavior for the remaining 17 percent of taxes not paid voluntarily, as well as a credible threat of sanctions for noncompliance, are nonetheless essential for effective and equitable tax administration, given that paying is mandatory, not optional.

Some of the complaints against the IRS are based on perception as much as on reality. For example, two-thirds of the taxpayers in 2012 did not itemize their deductions, but instead took the standard deduction, greatly reducing the complexity of their tax returns. However, a large and increasing number of those filling out even these simplified returns turn to professional tax preparers.

TAX SIMPLIFICATION

The U.S. tax structure is highly complex. Hence, a major impetus of tax reform has been tax simplification.

The complexity:

- Gives rise to high costs of compliance, especially record keeping.
- Creates uncertainty in calculating tax liabilities, even with the help of tax professionals.
- Reduces revenue by opening up opportunities for tax avoidance, particularly for wealthy taxpayers.
- Decreases transparency and increases perceptions of inequities, undermining incentives for voluntary compliance.
- Is caused partly by government's attempt to reduce tax avoidance, partly by attempts to achieve social and economic objectives through the tax system.

The more fundamental problem is that the IRS has been unable to keep pace with the technological revolution in the rest of the economy. In principle, with W-2s and interest and dividend payments submitted directly to the IRS (with an increasing fraction submitted electronically), the IRS should be able to perform the tax calculations for all taxpayers not itemizing deductions; all these taxpayers need to provide is information about their family status. Even most itemized deductions could be handled by the IRS (especially state and local income taxes and mortgage payments). For example, in several Scandinavian countries, income tax authorities now use advances in information technology to automatically fill in and file tax returns for most of their taxpayers, so less than 10 percent of tax filers use tax preparers, as opposed to 60 percent in the United States. In fact, however, the IRS's computerization program has been less successful than one might have hoped.

THE 1986 TAX REFORM

The 1986 tax reform had as one of its major objectives the simplification of the tax code. In this, by most accounts, it failed. The major simplification that the supporters of the reform pointed to, the reduction in the number of tax brackets, was a superficial one. After one's taxable income has been calculated, looking up in the tax table the tax that is due is an easy task, regardless of the number of brackets.

In fact, the new distinctions introduced in the tax law between various categories of income made the tax law more complicated. As we said in Chapter 24, distinguishing among passive, investment, and ordinary income has required a host of definitions, regulations, and court cases. Indeed, the 1986 tax law did not even result in any simplification of tax forms; some forms became more complicated, and several new forms were introduced.

Similarly, the 1993, 1997, 2001, 2003, and 2012 tax bills offered no progress toward a simpler tax code. Indeed, because all these bills introduced important new elements of preferences, they served significantly to further complicate the tax system.

TRANSITION ISSUES AND THE POLITICS OF TAX REFORM

Changes in tax laws always entail problems; indeed, there is a maxim that "old taxes are good taxes." Each new tax law requires a myriad of court cases to determine how each provision of the new law is to be interpreted. Until these interpretations are clear, there is uncertainty, and uncertainty has its own economic costs.

The most difficult problems are those associated with capital (investments). As we saw earlier, tax provisions are, to a large extent, capitalized—reflected in the value of assets. Accordingly, changes in tax laws result in changes in capital values. Eliminating the special provisions for commercial real estate might drastically reduce the demand for such real estate. In the short run, prices would fall. Owners would suffer a capital loss. At the lower prices, construction would stop. In the long run, prices of buildings would be restored, as the excess supply dried up. Meanwhile, however, both owners of commercial real estate and construction firms would suffer. This is precisely what happened in the aftermath of the 1986 tax reform.

Such changes raise fundamental equity issues as well as political problems. Is it unfair to change rules? As we have noted before, investments are made with certain expectations concerning taxes. Changing taxes provides windfall gains to some and windfall losses to others, even when the taxes do not *directly* affect the investment in question. A tax benefit for owner-occupied housing reduces the demand for rental housing, and, accordingly, has adverse effects on the price of multifamily rental properties.

Those who bought a house expecting that they could deduct the interest payments on their mortgage and their property taxes from the income taxes will rightly feel aggrieved if those benefits are taken away—they may even have a hard time making ends meet. That is why when new tax laws are introduced, there is often a slow transition; for instance, those with existing mortgages might be able to deduct interest for up to ten years.

TRANSITION ISSUES

Transitions give rise to equity issues—changes in asset values, windfall gains to some, losses to others. There are efficiency costs associated with adjusting to new tax laws.

General principles:

- Try to avoid frequent tax law changes.
- Make changes gradually that affect capital asset values.
- Better to be approximately right than precisely wrong —try for simple over optimal.
- Formulate tax policies in the context of tax administration, as the impact of tax laws is largely determined by how they are implemented.

Politics of tax reform:

- Revenue-neutral taxes have as many winners as losers.
- Losers often feel losses more acutely.
- Not zero-sum—the majority could be better off.
- Gimmicks are often required to garner enough support.
- To reduce scope of conflict, reforms often focus on revenue-neutral and distribution-neutral tax changes.

Although such transitional issues are of concern, they clearly cannot dominate economic policy making: otherwise, one could never abolish any provision of a special interest group, no matter how distortionary. Those who are hurt by a new tax law often get much more concerned about the inequities than those who benefit; they lobby Congress for special transition rules. In the final days of the passage of every tax bill, a host of special provisions, worth billions of dollars, are adopted to the benefit of certain companies, particularly companies in the congressional districts of the bill's proponents. To some, this is just a reflection of the politics of taxation. To others, it is just a way to enhance the likelihood that the tax reform is a Pareto improvement—many of the special provisions are nothing but the side payments required to compensate the losers. To still others, the special provisions symbolize the defeat of the principles of tax reform.

The problems presented by transition lead to two general maxims: try to make tax reforms relatively seldom; and try to implement reforms gradually, especially reforms that have adverse effects on capital values.[13] The frequent revisions of the tax code over the past three decades run counter to the first maxim.

When President Reagan sent Congress his original proposal for tax reform, he made it clear that his objective was to change the *structure* of the income tax, not the level of taxation. Thus, he insisted that it be **revenue neutral**; that is, it raise the same amount of revenue as was being raised under the existing laws. Indeed, he even insisted that the overall degree of progressivity—the percentage of taxes paid by those in different income categories—not be substantially changed. The concern was that nothing would come of an open-ended discussion, in which all the structural issues were up for debate. He wished the discussion to focus on particular inefficiencies and inadequacies of the tax code.

To a large extent, tax reforms change the burden of taxation; there are many losers and winners. The fact that there are so many losers makes reform difficult.

Most economists, however, believe that important tax reforms are not "zero sum," that is, they are more than just redistributing income; more is at stake than just how the economic pie is to be divided. By improving economic efficiency, by eliminating distortions, reforms increase the size of the economic pie. The problem, however, is that those who worry that the tax law change will hurt them feel the loss far more intensely than those who might gain; those who reform advocates say will gain are often skeptical that the gains will actually be realized.

---

[13] Concern about these equity effects often leads to provisions called "grandfathering," which extend preferential treatment to those currently receiving the benefit but not to others. Grandfathering thus reduces one sense of inequity—that arising out of unanticipated changes in tax treatment—but it introduces a new source of inequity, between those who have access to the preferential treatment and those who do not.

In the discussions of the 1986 tax reform, to make the increase in the size of the pie look even bigger, corporation income taxes were increased so individual income taxes could, overall, be reduced. Because most people do not accurately see the burden imposed on them by the corporation tax, this enabled more individuals to see themselves as net winners.

Any tax bill is viewed by special interest groups as an opportunity to get favorable treatment. Congress has tried to reduce this pressure by requiring that any congressperson who proposes an amendment granting tax relief must specify how the offsetting extra revenue would be raised—that is, which taxes would be increased, or which loopholes would be eliminated.

TAX REFORMS FOR THE TWENTY-FIRST CENTURY

Considerable dissatisfaction with the tax system remains. Two groups of reforms have been under discussion recently: minor reforms in the current tax system, and major new reforms.

REFORMS WITHIN THE CURRENT FRAMEWORK

Recent reforms have struggled to achieve a balance among competing objectives.

1. High marginal tax rates increase deadweight loss, reduce tax compliance, and increase incentives for tax avoidance. Low marginal tax rates reduce progressivity.

2. Base broadening allows the reduction of tax rates. Preferential treatment of certain categories of income and expenditures introduces complexity, inequity, and opportunities for tax avoidance. To the extent that preferential treatment is introduced to enhance equity or to promote socially desirable objectives, eliminating preferential treatment reduces the tax system's instrumentality for achieving objectives other than raising revenues.

PROMOTING VALUES THROUGH THE TAX SYSTEM   In the 1990s, there was increasing reliance on tax policy as an instrument for pursuing economic and social goals, simply because of the strong resistance to new taxes and increased direct expenditures. Tax expenditures appear to be more a politically acceptable means than direct expenditures, even

if they are less well targeted and introduce more economic distortions. Rather than focusing on base broadening, the reforms of 1993 and 1997 focused on using the tax system to express values—even when the provisions themselves have limited economic impact.

A number of reforms have focused on *family values*—encouraging adoptions, the child care credit, and the reduction of the marriage penalty. Some want to encourage more women to participate in the labor force, by expanding the child care credit; others want to encourage mothers to stay home and take care of their own children, by reducing the child care credit. Other reforms, such as the tuition tax credit, have focused on encouraging education.

PROMOTING SAVINGS AND INVESTMENT Still other reforms have focused on encouraging savings—especially the steady expansion in IRAs over the past decade and the preferential treatment of capital gains in the 2001 and 2003 tax reforms. However, whereas some reformers would like to extend the preferential treatment of the return to savings, taxing only consumption or wage income, others want to eliminate the preferences currently extended.

TAX SIMPLIFICATION A number of reforms concern simplifying the tax system. This was one of the motivations for the provision that exempted up to $500,000 in capital gains on owner-occupied housing. (The provision's political appeal did not go unnoticed by its advocates in both parties, however; the fact that the provision may have exacerbated the distortions of the tax system, by further encouraging investment in housing, was not of much concern.)

One way of reducing the taxpayers' burden would be for the government to assume responsibility for calculating taxes for the three-fourths of individuals who have simple tax forms—only wage income, interest, dividends, and capital gains on stocks.

Many of the reforms that would broaden the tax base by eliminating special provisions would also help simplify the tax system. On the other hand, recent reforms that have tried to use the tax system to advance social and economic policies have moved in the opposite direction.

MAJOR NEW REFORMS

There are two important sets of major reforms that have been discussed in recent years. One continues the tradition noted earlier of trying to use the tax system to correct economic distortions and address major societal problems, while the other focuses more on reducing complexity and distortions introduced by the tax system itself.

ENVIRONMENTAL TAXES   Another major reform is to shift more of the burden of taxation onto items that society wants to discourage and away from areas that should be encouraged. Thus, taxes on cigarettes and alcohol (sometimes called "sin taxes") and on gasoline and other goods that pollute the environment would be increased, and taxes on work and savings would be reduced. Taxes designed to address externalities, like pollution, are called *corrective* or *Pigouvian* taxes, after Arthur Pigou, the famous Cambridge economist who advocated such taxes almost a century ago. Such taxes not only raise revenues, but they also help align private incentives with social returns, and, in that sense, actually work to improve the overall efficiency of the economy.

To be sure, such taxes may have adverse effects on GDP (the measure of national output), *as conventionally measured*, because conventional measures do not include environmental benefits such as clean air or water. However, there is a consensus that standards of living, and especially health, are improved when smog does not cloud the skies three hundred days a year.

The most important of such taxes are those that affect energy consumption in general, and the use of coal and other hydrocarbons in particular. With the accumulation of scientific evidence concerning the potential of global warming and other changes in climate from greenhouse gas concentrations, there has been a concerted international effort to reduce greenhouse gas emissions. This effort began in 1979 with the first World Climate Conference in Geneva, and was followed by establishment of the Intergovernmental Panel on Climate Change in 1988, adoption of the United Nations Framework Convention on Climate Change at the Earth Summit in Rio de Janeiro in 1992, adoption of the Kyoto Protocol in 1997, and the search for an international agreement after the Kyoto Protocol expires in 2012. Greenhouse gas emissions are much greater (per unit energy generated) from coal than from oil, which, in turn, has greater emissions than natural gas. A tax on greenhouse gas emissions would discourage the use of coal and oil, and thus reduce greenhouse gas emissions. In practice, the emissions cannot themselves be directly measured, but a **carbon tax**, which would set the tax rates on coal, oil, and natural gas in relationship to their carbon content (which determines emissions), would have much the same effect. Carbon taxes have received considerable attention in Europe and have been instituted in Denmark, Finland, and the Netherlands. In 1993, a carbon tax was considered briefly by the U.S. government, but opposition from coal-producing states quickly forced the consideration of an alternative energy tax, called the **BTU tax**. A BTU (British thermal unit) is the standard way of measuring energy output. The BTU tax was designed to be a tax on energy consumption; reducing energy consumption would reduce greenhouse gas emissions,

although not so efficiently as the more targeted carbon tax. The BTU tax actually proposed by the Clinton administration in February 1993 was a hybrid between a straight energy tax and a carbon tax, with rates on coal adjusted slightly higher and on natural gas slightly lower.

Not surprisingly, opposition from both energy-producing companies and high–energy-using firms (such as aluminum smelters) combined to defeat the tax, leaving in its place a slight expansion of the gasoline tax. Although this tax has some environmental benefits, these are far smaller than those of the original proposals. Moreover, the gasoline tax is a unit tax rather than an ad valorem tax, so its real value has declined precipitously over the years. Its impact has also diminished because a combination of high gasoline prices and more fuel-efficient vehicles has resulted in people reducing their total gasoline consumption.

Other categories of environmental taxes that are currently under discussion, and that, in the long run, have some chance of success, focus on pesticide use and recycling. A major source of water pollution today is from pesticides and fertilizers used by farmers (what are called *nonpoint sources* of pollution). Unlike pollution generated by large factories, these are very hard to control directly. One way to effectively control this form of pollution is to make it more expensive to use the pollutants by taxing them.

The failure to reuse resources contributes to excessive waste. It costs money to recycle, but it also costs money to dispose of wastes. There is increasing concern about the social costs of waste disposal. No one wants a dump site in their backyard, partly because of the danger of toxic substances. As a result, communities have found it harder and harder to find places to dispose of waste. Increased fees for disposing of waste will encourage greater recycling. However, there is concern that monitoring disposal is extremely difficult, and as fees for legal disposal increase there will be increased use of illegal means of disposal.

Another way of increasing recycling is to increase the cost of new versus recycled materials. For instance, a tax on virgin pulp (pulp from newly felled trees) will encourage more use of recycled pulp in manufacturing paper. The bottle recycling tax is now perhaps the most well-known tax designed to encourage recycling.

Currently, the United States uses a variety of regulations to encourage good environmental policies in general and energy efficiency in particular. Most notable are the corporate average fuel economy (CAFE) standards with which automobile manufacturers must comply, designed to increase the energy efficiency of cars. By switching from regulation to taxes, two benefits may be realized: (1) economic efficiency may be enhanced, and (2) revenues will be raised, reducing the burden that must be imposed in the form of distortionary taxation. This notion of the *double*

dividend to be realized through environmental taxation has gained considerable attention from economists and policy makers.[14]

FINANCIAL SECTOR TAXES   The Great Recession of 2008 focused attention on the activities of the financial sector. The financial sector is supposed to allocate resources—providing credit to those who can use it well—and manage risks. It had in fact badly misallocated investments and created risks, which brought the global economy to the brink of ruin. Much of its activities were centered around speculation, generating huge fees for the banks, but without any observable benefit for the "real" economy. The financial sector had growth to the point where it accounted for 8 percent of GDP and 40 percent of corporate profits, and was the source of much of the growing inequality. Not only did there seem to be excessive financial activity, some financial institutions—including some very big banks—had grown to the point where they were too big to fail: if they risked failure, the government would come to the rescue for fear of what their demise might do to the entire economy. These institutions had many advantages: because those who provided them with capital knew that the government would bail them out should they run into trouble, they could obtain capital at lower interest rates than competitors, not because they were more efficient, but because of the implicit subsidy. Moreover, these too-big-to-fail banks often proved (according to critics) too-big-to-jail: even when massive violations of laws and regulations were uncovered, the government was loathe to prosecute.

In response, there have been a number of proposals to impose financial sector taxes, which, like environmental taxes, would both raise revenue and improve the overall efficiency of the economy. The Obama administration at one time suggested a bank levy that would increase with the size of the bank, partially correcting the unfair advantage that accrue to large banks. Many European countries have adopted a financial transactions tax. Such a tax, imposed at a very low rate on every transaction, discourages short-term speculative trading, but have no significant effect on long-term investments. There is an increasing concern that financial markets have resulted in excessive short-termism—a focus on what happens in the next three months—while sustainable growth requires long-term

---

[14] Environmental taxes could lead to less spending on other taxed goods or lower labor supply or savings, thus offsetting the revenue gains that would otherwise be received. See, for instance, L. H. Goulder, "Environmental Taxation and the 'Double Dividend': A Reader's Guide," *International Tax and Public Finance* 2, no. 2 (August 1995): 157–183; W. E. Oates, "Green Taxes: Can We Protect the Environment and Improve the Tax System at the Same Time?" *Southern Economic Journal* 61, no. 4 (April 1995): 915–922; C. Carraro, M. Galeotti, and M. Gallo, "Environmental Taxation and Unemployment: Some Evidence on the 'Double Dividend Hypothesis' in Europe," *Journal of Public Economics* 62, no. 1–2 (October 1996): 141–181; and A. L. Bovenberg and L. H. Goulder, "Optimal Environmental Taxation in the Presence of Other Taxes: General-Equilibrium Analyses," *American Economic Review* 86, no. 4 (September 1996): 985–1000.

thinking. This tax, its advocates hope, might be a small nudge in that direction.

Not surprisingly, such taxes have encountered strong opposition from the financial sector. It is thus perhaps surprising that in spite of such opposition, some European countries have agreed to adopt such a tax. The United States and United Kingdom, however, with large and politically influential financial sectors, have been strong opponents of adopting such a tax on a global basis.

Other major reforms currently being debated embody the themes of simplification, base broadening, and promoting savings and investment.

FLAT-RATE TAX   We noted that some of the complexities of the tax system arise from its progressivity. This has led to a proposal to have a single, flat rate. Individuals would pay a given percentage of their income above a certain threshold level. The proposal, although discussed by economists for a long time, entered the political arena in 1996 when it was advocated by presidential candidate Steve Forbes in the Republican primaries. Under **flat-rate tax** proposals, rates at the upper end would be lowered, necessitating an increase in rates elsewhere (unless one believed that the tax rates at upper incomes were so high that they could be reduced without reducing tax revenues). As a result, these proposals have not been well received.

One of the main advantages of the flat-rate tax proposal is the ease with which it can be administered. Because everyone pays the same marginal tax rate, it makes no difference whether the tax is collected at the level of the individual (who receives the income) or the level of the business (say, as the owner pays wages). Because flat-rate tax proposals typically eliminate all special provisions, individuals could fill out the tax form on the back of a postcard. Note, however, that this simplification has little to do with the single rate, but rather deals with the elimination of special provisions.

CONSUMPTION TAXES   Many of the major tax reforms propose to tax consumption only. This is true of the flat-rate proposals just discussed and the value-added tax proposal discussed in the next section. As we saw in Chapters 21 and 24, there are strong arguments for taxing consumption only: much of the complexity of the tax code arises from the attempt to tax interest; a **consumption tax** imposes the same tax on individuals with the same lifetime incomes, whereas the income tax discriminates against individuals who prefer to consume later in life; it seems more appropriate to tax what individuals take out of society (their consumption) rather than what they contribute (measured by their income); and switching to a consumption tax might encourage savings, which remain abysmally low in the United States.

Critics of these proposals argue that given the insensitivity of savings to interest rates, switching to a consumption tax will not promote savings. Advocates point out that they are concerned with *switching* from an income tax to a consumption tax, and that what matters thus is not the interest elasticity, but the *compensated* interest elasticity, of savings—taking into account that as the tax on interest income falls, other taxes would have to increase.[15]

Advocates of the consumption tax believe that it would be easy to administer the tax. The simplest procedures focus on cash flow: an individual would be taxed on net cash flow (income) minus net additions to savings:

$$\text{Consumption} = \text{income} - \text{savings.}$$

As individuals' purchases and sales of assets, like stocks, are already recorded, little extra work would be required. Critics worry that there are many other assets for which purchases and sales are not so easily monitored (like works of art), and say that the transition into the system would be complicated by the fact that individuals have large amounts of assets already in their possession. Sales of these assets used to finance consumption might be hard to detect. Others argue that because the income with which these assets were purchased was already taxed once, it would be unfair to tax the consumption derived from their sale.

Politically, however, the major objection to adopting a consumption tax has been that to achieve the same degree of progressivity as under the current income tax system, marginal tax rates on consumption would have to be very high—many individuals would have to face marginal tax rates of 50 percent or more. The high marginal tax rates foreseen have so far prevented the proposal from receiving great popular attention.

VALUE-ADDED TAX   A **value-added tax (VAT)** is imposed at each stage of production on the difference between a firm's sales and its purchases from other firms—that is, on the value added by the firm. The VAT has become a major source of revenue in most European countries, with rates as high as 20 percent.

There are several ways of calculating a country's output. One way focuses on sales of final output; another, on adding up the value added at each stage of production; a third, on adding up the incomes of all individuals. All three are equivalent, and all three form bases of levying taxes. A sales tax imposed on final sales, a VAT, and an income tax can, in this

---

[15] The uncompensated interest elasticity of savings is low because of the offsetting income and substitution effects; a lower interest income tax means that individuals are better off, and this increases their consumption (reduces savings). However, if at the same time that the interest income tax is reduced, other taxes such as wage taxes are increased, the income effect is eliminated, or at least reduced, so the predominant effect is the substitution effect. The substitution effect unambiguously leads to more savings.

ORDINARY INCOME VERSUS CAPITAL GAINS

When capital gains are taxed at lower rates than ordinary income, individuals have an incentive to try to characterize income as a capital gain rather than as ordinary income. For instance, if a bond is issued paying an interest rate lower than the prevailing market rate of interest, the bond will sell at a discount; thus, if the market rate of interest is 10 percent, a bond promising to pay $100 in a year plus 5 percent interest will sell for approximately $95.

Such a bond will yield the owner a 10 percent return—5 percent interest plus 5 percent capital gain. The 5 percent capital gain is nothing but disguised interest. In recent years, tax laws have attempted to stop this form of tax gimmickry by subjecting *original issue discounts*—that is, discounts that occur at the time the bonds are issued—to ordinary income taxation, not the preferential capital gains rates.

sense, all be made equivalent, although there may be marked differences in the costs of administering the different taxes. When based on consumption, a VAT is similar to a retail sales tax that is collected incrementally rather than entirely at the final stage of production. The advantage of the VAT is that most of value added occurs within large corporations. A tax on final sales puts a large burden on retailers, many of which are small businesses. As a result, there is a high degree of compliance with the VAT (except in the informal economy).

A major advantage of the VAT is its relative simplicity. Critics, however, point out that it is a proportional tax, and thus less equitable than a progressive consumption or income tax. Although some degree of progressivity can be obtained by taxing different goods at different rates (for instance, exempting food), doing so reduces the tax's simplicity and introduces distortions—negating the great advantages of the VAT. However, when viewed in respect to lifetime consumption, the VAT is less regressive than it seems because the savings of wealthy people, which are not taxed when these individuals are young, is taxed when they consume out of savings in their later years, and recent field studies have shown that the VAT is actually progressive in many developing countries because it is not enforced where low-income families shop.[16] But without as effective gift and inheritance tax, the problem of regression remains.

[16] See, for example, G. Jenkins, H. Jenkins, and C. Y. Kuo, "Is the Value Added Tax Naturally Progressive?" Queen's Economics Department Working Paper No. 1059, April 2006.

COMBINING THE VAT WITH OTHER FORMS OF TAXES Most proponents of the VAT do not advocate that it should replace the income tax—in Europe, both are used. Rather, they believe that it can be used to reduce substantially the revenues to be collected from the income tax, and hence the high marginal tax rates; as much of tax avoidance is related to the level of marginal rates, advocates believe that it will decrease the distortions and increase the equity of the tax system. Critics point out that the total distortion of the tax system is related to the sum of the (marginal) tax on income and the VAT, and that unless this is reduced, the deadweight loss associated with the tax will not be reduced. Moreover, they say, the gains in compliance costs from the reduction in the marginal tax rate will be more than offset by the additional administrative costs associated with collecting the VAT. Finally, there is concern that because the VAT is collected in a piecemeal way, individuals will not be conscious of the full scale of the taxes they pay, and this will lead politicians to increase the overall tax burden. For those who would like to see a larger public sector, this is an advantage; but for those who would like to see a smaller public sector, it is a disadvantage.

> ## RECENT TAX REFORMS AND REFORM PROPOSALS
>
> Conflicting objectives:
>
> - Increasing progressivity vs. lowering marginal tax rates to reduce distortions
> - Simplification vs. use of tax code to achieve broad range of objectives
>
> Reforms under current framework:
>
> - Some of base broadening/simplification of 1986 undone by tax reforms of 1990s and 2000s
> - Taxes used to:
>   Promote family values
>   Promote savings and investment
>   Promote education
>
> Major reforms proposed:
>
> - Value-added tax
> - Energy taxes (taxing "bads")
> - Flat tax
> - Consumption tax
> - Financial taxes

Advocates of a combined approach argue that it could be used to greatly reduce administrative and compliance costs, while retaining a fair degree of progressivity. For instance, a 10 percent VAT combined with a $100,000 standard deduction and income tax rates adjusted downward to reflect the VAT payments, could generate approximately the same revenue as the current tax system and have about the same degree of progressivity, but could eliminate the necessity of filling out individual income tax forms for 96 percent of Americans. It would also virtually eliminate the distortions associated with the special provisions of the tax code (as almost all taxpayers would take advantage of the $100,000 standard deduction), without engaging in the political fights associated with their outright elimination.[17]

---

[17] Under the current system, the very poor pay no tax, whereas with the VAT, they do. The effects of this could be offset by a broadened earned income tax credit, but doing so would reduce some of the simplification provided by the VAT system. On the other hand, with improved computerization, the IRS could determine an individual's eligibility for benefits, as long as the individual provided the requisite household data and certified that nonwage earnings did not exceed a threshold level.

IRAs AND NATIONAL SAVINGS

Even though IRAs are designed to encourage savings, there is considerable controversy about whether they actually do. As they are currently designed, there is typically a maximum contribution, so the marginal return to savings is unaffected for anyone saving more than the threshold level. Anyone with assets in a portfolio (such as a non-IRA savings account) can simply transfer those assets into an IRA account. The tax provision encourages people to transfer their assets, not to save more.

Moreover, the special provisions result in lowered government tax revenue—and thus lower government savings. Unless private savings increases by a greater amount than the losses in taxes, national savings will actually be reduced.

In spite of this, there is some evidence that IRAs may encourage personal savings, perhaps because banks and mutual funds have used the lure of the IRA to help advertise the virtues of savings.

INTEGRATING STATE AND LOCAL INCOME TAXES   Critics of all these reforms point out that not much savings will occur unless there are commensurate reforms in state and local income taxes. Even though the distortions associated with unreformed state taxes are likely to be small (simply because state tax rates tend to be low), the administrative costs can remain high; the time to fill out complex tax forms does not, in general, depend on the tax rates, so unless states reform, households and firms will continue to waste huge amounts of time filling out complex tax forms. Traditionally, state and local authorities have followed the leadership of the federal government, but there are some notable exceptions. California has its own depreciation formula, for instance, necessitating that firms in California perform two different sets of calculations.

In addition, if a national VAT were to be introduced, it would have to be integrated into the current system of state and local sales taxes because the tax base would essentially be the same. This is one of the main obstacles to adopting a VAT in the United States, as it is perceived by many subnational jurisdictions as impinging on local government rights under the federal system's principle of fiscal subsidiarity. In this context, subsidiarity means that the federal government should be responsible only for taxes that cannot be administered effectively by a lower level of government.

REVIEW AND PRACTICE

SUMMARY

1. High administrative costs and low levels of compliance, the complexity of the tax code, the perceived inequities of the tax structure, and the large distortions that some argue are associated with current marginal tax rates, all have contributed to the impetus for tax reforms.

2. Many of the problems of the income tax system have arisen because too much has been asked of it. It was supposed to provide economic incentives (e.g., to invest, to save, to encourage health insurance, to support state and local governments) and redistribute income, as well as raise revenue.

3. The 1986 Tax Reform Act decreased the highest marginal tax rate, removed those in poverty from the tax rolls, and made tax avoidance more difficult. However, most Americans did not face greatly reduced marginal tax rates, and for the most part, the reductions simply reversed the effects of inflation of the 1970s.

4. The 1986 tax law did not simplify the tax code (in some ways it greatly increased the complexity of the tax system), nor did it really reduce tax avoidance activities, since it left some important loopholes. It also did not substantially broaden the tax base. To the extent that it "leveled the playing field" for investments, it did so primarily by eliminating the special investment incentives that had been introduced in 1981.

5. The 1993 tax bill increased progressivity by increasing marginal tax rates on the highest-income individuals and substantially increasing the earned income tax credit. As a result, the goal of ensuring that all families with one full-time earner would be out of poverty was almost attained. The 1997, 2001, and 2003 tax reforms reduced progressivity by restoring preferential treatment of capital gains, most of which accrue to the very wealthy. The 2003 tax bill further reduced progressivity by lowering taxes on dividends.

6. The 1993 and 1997 tax laws introduced special provisions for investments in small and new businesses, expanded IRAs, introduced new tax credits for education, and increased the child care credit.

7. In the current political climate, tax laws are being used to encourage specific activities (e.g., acquiring education) through tax expenditures because increases in direct expenditures are hard to obtain, and as an expression of values (e.g., encouraging families) by eliminating the marriage penalty and encouraging adoption.

8. Other major thrusts of tax reform under the current system include those to promote savings and simplify some of the provisions contributing to tax complexity.

9. Major reforms include taxing energy (carbon) usage or other activities detrimental to the environment, as well as taxing financial institutions and transactions. Other reforms include the flat tax, the value-added tax, and basing taxes on consumption rather than income.

KEY CONCEPTS

Base broadening

BTU tax

Carbon tax

Consumption tax

Flat-rate tax

Revenue neutral

Underground economy

Value-added tax (VAT)

QUESTIONS AND PROBLEMS

1. There is a widespread view that a consumption tax would hurt the poor. Is this necessarily the case?

2. The adoption of a *consumption tax* would have a significant effect on the market value of certain

assets. Which assets are likely to decrease in value? Which are likely to increase in value? Should the government do anything to compensate the losers or to tax the gainers?

3. There is a widespread view that the appropriate basis for taxation is an individual's lifetime consumption (or lifetime income). Discuss the inequities and inefficiencies that would arise from a consumption tax that had increasing marginal tax rates but no provisions for averaging years of low consumption with years of high consumption. Would these problems also arise under a flat-rate consumption tax?

4. In Chapter 22 we discussed the problems associated with choosing the appropriate unit for taxation (family versus individual). How would these problems be affected by the adoption of a consumption tax? A flat-rate income tax?

5. Many economists and politicians have argued that moving from an income tax to a consumption tax would be unfair unless a wealth tax or an inheritance tax were enacted at the same time. Explain the arguments on both sides of this issue.

6. Explain why the repeal of the investment tax credit might have increased the value of existing capital goods.

7. Assume that contributions to an IRA account in excess of 10 percent of income are tax deductible. Use a two-period diagram to show the budget constraint. Explain why this form of IRA account is more likely to increase savings than the traditional form.

8. Assume that the supply of oil is inelastic. What is the effect of a tax by all countries on oil? Does it reduce oil consumption? Assume that only Switzerland imposes a tax on oil. What happens to world oil consumption? To Swiss oil consumption?

9. Consider a two-period model in which oil (still in inelastic supply) can be used in either this period or the next period. In equilibrium, the price in the next period must be higher than the price in this period, if an owner of oil is to be willing not to sell it during this period. Explain why if the interest rate is 10 percent, the price must be 10 percent higher. (The principle that price must rise at the rate of interest is called *Hotelling's principle*, after Harold Hotelling, a distinguished professor of statistics at Columbia and North Carolina State, who first enunciated it almost three-quarters of a century ago.) What are the consequences of imposing a tax on oil at the same rate in both periods?

10. Using the median voter model discussed in Chapter 9, discuss what might be the consequences of eliminating the deductibility of all state and local taxes for the level of expenditures at the state and local level. How does your answer depend on whether the median voter itemizes deductions?

FURTHER ISSUES

The United States has a federal system of government, with some activities being undertaken at the state and local level, others at the national level. Chapter 26 explains the rationale for a federal system and some of the important fiscal interactions between the federal government and state and local governments. It also explores the role of competition among communities in ensuring that the correct levels and kinds of public goods are produced and that they are produced efficiently.

Chapter 27 briefly describes expenditures and taxes at the state and local levels. It focuses particularly on the incidence of taxes and expenditure programs in situations in which capital and labor are highly mobile.

Chapter 28 addresses a major issue of the past three decades: America's budget deficit, which soared during the 1980s, was brought under control in the late 1990s, but has since risen precipitously. The chapter analyzes the impact of the deficit, as well as taxation and government expenditures, on economic growth and stability.

26

INTER-GOVERNMENTAL FISCAL RELATIONS

FOCUS QUESTIONS

1. How are the responsibilities for providing public goods and services shared between the federal government and the states?

2. What are the economic principles that ought to govern the assignment of responsibilities? When can there be efficient decentralization of decision making concerning the provision and financing of public services?

3. What role should the federal government undertake in redistributing income from rich states to poor states? How does the federal government subsidize states now, and how effective are these subsidies?

The Constitution of the United States stipulates that powers not expressly delegated to the federal government (e.g., providing for the national defense, printing money, and running the post office) rest with the states. For a long time, the prevailing view was that this seems to leave responsibility for the provision of most public services (e.g., education, police and fire protection, and roads and highways) with the states. The Constitution is a flexible document, however, and court interpretations of it have essentially freed the federal government to provide many other services.

There has been an ongoing debate about **fiscal federalism**, the division of economic responsibilities between the federal government and the states and localities. Federalism, of course, spans issues that go beyond economics. For example, in the 1960s, 1970s, and 1980s, civil rights advocates urged a more active role for the federal government, while those who resisted emphasized states' rights. This chapter focuses on economic issues, for example, what goods and services should be provided locally or nationally. This issue has surfaced periodically; in his 1982 State of the Union message, President Reagan called for a "New Federalism," giving

states increased authority in welfare, with the federal government taking on more of the burden of paying for Medicaid. Critics argued that the New Federalism was a ploy for justifying cutbacks in federal assistance to states and localities, as a way of reducing the size of the federal government. Indeed, between 1980 and 1986, federal grants decreased from 3.3 percent of GDP to 2.6 percent. During the mid- and late 1990s, with a Republican majority in Congress, there were renewed demands for state control of federal programs that helped fuel the welfare reform of 1996.

This chapter briefly describes the broad division of responsibilities, then focuses on the central economic issues in fiscal federalism. It concludes with a few brief remarks about the underlying politics and philosophy in the debate.

THE DIVISION OF RESPONSIBILITIES

The relationships between the federal government and the states and localities are complex, and are not well described by a simple look at expenditures. There are two key issues: Who makes the decisions about the programs, and who pays for them? In some cases, the federal government pays for a program, and gives broad discretion to the states as to how to carry out the mandate. In other cases, the federal government essentially dictates all the terms, and the states simply administer the program.

For instance, in the SNAP food stamp program, eligibility standards and amounts are determined federally, and the states just administer the program. In some cases, the federal government gives **matching grants**— the state determines the level of expenditure (within limits) and the federal government pays a portion of the costs, which may depend on the per capita income of the state. In other cases, the federal government provides a **block grant**—a fixed amount of money subject to general expenditure guidelines. The state then bears the full costs of any expenditures above that amount. At one time, the federal government provided **general revenue sharing**—block grants that could be used for any purpose. (The federal government was sharing its revenues with the states, based on the presumption that the federal government could raise revenues more efficiently but states could make certain types of expenditure decisions more effectively.) Today, it no longer does this, but there are efforts to convert matching grants for specific purposes such as welfare into block grants

for those purposes. In 1996, the AFDC (Aid to Families with Dependent Children) program was replaced with a block grant program, TANF (Temporary Assistance for Needy Families), but the food stamp program remained federally financed, and the Medicaid program continued under a matching system.

Figure 26.1 shows the fraction of government expenditures for various categories financed at the federal level. (Bear in mind that expenditures give only a partial view of the role of each level of government in each activity.) States and localities retain most responsibility for education (primary, secondary, and public tertiary), public safety (police and fire protection), and transportation (public roads and mass transit). On the other hand, the federal government bears major responsibility for Social Security and income support, housing and community services, medical care, and agriculture. These patterns can be seen slightly differently in Figure 26.2, which shows how states and localities spend their money. Education is the single largest expenditure, followed by medical care and public safety.

Just as there is a division of responsibility between the federal government, on the one hand, and state and local governments on the other,
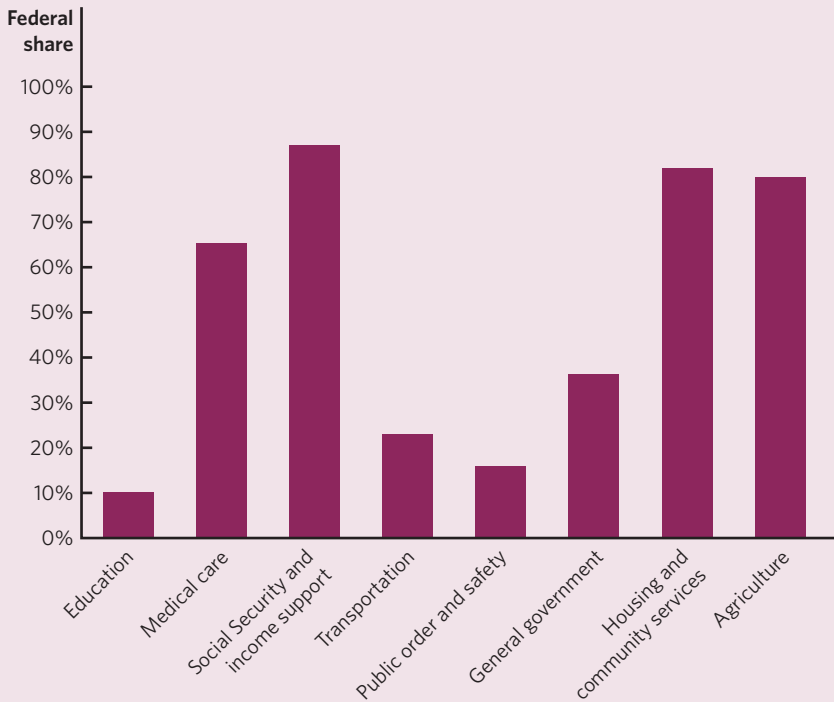
FIGURE 26.1

FRACTION OF GOVERNMENT EXPENDITURES FOR SELECTED CATEGORIES FINANCED AT THE FEDERAL LEVEL, 2008

The federal government finances most public sector spending on Social Security and income support, medical care, housing and community services, and agriculture, but state and local governments finance most public sector spending on education, policing and firefighting, and transportation.

SOURCE: U.S. Department of Commerce, Bureau of Economic Analysis, *National Income and Product Accounts Tables*, Tables 3.16 and 3.17.

- Interest payments (5.4%)
- Other (4.2%)
- General government (8.4%)
- Transportation (6.0%)
- Social Security and income support (7.5%)
- Education (35.0%)
- Medical care (20.5%)
- Public order and safety (13.0%)

so there is a division of responsibility between state governments and local governments. The division is a complicated one, involving financing, regulation, and administration. Thus, elementary and secondary schools are almost all run by local communities, but half the financing comes from the states, which also impose a variety of regulations. This is part of a complex set of intergovernmental transfers between states and local governments—almost one-third of total local government revenue comes from state grants. Thus, whereas almost 90 percent of public welfare and 60 percent of highway expenditures occur at the state level, almost all water and sewerage, solid waste, housing and community development, firefighting, and police expenditures occur at the local level.[1]

When the United States is placed in an international comparative context, no clear pattern emerges among high-income countries regarding division of responsibilities between different levels of government for most functions—the diversity is remarkable, with the exception of social protection, which is predominantly a central government responsibility (see Figure 26.3). For example, although health care is usually a central government responsibility, in Sweden, most health expenditures are by local government, and in Spain, health care is primarily a state-level responsibility. Similarly, in most countries, public safety is a central government responsibility, but in the United States and Germany, most public safety expenditures are at the subnational level.

---

[1]U.S. Census Bureau, "State and Local Government Finances and Employment," *2012 Statistical Abstract*, Tables 454–456.

OTHER INTERACTION BETWEEN THE FEDERAL GOVERNMENT AND THE STATE AND LOCAL GOVERNMENTS

The federal government affects the states and localities in a variety of other ways, besides providing grants. Its regulations and the federal tax code affect states and localities, just as they affect private businesses.

REGULATION   The Constitution restricts the laws that states can pass. The states cannot enact legislation that deprives an individual of the right to a trial, no matter how heinous a crime he or she may have committed, nor can states use racial or religious grounds to bar an individual from holding a job. Many Supreme Court decisions in recent years have countered state actions that are in violation of the Constitution.

State and local agencies may also be subject to the same pollution and environmental regulations that apply to private firms and individuals. In some cases, the federal government has mandated that state and local governments provide certain services (such as access facilities for the handicapped) without providing the requisite funds. Not surprisingly, the states and local communities have complained that if the federal

government attaches such importance to these services, it should also finance them (see case study, "Unfunded Mandates").

INCENTIVES Sometimes the federal government imposes its will through eligibility requirements for grants. For instance, until December 1995 transportation funds were made contingent upon maintaining speed limits of 55 miles per hour in urban areas and 65 in non-urban areas; educational funds are still made contingent upon having adequate affirmative action programs.

TAX EXPENDITURES One of the important ways in which the federal government affects state and local expenditures is through the tax expenditures associated with the personal and corporate income taxes. These expenditures were estimated at approximately $100 billion in 2011. For instance, interest on state and local bonds is not subject to federal taxation, and state and local income and property taxes are deductible from individual federal income taxes. As we shall see, this not only can be thought of as a subsidy to states and localities, but also provides an incentive for greater expenditures at the state and local level.

THE SIZE OF FINANCIAL TRANSFERS

Chapter 2 emphasized that the magnitude of governmental expenditures does not provide a complete picture of government's role in the economy. Similarly, the magnitude of federal transfers to states and localities does not show the extent to which state and local government expenditures are affected by federal activities. Still, several features of these transfers are worth noting. First, they grew immensely between 1929, when they amounted to just 1.3 percent of state and local government revenues, and 1970, when they amounted to one-fifth. There were some fluctuations over the next three decades, but their share had risen to one-fourth by 2005 and stands at 28 percent in the aftermath of the Great Recession. As a share of total federal expenditures, grants to state and local government have risen from 12 percent in 1970 to 17 percent today. Second, federal aid appears to be more important at the state level than at the local level, accounting for slightly more than one-fourth of state revenue but just over 3 percent of local revenue.[2]

These figures are deceptive, however. Much of the money granted to states is passed through to local governments. Transfers from state to local governments account for about one-third of total local revenue.

---

[2] Executive Office of the President, Office of Management and Budget, *Budget of the United States Government, Fiscal Year 2013, Analytical Perspectives* (Washington, DC: Government Printing Office), Table 18-2; and U.S. Census Bureau, "State and Local Government Finances and Employment," *2012 Statistical Abstract*, Tables 452 and 455.

UNFUNDED MANDATES

By 1994, the issue of unfunded mandates had been festering for a long time. Even pieces of legislation for which there was widespread support imposed financial burdens on states and localities that they found difficult to meet. For instance, the Americans with Disabilities Act required that states and municipalities ensure access to public buildings and public transportation by the handicapped. The reauthorization of the Clean Water Act brought the issue to a head. Communities realized that they would have to pay millions for new sewage treatment facilities.

Critics of unfunded mandates argued that just as Congress had committed itself not to pass additional expenditures unless it could finance them, so too it should commit itself not to pass additional mandates on states and communities unless it paid for them. The unfunded mandates issue was also seized upon by those who wanted to cut back government activities in general. They saw it as one way of ensuring, for instance, that no new environmental legislation could be passed. It was precisely this aspect that made the unfunded mandates issue so alarming.

Economists pointed out that some of these "mandates," such as those involved in the Clean Water Act, did not ask localities to assume an "unfair" burden; they mandated only that localities not impose an externality on other communities. By failing to treat sewage adequately, some communities were imposing costs on other communities.

The compromise adopted by Congress in 1995, the Unfunded Mandates Reform Act (UMRA), required that legislation that imposed costs on communities be accompanied with an estimate of the magnitude of those costs. Then, at least, Congress could judge whether those costs were reasonable, and whether the benefits of the legislation commensurate with those costs.

UMRA is still in effect today, and the more than 10,000 cost estimates of intergovernmental mandates submitted to Congress by the Congressional Budget Office since January 1996 have had a modest positive impact by making the costs of proposed federal mandates transparent.

SOURCE: Robert Jay Dilger and Richard S. Beth, *Unfunded Mandates Reform Act: History, Impact, and Issues* (Washington, DC: Congressional Research Service, June 2014).

Similarly, 64 percent of federal money transferred to state and local governments is passed on to individuals, primarily through Medicaid and income security expenditures. Medicaid reimbursements account for a growing proportion of federal aid to states. In 1973, they accounted for 11 percent of federal grants; in 2011, Medicaid accounted for 45 percent of transfers to state and local governments.[3]

Even these Medicaid statistics are misleading. States have learned how to use the Medicaid program as a form of general revenue sharing. To see how the "scam" works, assume some hospital increases its charges

---

[3]Executive Office of the President, Office of Management and Budget, *Budget of the United States Government, Fiscal Year 2013, Analytical Perspectives* (Washington, DC: Government Printing Office), Table 18-1; and U.S. Census Bureau, "State and Local Government Finances and Employment," *2012 Statistical Abstract*, Tables 454 and 455.

by $1000, with the state picking up $500 and the federal government picking up $500. The hospital could then rebate, say, $800 to the state. The hospital is better off (it has $200 more to spend), the state is better off (it has $300 to spend), and only the federal government is worse off. Expenditures for health care for the poor go up, but not because services go up or because more resources are devoted to the poor.

Financial transfers are also very important in developing economies, averaging over 40 percent of subnational expenditures and reaching twice that share in some countries. The purpose of these transfers is to close a fiscal gap created by central government assignment of more expenditure responsibility than subnational governments can pay for: subnational expenditures average about one-fourth of total public sector expenditures in developing economies, but subnational revenue averages only one-sixth of total public sector revenue. These central government transfers are used primarily to finance the key subnational responsibilities of education, health, and general government services.[4]

PRINCIPLES OF FISCAL FEDERALISM

In the previous section, we described the various activities that are undertaken at each level of government: the division of responsibilities. To a large extent, this division of responsibilities evolved over time. The Constitution, which set forth the framework within which the division of responsibilities occurs, was written more than 200 years ago, well before the development of the modern theory of public finance, and before notions of public goods even existed.

With the development of the modern theory of public finance, however, we can ask: What principles should guide the assignment of responsibilities? Are some assignments more likely to lead to efficiency, or to decisions about the level or kind of public goods being produced that are more in accord with the preferences of citizens? This section sets forth some of the key principles.

NATIONAL PUBLIC GOODS VERSUS LOCAL PUBLIC GOODS

For some kinds of goods, there is a strong presumption for federal provision. These are **national public goods**, whose benefits accrue to everyone

---

[4] A. Shah, "Fiscal Decentralization in Developing and Transition Economies," World Bank Policy Research Working Paper 3282, Washington, DC, April 2004.

in the nation. In contrast, the benefits of **local public goods** accrue to residents of a particular community. National defense is a national public good; traffic lights and fire protection are local public goods.

Just as most goods publicly provided at the national level are not pure public goods, most goods publicly provided at the local level are not pure public goods. For some goods, such as public libraries, exclusion is easy but undesirable, as the cost of providing access to an additional individual is almost zero. Some goods that local governments provide, such as education and public hospitals, are essentially private goods; exclusion is easy, and the costs of providing services to additional individuals are significant (see Chapter 5). By the same token, some local public goods are not purely local; some of the benefits accrue some of the time to those living in other communities.

The argument that if there is to be an efficient supply of public goods they must be provided publicly implies, by extension, that if there is to be an efficient supply of *national* public goods, they must be provided at the national level. If it were left up to each community to provide for national

INTERNATIONAL PUBLIC GOODS

Just as there are some public goods whose benefits accrue only to those living inside a particular community, there are some public goods whose benefits accrue to people living all over the world. Just as there will be an undersupply of national public goods if the provision is left to local communities, there is likely to be an undersupply of **international public goods** if the provision is left to national governments.

There are at least four important categories of such international public goods. The first, the global environment—the earth's atmosphere and the oceans that surround the continents—is perhaps the most obvious. Even though increases in greenhouse gases are likely to affect different countries differently, the overall concentration of greenhouse gases is the result of the cumulative actions of all the individual countries. The second, international security, can potentially affect almost everyone in the world, as it did during the two world wars of the twentieth century. The third is knowledge. The marginal cost of an additional person *anywhere in the world* having a bit of knowledge is zero—it does not subtract from what others know (although it may reduce the economic rents that they can obtain from the knowledge, and the marginal cost of *transmission* of knowledge is not zero). Furthermore, at least for many types of knowledge, exclusion is difficult, if not impossible. The fourth is international economic stability. An economic crisis in one country can spread to other countries, just as a disease in one person can infect others. (In fact, economists refer to the process as *contagion*.) Thus, maintaining international economic stability and containing the impact of crises is viewed as an international public good of first-order importance, and the international community has set up international financial institutions—the International Monetary Fund and the World Bank—to provide assistance in the event of a crisis.

LOCAL, NATIONAL, AND INTER-NATIONAL PUBLIC GOODS

- *Local public goods:* public goods whose benefits are limited to those living in a locality
- *National public goods:* public goods whose benefits accrue to everyone in the nation
- *International public goods:* public goods whose benefits are global in nature

public goods, there would be free rider problems, just as there would be if the provision of national public goods were left up to individual households and firms.

DO LOCAL COMMUNITIES PROVIDE LOCAL PUBLIC GOODS EFFICIENTLY?

Although there is a presumption that the federal government should provide national public goods, the question remains: Should the provision of local public goods be left up to states and localities?

In a remarkable article written almost sixty years ago, Charles Tiebout of the University of Washington argued that one could think in terms of local communities' competing with each other to supply local public goods to citizens—efficiently, in the quantities and forms they want—just as firms compete to supply conventional private goods.[5] He argued that just as competition among private firms leads to the efficient provision of private goods, so, too, competition among local communities leads to efficiency in the provision of local public goods. This hypothesis is called the **Tiebout hypothesis**. The following section explores the Tiebout hypothesis—including its limitations—in greater depth.

TIEBOUT HYPOTHESIS

Chapter 3 discussed the rationale for government activities. The fundamental theorem of welfare economics—Adam Smith's "invisible hand"—implies that in the absence of a market failure, such as public goods, the economy will be Pareto efficient. Individuals, all acting in their own self-interest, will make decisions that lead to Pareto efficiency. Competition among producers leads them to supply the goods individuals want at the lowest possible cost.

An analogous argument can be made for the provision of local public goods and services by state and local governments, as distinct from the federal government. Competition among communities, it is argued, will

[5] See C. Tiebout, "A Pure Theory of Local Expenditure," *Journal of Political Economy* 64 (1956): 416–424.

result in communities' supplying the goods and services individuals want and producing these goods in an efficient manner.

Tiebout was originally concerned with the problem of *preference revelation* discussed in Chapter 9: individuals reveal their preferences for private goods simply by buying goods, but how are they to reveal their preferences for public goods? When individuals vote, they choose candidates who reflect their overall values, but they cannot express in detail their views about particular categories of expenditures. Only limited use of referenda is made in most states. Even if individuals were asked to vote directly on expenditures for particular programs, the resulting equilibrium would not, in general, be Pareto efficient.

Tiebout argued that individuals could "vote with their feet"—their choice of communities revealed their preferences toward locally provided public goods in the same way as their choices of products revealed their preferences for private goods. Moreover, just as there are incentives for firms to find out what commodities individuals prefer and to produce those commodities efficiently, so are there incentives for communities to find out what kinds of community-provided goods individuals prefer, and to provide them efficiently. This is seen most strongly in the case of community developers. In recent years these developers, recognizing that many individuals would like more security and more communal facilities (swimming pools, tennis courts) than are provided by the typical city, have formed large developments providing these services. Because these communities meet the needs of the individuals better than the available alternatives, individuals are willing to pay higher rents (or spend more to purchase homes in these communities). This gives developers a return for their efforts to ascertain what individuals want and to meet these desires.

More generally, communities that provide the services individuals like and provide them efficiently will experience an influx of individuals; communities that fail to do so will experience an outflux. Such migration (with its consequent effect on property values) provides essentially the same kind of signal to city managers that the market provides to a firm's managers (a firm that fails to provide a commodity individuals like will find its sales declining; a firm that succeeds will find its sales increasing). Politicians, sometimes under pressure from the electorate, respond

TIEBOUT HYPOTHESIS

Competition among communities ensures efficiency in the supply of local public goods, just as competition among firms ensures efficiency in the supply of private goods.

Limitations:

- "Market failures"

    Externalities: decisions of community have effects on others

    Imperfect competition: limited number of communities

- Tax competition may simply lead to lower taxes on businesses

- Redistribution—with free migration and local competition, there will be no (or, at most, limited) redistribution at local level

to these signals in much the same way as a firm's managers respond to market signals.

The analogy is an instructive one. Under certain assumptions, the separate decisions of each community lead to a Pareto efficient allocation, just as the separate decisions of firms and individuals concerning private goods lead to Pareto efficiency.

But these assumptions generally do not hold. And even were they to hold, the inequality in the distribution of welfare across communities might be unacceptably large.

The qualifications to the Tiebout hypothesis closely parallel those we discussed in Part 2, concerning the circumstances in which market allocations might not be Pareto efficient or, even if efficient, might not be desirable.[6] The two most fundamental qualifications are the presence of market failures and dissatisfaction with the distribution of income.

MARKET FAILURES

The most important "market failures" have to do with externalities and imperfect competition.

EXTERNALITIES The actions of one community may have marked effects on other communities. If a community constructs a smelly sewage plant or allows the development of an industrial area at its boundary, in a location such that the winds blow the bad odors over the neighboring communities, an important externality results. We sometimes refer to these externalities as *spillovers*. Not all spillovers are negative. Some economists believe that there are important public benefits from having an educated citizenry, and that they provide some justification for public support of education. To the extent that this is true, and to the extent that individuals move away from the community that provided them with a free education, there are spillovers from a local community's public education system.

Migration and location inefficiencies may be thought of as a particularly important class of externalities. Individuals who move into a community bring both benefits and costs: they may increase the tax base, but also

---

[6] Since Tiebout, an extensive literature has developed evaluating the conditions under which the result is valid. See, in particular, J. E. Stiglitz, "Public Goods in Open Economies with Heterogeneous Individuals," in *Locational Analysis of Public Facilities,* ed. J. F. Thisse and H. G. Zoller (New York: Elsevier–North Holland, 1983), pp. 55–78; J. E. Stiglitz, "Theory of Local Public Goods," in *The Economics of Public Services,* ed. M. Feldstein and R. Inman (New York: Macmillan, 1977), pp. 274–333; T. Bewley, "A Critique of Tiebout's Theory of Local Public Expenditures," *Econometrica* 49 (1981): 713–740; and G. R. Zodrow and P. Mieszkowski, "Pigou, Tiebout, Property Taxation, and the Underprovision of Local Public Goods," *Journal of Urban Economics* 19 (1986): 356–370.

may lead to increased demands on public services and greater congestion, for instance, of roads and parks. Because in many cases they neither pay for these costs nor are compensated for the benefits they confer, there are likely to be inefficiencies in location decisions. Many countries have become increasingly concerned about what they view as excessive concentration of population in the major cities (London, Paris, Mexico City), and have developed decentralization policies to attain what they consider a more efficient pattern of location.

COMPETITION AND PROFIT MAXIMIZATION   A central assumption underlying the results concerning the efficiency of market economies is that there are many profit-maximizing firms. The Tiebout hypothesis similarly assumes the existence of many competing communities.[7] In most areas, there is only a limited number of competing communities; there is, in effect, only limited competition. Moreover, communities do not decide which goods and services to provide on the basis of any simple profit-maximization criterion, but by a political process along the lines discussed in Chapter 9. The kinds of inefficiencies to which this may give rise will be described in the next chapter. Here, we simply note that limited competition provides an explanation for why we should be skeptical about the Tiebout hypothesis.

TAX COMPETITION   Tiebout's model suggests that competition among communities is not only healthy, but also necessary to attain Pareto optimality. However, there is another view of competition among communities that is far more negative. This view sees different communities competing to attract businesses, with the associated tax base and employment opportunities. Gains in one community are partly at the expense of losses in other communities. More generally, the competition to attract businesses results in lower taxes for businesses: in the end, businesses are the ultimate beneficiaries. In this perspective, it would appear preferable for communities to agree not to compete.

Earlier, we pointed out that the incidence of taxes imposed by local communities had to be on immobile factors, as the mobile factors could move to escape taxation. Capital (and businesses, more generally) is mobile; the competition to attract businesses through tax concessions is just a reflection of this reality. If communities agreed to give no tax concessions, competition would almost surely take other, probably more wasteful, forms,

---

[7] Indeed, there must be so many that all residents within each community who have the same skills also have the same tastes for public goods. Another implication is that (provided voters are rational) there would be complete unanimity in voting. Obviously, neither of these conditions is satisfied. See R. W. Eberts and T. J. Gronberg, "Jurisdictional Homogeneity and the Tiebout Hypothesis," *Journal of Urban Economics* 10 (1981): 227–239; and H. Pack and J. Pack, "Metropolitan Fragmentation and Local Public Expenditure," *National Tax Journal* 31 (1978): 349–362.

such as providing enhanced public facilities like roads for the businesses. (If even this were somehow stopped, communities that had higher tax rates because they had a lower tax base would find it impossible to attract businesses.) From this perspective, then, trying to stop **tax competition** is like trying to stop competition elsewhere in the economy. Not only are such attempts likely to be ineffective, but to the extent they are effective, they are also likely to lead to other problems, including inefficiencies.

Nonetheless, an increasingly popular alternative to this "race to the bottom" characteristic of escalating wars of tax incentives between dueling local governments is **economic gardening**: nurturing local businesses and making investments to support both existing and new businesses. The premise of economic gardening is that tax competition is usually expensive and counterproductive—it either fails to attract businesses, or attracts businesses that would have come without tax incentives because of market-related factors and the quality of local public infrastructure and services. Moreover, proponents of economic gardening claim that tax incentives deprive local government of the resources needed to serve all businesses in its jurisdiction, and predicate future success on maintaining a subsidized cost advantage that will be increasingly difficult to maintain.

As in other areas, complete collusion, were it successful, could have real effects. If all communities were able to agree not to compete for business, and agreed, say, to impose a uniform tax on business, then the communities would gain at the expense of businesses. Such a tax would be equivalent to a federally imposed tax.[8] The debate on tax competition illustrates the marked advantages the federal government has in imposing taxes.

REDISTRIBUTION

Redistribution—the second basic qualification to the Tiebout hypothesis—may be a more important explanation of the role of the federal government than the market failures we have just described. There is concern about the distribution of income both among individuals and across communities.

INEQUALITY AMONG INDIVIDUALS Should the extent of redistribution—the level of welfare payments—be a local or national decision? Is "redistribution" a local public good? Assume that individuals in some community believe strongly that no individual should live in a slum, and so they provide a good public housing program, whereas individuals in some other community have different ethical concerns. Is there any

---

[8] There remains the problem of tax competition among countries.

reason why individuals in the first community should attempt to impose their ethical beliefs on the second, by attempting to make minimal housing standards a national rather than a local issue?

The answer is yes. The reason is that, with relatively free migration, the extent of redistribution that is feasible at the local level is very limited. Any community that decides to provide better housing for the poor, or better medical care, might find itself faced with an influx of the poor. Communities have an incentive to try to appear unattractive to the poor, so they will move on to the next community. Some communities, for instance, do this by passing zoning laws that require multiacre lots.[9] Others do it by limiting the provision of certain public services that are particularly valued by the poor and for which the wealthier have good private substitutes, such as bus services.

Indeed, if there were perfect competition among communities, the efforts to provide local public services at least cost to the taxpayers would result in taxpayers' paying taxes only commensurate with the benefits they themselves received. A community that had no welfare program and succeeded in excluding most of the poor would be able to provide public services (e.g., education, sewage treatment, libraries, etc.) at lower tax rates than a community that had an ambitious welfare program (e.g., public housing, good medical care, etc.) and educational programs aimed at disadvantaged children. The fact that competition is frequently limited, migration is slow, and decisions concerning public services are made politically means that there often are local (and state) redistribution programs—but these remain limited.

INEQUALITY ACROSS COMMUNITIES   We have already noted that there are marked differences among the states in per capita income. For a poor community to provide the same level of services as a rich one requires that it levy much higher tax rates. Indeed, we see enormous variation in per capita expenditures and tax rates across the United States. Per capita expenditures at the state level in 2008 varied from lows of $4079 in Texas, $4146 in Nevada, and $4190 in Florida, to highs of $14,701 in Alaska, $9534 in Wyoming, and $8182 in Hawaii, with an average for the country of $5696.[10] Total state and local taxes paid in 2008 by a typical family of four earning $75,000 ranged from the high end at 12.4 percent of income in Philadelphia, 11.2 percent in Detroit, and 10.9 percent in New York City, to the low end at 4.4 percent of income in Jacksonville, 5.5 percent in Las Vegas, and 5.6 percent in Houston.[11]

---

[9] Courts have recently restricted the use of zoning as an exclusionary device.
[10] U.S. Census Bureau, "State and Local Government Finances and Employment," *2012 Statistical Abstract*, Tables T-13 and 454.
[11] U.S. Census Bureau, "State and Local Government Finances and Employment," *2012 Statistical Abstract*, Table 447.

Why should we be more concerned with the inequality associated with locally provided public goods (and tax rates) than we are with inequality in general? Is there any reason why there should be specific federal programs directed at reducing this particular kind of inequality? If we want more redistribution, why not simply impose a more progressive federal tax, letting individuals then choose how to spend their money? If they wish to live in communities that spend more or less on local public goods, why not let them? The issues are analogous to those that arose in earlier chapters concerning whether the government should have specific policies directed at decreasing the extent of inequality of access to specific goods, such as medicine, food, and housing. The concept of *specific egalitarianism* was introduced—the view that the consumption of certain commodities should not depend on one's (or one's parents') income or wealth. Education, the most important locally and publicly provided good, is one of those goods for which the strongest argument for equality of access can be made.

Several arguments, however, can be made against providing programs aimed at reducing inequality in the provision of local public services.

1. *Consumer sovereignty*. The first is the standard "consumer sovereignty" argument: individuals should be allowed to choose the goods they prefer. The argument is that the federal government should not force its preferences—for food, housing, or education—on local communities. Programs aimed at reducing inequality in the provision of local public goods (to the extent that they are effective) distort consumption patterns; they may result in greater consumption of "local public goods" and less consumption of private goods than a redistributive program providing cash to individuals. Categorical grants (again, to the extent that they are effective) cause a distortion in the mix of locally provided goods; they may, for instance, result in more education and urban redevelopment and less frequent sewage collection. Whenever there are such distortions, there is a deadweight loss.

   This consumer sovereignty argument, though relevant, is somewhat less forceful for some locally provided goods than for others. For instance, decisions concerning elementary and secondary school education are made not by the individual but by his or her parents; and decisions concerning local public goods are made by a political process, which need not yield efficient outcomes, as we saw in Chapter 9.

   Moreover, much of the consumer sovereignty argument is predicated on the belief that individuals are well informed and rational, but in many cases, these assumptions are not valid. Parents seldom have information about the quality of schools and typically send their children to the nearest school, even when they have a choice.

2. *The difficulty of targeting communities for redistribution.* A second argument against programs aimed at redistributing income across communities (localities, states) is that such programs are not well targeted. Most communities contain a mix of poor and rich individuals. A program aimed at redistributing resources to a community whose average income is low may simply result in a lowering of the tax rate; the program's main beneficiaries will thus be the rich individuals within the poor communities. On the other hand, certain specific programs, such as the school lunch program, may be more effective in redistributing income to *children* than programs aimed at redistributing income among families.

3. *Location inefficiencies.* A third argument is that programs redistributing income across communities result in location inefficiencies. They distort the decisions of individuals about where to live and the decisions of businesses about where to locate.

The United States is a very mobile society; we move often, and frequently quite far. There have been large migrations from the rural South to the urban North, and from the Snow Belt to the Sun Belt. A variety of reasons induce individuals to move, but economic considerations are among the more important. These include not only individuals' opportunities for employment and the wages they receive, but also the taxes that are imposed and the public goods that are provided. As demands and technologies change, economic efficiency requires that individuals move to where they can be more productive. This will necessitate that some localities, and indeed even some regions, experience declining populations, whereas others experience rapidly rising populations. Federal aid aimed at redistributing income from one locality to another may interfere with the efficient allocation of labor and capital. The level of taxes and public services provided by one community will not correctly reflect the economic potential of that community. The inefficiencies to which this gives rise may be small in the short run but may become large in the long run. Individuals will be encouraged to stay where they are rather than move to more productive localities. Indeed, it might be better to use the same funds to subsidize emigration out of the unproductive areas. Similarly, with new highway systems, it may no longer be efficient to have the larger agglomerations of population associated with inner cities. Thus, aid to central cities may serve to perpetuate these inefficient patterns of location.[12]

Note that these inefficiencies arise from attempts to redistribute income among communities. If our basic concern is with inequality

---

[12] On the other hand, the aid may compensate for positive externalities produced by the inner cities.

among individuals, redistribution should be aimed at individuals, not at regions or localities.

In addition, specific redistributive programs, if they are not well designed, may give rise to large distortions. A program aimed at remedying measured housing shortages among the very poor, by providing federal subsidies, may encourage communities to undertake actions that exacerbate these shortages (such as rent control). Critics worry that a program to bail out cities that have borrowed excessively and appear to be in danger of defaulting on their bonds may encourage other communities to borrow more than they otherwise would, knowing that if they get into trouble the federal government is there to rescue them.

However, there is little evidence that such adverse incentives (sometimes called *moral hazard*) have played an important role. Some of the most widely noted instances of communities that have gotten into financial troubles are connected with exploitation by Wall Street, in its selling of inappropriate financial products. Others are connected with declines in America's conventional manufacturing—its deindustrialization—with particular adverse effects on communities in which such industries have been central. These financial problems have been exacerbated by the fractionalization of metropolitan areas, which has left many inner cities in very bad financial straits even as their surrounding suburbs prosper.

An important consideration in ascertaining appropriate policies for cities in distress is the consequences for "innocent" victims (e.g., children who will get an inadequate education unless something is done to provide these cities with some assistance).

OTHER ARGUMENTS FOR LOCAL PROVISION

The concept of local public goods provides the central argument for the provision of certain public goods locally, but several other arguments have been put forward for assigning greater responsibility for the provision of collective goods to the local level—even when the goods are not pure local public goods, or even when doing so may limit the scope for redistribution. These arguments have played an important role in recent political debates in the United States. One is that by delegating more responsibility to local communities, there can be more adaptation to the circumstances and preferences of those who benefit from the good. Moreover, there is more likely to be active involvement of citizens—for example, in schools—when they are the responsibility of local communities; this involvement leads to higher-quality public services.

Moreover, at the local level, individuals see more clearly the link between benefits and costs (what they have to pay in taxes); when people sense clearly the link between government benefits and taxes, they are less likely to ask for benefits that are not worth the costs, and are more likely to demand efficiency in the provision of benefits. Moreover, there can be more experimentation, which provides information that is particularly valuable in designing programs in areas such as welfare, where there is a general sense of major need for improvement.

Besides these analytic arguments, some may push for delegating more responsibility to states and localities because they believe that the political process will result in decisions that are different—and more to their liking—than if the decisions are made at the national level. For instance, many believe that assigning states and localities more of the responsibility for decisions concerning welfare programs will result in a more effective containment of costs.

However, local and state politics may play out in other ways as well. In many states, rural areas have a political weight that is out of proportion with their population. As a result, state politics can be much more conservative, biased against urban areas, and less concerned about issues of inequality. For instance, states were left with discretion in signing up for the expanded Medicaid provisions under the Affordable Care Act. The expansion was intended to reduce the number of low-income Americans who did not have access to health care. But even though the federal government was paying almost all of the costs, several states with the greatest problem of lack of coverage chose to opt out. Clearly, they did not have the interests of their poorer citizens in mind.

PRODUCTION VERSUS FINANCE

Many of the arguments typically made for local provision of public goods—that local communities are more responsive to the needs and preferences of those who actually receive the goods, that local communities have greater incentives for efficiency, and that devolving responsibility to local communities provides greater opportunities for experimentation—are mainly arguments for local *control* (local decision making) rather than local *finance*. However, there are good reasons for concern about separating finance from control. If voters of the country as a whole believe that their tax dollars should be used to finance welfare expenditures for the poor, they want to be sure that their money actually goes for this purpose, and not to finance suburban swimming pools. Some controls on expenditure are necessary. The issue is one of degree: how much control? By

imposing more controls, there may be greater assurance that the money is used in the way intended, but there is a cost—more bureaucracy, less adaptability to local circumstances, and less experimentation. In the case of the welfare program, prior to the 1996 reforms, there was a consensus that more local autonomy was needed, and the federal government granted the vast majority of states **waivers** of federal rules to allow them to introduce specific experiments. A similar debate is now under way regarding the appropriate degree of state discretion in the context of national health care reform.

When responsibility for decision making devolves downward from the federal government, there is a question of how far downward: to the states, to subunits of the states such as cities or counties, or directly to individuals? Many of the arguments for devolving responsibility suggest that the lower the level, the better. A housing program is more likely to be responsive to local needs if responsibility is given to the city or neighborhood, rather than to the state. Many argue, why involve intermediary levels of government at all? Why not simply give poor individuals housing vouchers—certificates that they can use to buy housing anywhere—giving them the decision-making responsibility over the kind of housing they want?

The discussions earlier in this chapter and in previous chapters have provided a number of reasons for not relying on vouchers or cash transfers. Promoting a society that is concerned about children and providing good public schools for all may be a more effective way of ensuring quality education and educational opportunity than giving money to parents and hoping that they will make the best decision for their children. We know that private markets often engage in exploitation, taking advantage of imperfect and incomplete information. There are hosts of other market failures that government provision may help address.

Assigning responsibility for decision making to local communities does not mean that they actually have to do the production themselves. Just as the federal government can produce goods and services directly or purchase them from private firms, the same is true at the local level. Typically, local communities are involved in the production of most of the goods and services that they provide—from police and fire protection to schools. However, there are some areas—most notably, garbage collection—that many communities contract out to private providers. (In still other localities, garbage collection is treated as a private good, with the community taking on no role at all.) The discussion of public versus private production at the local level parallels that at the national level (see Chapter 8). More recently, states and localities have been exploring new possibilities for both contracting out and complete privatization—including performance of administrative

services, lease or sale of public infrastructure (especially buildings and toll roads), charter universities and charter schools, development agencies, prisons (see case study, "Privatizing Prisons" in Chapter 8), and even state lotteries.

EFFECTIVENESS OF FEDERAL CATEGORICAL AID TO LOCAL COMMUNITIES

The intention of federal categorical aid to local communities is to encourage local spending on particular public services. Aid to bilingual education, to vocational education, and to school libraries is intended to result in an increase in expenditures in each of these categories. How effective is this aid? Do federal funds just substitute for local funds, or do they actually result in more expenditures for the intended purpose?

From a theoretical perspective, the issue is precisely the same as one that we discussed in Chapter 10. How effective is categorical aid to individuals in encouraging expenditures, say, on food or housing? The answer depends on whether there is a substitution effect or just an income effect.

We wish to compare three types of federal aid to local communities—a block grant not tied to any specific use, a block grant tied to a specific purpose, and matching aid for a specific purpose.

Figure 26.4 shows the budget constraint of the community. (We simplify by assuming all individuals within the community are identical, so that we can ignore questions concerning differences in tastes.) The community would choose point E, the tangency between the budget constraint and the indifference curve of the representative individual. Now assume that the federal government provides a block grant to the community. This shifts out the budget constraint, to line $B'B'$. There is now a new equilibrium, E^*. It entails a higher level of expenditure on local publicly provided goods and a higher level of per capita consumption of private goods. That is, the federal aid has, in fact, resulted in lowering the tax rate imposed on individuals. The federal money has partially substituted for local community money; the community, because it is better off, spends more on publicly provided goods as well as privately provided goods.

Now assume, however, that there are two different publicly provided goods, garbage collection and education, on which the community can spend funds. We represent the allocation decision of the community between the two goods by the same kind of diagrammatic devices we have used to represent the allocation between private and publicly

FIGURE 26.4

EFFECT OF BLOCK GRANTS

A lump-sum transfer to a community will result in an increase in public expenditures, but by an amount less than the transfer; local taxes will go down.
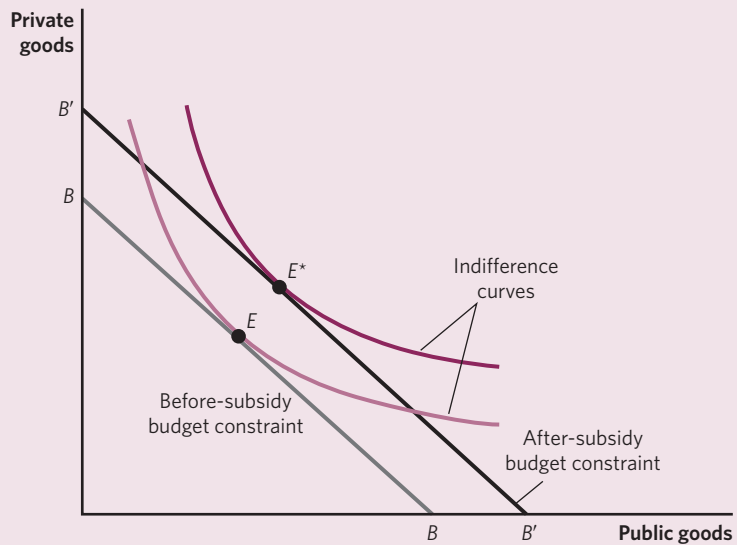
provided goods.[13] The community has a budget constraint; it needs to divide its total budget between the two goods, as represented by Figure 26.5. The community also has indifference curves between the two goods. The initial equilibrium is represented in Figure 26.4 by E. Now with the federal aid, the budget constraint has moved out, and the new equilibrium is E^*. Does it make any difference whether the government specifies that the funds be allocated to one public good or the other? Not usually. As long as the amount of federal aid that is tied to a good is less than the amount that the community wishes to spend on it, federal aid will substitute for local support for this particular good, on an almost dollar-for-dollar basis. That is, if the community spends, say, 5 percent of any additional increase in its wealth on education and 5 percent on garbage collection, a federal grant of $1 million will result in $50,000 additional expenditure on education and $50,000 on garbage collection. The remaining $900,000 will be used to lower the tax rate. However, it makes no difference whether or not the government stipulates that the money it gives be used for education, as long as the community was previously spending more than $1 million on education. If it were not spending this amount, then, of course, there would be greater effect on

---

[13] This kind of analysis assumes that we can separate the allocation decision among publicly provided goods from the allocation decision between private and public goods. This kind of separation is possible only under a fairly stringent mathematical condition on preferences known as *separability*, in which we assume that the marginal rate of substitution between public goods 1 and 2 does not depend on the level of consumption of other goods.

its education budget; expenditure would increase by the amount that the federal aid exceeded the amount previously expended.[14]

These results for block grant categorical aid need to be contrasted with a government program of *matching* local expenditures, for instance, on libraries. Suppose the federal government matches local expenditures on a dollar-for-dollar basis. If the local community wishes to buy a book that costs $10, it costs the community only $5, as the federal government provides the other $5 with a matching grant. This arrangement obviously creates a considerable inducement to spend more on these services, as illustrated in Figure 26.6. The new budget constraint, with the subsidy for local government expenditures, is rotated around point *B*. If the community were to decide to spend nothing, it would not receive federal aid. For every dollar of privately provided goods that the community gives up, it can obtain twice as many publicly provided goods as previously. Thus, the budget constraint is much flatter. This outward shift in the budget constraint has an income effect as before, but now there is, in addition, a substitution effect. Because publicly provided goods are less expensive, the community will wish to spend more. The equilibrium will change from *E* to *E**.

Figure 26.6 also shows the community's budget constraint with a block grant that provides the community with the same welfare as the

---

[14] A full analysis of this problem requires a three-dimensional diagram with education, garbage collection, and private goods on the three axes.
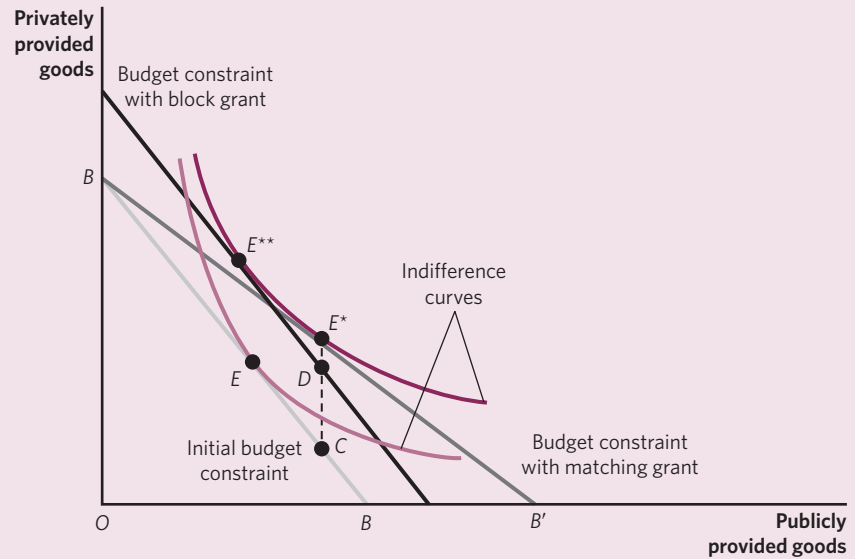
matching grant. (This budget constraint is clearly parallel to the before-subsidy budget constraint, and the new equilibrium at E^{**} is on the same indifference curve as E^*.) Two things should be noted: the equilibrium level of public expenditure on the public good is lower than with the matching grant, and the cost to the federal government is lower. There is a deadweight loss associated with the matching grant (of DE^*, in terms of privately provided goods).

If the matching funds are provided for a particular good, the federal aid will have a marked effect on the composition of the community's budget. It will encourage goods whose prices are lowered (perhaps partly at the expense of other publicly provided goods, whose *relative* prices can now be viewed as being higher). By the same token, it should be clear that for any given level of federal grants, if the object of the federal government is to encourage the provision of particular goods, a system of matching grants is far more effective than block grants—a lump-sum subsidy—whether restricted or not.

THEORY AND PRACTICE Evidence from actual government behavior supports our prediction that matching grants are more stimulative for local governmental spending than block grants. However, it does not support our prediction that nonmatching grants for specific purposes

have the same effects as a lump-sum increase in private income; the evidence suggests that categorical programs do have an effect on government budgets.[15] This has been referred to as the **flypaper effect**: money sticks where it hits.[16] Several explanations have been offered. One argument is that voters do not perceive the true marginal price of public expenditures when nonmatching grants are made; marginal costs exceed average costs, and voters are more aware of the latter than the former. Another explanation is that, at least in the short run, government bureaucrats have considerable discretion over their budgets. If they receive additional funds, the voters do not immediately know about it; and even if they did, they do not have the means to force the bureaucrats to pass the money back to them. A third argument posits that federal administrators can ensure that the money is spent in an incremental manner; they have enough discretion to withdraw funds if they believe that the federal funds are simply being used to substitute for state funds. This argument is supported by the "maintenance of effort" requirement of many federal grant programs, whereby grant recipients must spend their state and local funds at pre-grant levels for grant-funded activities, so that federal dollars supplement normal activities rather than supplant them.

> ## FEDERAL AID TO COMMUNITIES
>
> *Block categorical grants*
>
> - Fixed amounts for certain categories of expenditures
> - Effect much like a lump sum grant, provided the grant is less than would otherwise have been spent
>
> *Matching grants*
>
> - Amount received by state or locality depends on level of expenditure
> - Has both substitution and income effects—and therefore likely larger effect than comparable size block grant
> - Has distortionary effect—same level of community welfare can be attained at lower cost with a block grant

THE FEDERAL TAX SYSTEM AND LOCAL EXPENDITURES

The federal government affects local expenditures not only directly through its aid programs, but also indirectly through the federal tax system. Two provisions of the income tax code have an important effect on local communities. The first is that interest on state and local bonds is completely exempt from taxation by the federal government. This means that if an individual faces a 35 percent marginal tax rate, a 6.5 percent return

---

[15] See E. M. Gramlich, "Intergovernmental Grants: A Review of the Empirical Literature," in *The Political Economy of Fiscal Federalism*, ed. W. E. Oates (Lexington, MA: Lexington Books, 1977), pp. 219–239.
[16] P. N. Courant, E. M. Gramlich, and D. L. Rubinfeld, "The Stimulative Effect of Intergovernmental Grants: Or Why Money Sticks Where It Hits," in *Fiscal Federalism and Grants-in-Aid,* ed. P. Mieszkowski and W. Oakland (Washington, DC: Urban Institute, 1979), pp. 5–21.

on a tax-exempt local government bond is equivalent to a 10 percent return on any other bond. After taxes, a 10 percent return yields 6.5 percent: 10% $(1 - 0.35) = 6.5\%$. This tax exemption for state and local bonds clearly lowers the cost to state and local authorities of borrowing funds.

The second provision is that state and local income and property taxes are deductible from the federal income tax. That is, if an individual has an income of $40,000 and pays $1000 in property taxes, he or she can deduct that amount from income, and pay taxes on only $39,000. This means that if the individual is in the 35 percent tax bracket, the property tax reduces net income (what the individual can spend to buy consumption goods) by only $650. Of the $1000 in property taxes, the federal government is, effectively, paying more than one-third.

These tax benefits increase the level of expenditure on local public goods, encourage expenditures on capital projects, and induce some communities to finance their investments by debt.

Consider an idealized community in which all individuals are in the 35 percent tax bracket. If the community increases expenditures per family on education by $1000 and raises income and property taxes to finance the increased expenditures, the after-federal income tax cost to the individual is only $650. It is as if there is a federal matching grant for local public goods. The budget constraint facing the individual is identical to that depicted in Figure 26.6.

In most states, communities can borrow only to finance capital projects. If this restriction is binding (as it frequently is), the tax exemption of interest on local bonds implies that the effective cost of capital goods is lowered relative to that of current services (labor and materials); this results in a bias toward capital projects.

INEFFICIENCY OF TAX BENEFITS TO LOCAL COMMUNITIES   There are four reasons why providing aid to local communities through the federal income tax system may be inefficient. We have just discussed the first: aid provides a large incentive for the public provision of goods, regardless of the efficiency with which the local communities are able to deliver these goods and services. The second reason we discussed in Chapter 21: a significant fraction of the benefits of interest exemption accrue not to the communities, but to wealthy taxpayers.

The third reason why tax exemption may not be an efficient way of subsidizing local communities is that because of competition among communities, some of the benefits may accrue to industries within the communities rather than to the communities themselves. Local communities can issue tax-exempt bonds to help finance some of the capital costs required to provide the infrastructure to attract firms. If one community does this,

however, other communities respond, either by trying to attract firms to their community or by trying to prevent firms from leaving. The net effect is that the level of public goods provided to businesses may be higher than it would be otherwise. If only one community provided the higher level of public goods, it would be reflected in the price firms were willing to pay for land in that community. When all communities increase the level of public goods they provide, however, it may leave the total demand for land, and hence the level of rents, relatively unaffected.

The fourth consideration in an evaluation of federal tax and interest provisions is the inequities they create for individuals with different tastes and incomes. We have already noted that these provisions represent a considerable subsidy to the public provision of goods. Individuals who have a relatively strong preference for goods that tend to be publicly provided at the local level benefit by such measures, at the expense of those who have a weak preference for those commodities.

Because the magnitude of the reduction in effective costs of publicly provided goods depends on individuals' marginal tax rates, those who face a higher tax rate—usually wealthier individuals—receive a larger subsidy, and a larger reduction in their effective price of publicly provided goods. (To some extent, the "taste" effect and the pure income effect offset each other. Although they may receive a larger subsidy for each dollar spent by their local government, communities with wealthy individuals may actually spend less on at least certain categories of goods: for example, wealthier individuals are more likely to send their children to private schools.)

As persuasive as these arguments are to many economists, the political support for deductibility of state and local income and property taxes and preferential treatment of interest on state and local bonds remains strong enough that major changes in these deductions are unlikely in the foreseeable future.

CONCLUDING REMARKS

Some of the arguments we have given against federal assistance to communities, as opposed to individuals, may be overstated. We observed earlier that those who provide money at the federal level often have in mind specific uses for the funds. Much of the national consensus about government programs centers around access to certain basic goods and services, especially for certain groups in the population, wherever they are located: no child should go hungry, every child should have access to education (no child should be left behind), or no elderly should be left in destitution.

Although states have managed to assert their rights to administer many of these programs—and in some cases, because of their closeness to the situation, they might be able to better administer them—there remains a strong federal interest in insuring that the money is spent in the manner intended. That is why block grants, with no restriction, are limited.

But the same reasoning explains why money is given to states and localities rather than to individuals. Giving money to communities for specific purposes, for instance, to ensure that children are protected from the worst consequences of poverty, may be a far better way of achieving these objectives than giving the money to parents. Also, the goods and services for which there is a concern about access, like education and health, are often publicly supplied by local communities, especially to the poor. In some cases, like private, for-profit higher education, market provision has excelled more at identifying the poor individuals who can be exploited than at providing high-quality education at low costs. (Earlier chapters explained why many of these services have traditionally relied heavily on public provision.)

Moreover, the worry about adverse incentives is, for the most part, greatly exaggerated. Communities typically do not put themselves in the kind of dire straits that necessitate, or elicit, federal help. And the federal government often imposes "maintenance of effort" requirements—communities or states only get assistance if they have maintained certain efforts to provide benefits to their citizens.

In a dynamic economy, some communities will suffer as the demand for the products produced by their industries wanes or as patterns of living change. But it may be counterproductive to simply abandon these communities, as we have done with Detroit, MI, and Gary, IN. On their own, these communities have insufficient resources to restructure themselves, and they get set on a downward spiral: as people and businesses move out, their tax base shrinks; public services get cut back, inducing more outmigration. In these cases, federal assistance may help the communities "reinvent themselves."

Politics often play out differently at the local and state level than at the national level; local elites may have disproportionate influence. While it is often argued that governments that are closer to the people are more responsive to their needs, this does not always seem to be the case, even in democracies. The response to Obamacare provides a dramatic illustration. The poorest states—the states with the most poor people lacking health care coverage—decided not to accept the expanded Medicaid program, even though the federal government was paying 90 percent of the costs. These include states in which a century after the freeing of slaves, there was an active policy of disenfranchisement—for instance, denying or discouraging African Americans from voting with force. While the

civil rights bills improved matters, it did not lead to full participation of African Americans. Clearly, the decision to reject a virtually free transfer of medical costs to the federal government was not made in the interests of the large number of the poor and uninsured living in these states.

Generally, political processes in states give more political power to rural areas, than would be reflected simply in the proportion of the population that they represent. (This is the case, for instance, if one of the two chambers of the state legislature has representation by counties, and the urban population, though large, is concentrated in a few counties.) Since rural areas are often more conservative than urban areas, the political outcomes reflect these differences in beliefs and interests. At the national level, there is often a more active civil society based on coalitions of cities, and more liberal interests countervailing the conservative bias of state politics. Much of the debate about decentralization and devolution of power reflects these differences in political powers at different levels. Similar issues arise in other countries as well.

Moreover, there is often a difference of competency in administrative/ bureaucratic processes at different levels. It is not a surprise that many of the more talented people seek a bigger stage, making it easier to recruit good talent to be federal civil servants. The process can be self-reinforcing, since talented people often prefer to work with other talented people. And simply because of its larger size, the federal government can engage in more research and evaluation to assess on a more scientific basis its different programs in an attempt to improve them. (Such expenses are basically fixed costs; with larger programs, it is optimal to invest more to ensure their quality and effectiveness and promote evidence-based policies.)

All these observations provide an important caveat to the arguments concerning the benefits of decentralization, even in the case of purely local goods. But in a highly integrated economy with high mobility, there can be important externalities even from what otherwise would be purely local public goods. Less educated citizens may contribute less to the overall tax revenues and may be more likely to impose greater demands on public funds, such as for unemployment. Countries compete on the quality of their labor force and the goods produced, and a more poorly educated labor force may produce lower quality goods.

In short, the economics and politics of decentralization are complex. Various countries have experimented with different forms and degrees of decentralization, devolving various powers to lower administrative units, and giving different degrees of political (electoral) control at different levels with different financing arrangements. Some have employed the simple guidelines set forth earlier in this chapter, with considerable disappointment.

REVIEW AND PRACTICE

SUMMARY

1. The federal government regulates and subsidizes states and localities. It subsidizes them through both matching and block grants for specific purposes. In the past, it also provided general revenue sharing. In matching grants, the amount received by states and localities depends on the amount they spend. Indirect aid is provided by the exemption from taxation of interest on state and local bonds and the tax deductibility of state and local income and property taxes.

2. The arguments favoring local over federal provision of public goods are that local governments will be more responsive to the community's needs and preferences and have greater incentives to provide services efficiently. But differences in the way that state and local versus national politics play out provide a strong argument for national provision, especially in the context of growing inequality.

3. Local public goods are public goods whose benefits are limited to those living in a particular locality. The Tiebout hypothesis postulates that competition among communities results in an efficient provision of local public goods. The reasons why federal intervention may be required include market failures (externalities, particularly those associated with choice of location, and limited competition) and redistribution (the limited ability to redistribute income at the local level).

4. There are marked disparities in income per capita and in the provision of local public services across states and localities. Whether government policy should be directed at reducing inequalities across communities (rather than inequalities across individuals) is debatable.

5. Matching grants are more effective in encouraging expenditures in the direction desired, but there is a deadweight loss associated with their use. Although traditional theoretical arguments suggest that block grants, even for specific purposes, should have just income effects, and thus be equivalent to equal direct grants to the

members of the community, the empirical evidence suggests the presence of a flypaper effect.

6. Tax subsidies, including the tax exemption of interest on local and state bonds, lead to increased expenditures on publicly provided goods and increased capital investment by state and local governments.

7. Tax exemption of interest on state and local bonds and other forms of tax subsidies are an inefficient way of subsidizing state and local communities. Some of the benefit accrues to wealthy investors rather than to the communities, some of the benefit is passed along to businesses rather than to the residents of the communities, and the tax subsidies discriminate in favor of high-income individuals and individuals who have a strong preference for publicly provided goods.

KEY CONCEPTS

Block grant
Economic gardening
Fiscal federalism
Flypaper effect
General revenue sharing
International pubic goods
Local public goods
National pubic goods
Matching grants
Tax competition
Tiebout hypothesis
Waivers

QUESTIONS AND PROBLEMS

1. Discuss the advantages and disadvantages of state versus national determination of eligibility standards and benefits for food stamps, Medicaid, unemployment insurance, Temporary Assistance for Needy Families, and Old-Age and Survivors Insurance.

2. In President Reagan's State of the Union message in 1982, he proposed a trade with the states: in return for their taking over responsibility for the full costs of food stamps and Aid to Families with Dependent Children, the federal government would take over responsibility for Medicaid. In addition, he proposed phasing out most categorical grant programs, possibly substituting increases in block grants. The proposal was never adopted. Evaluate these proposals using the analysis of this chapter.

3. If the income elasticity of demand for education is 1, what will be the effect on expenditures on education of a small block grant of $100,000 if the community currently spends 5 percent of its total resources on education?

4. Many matching grant programs specify that the federal government matches, on a dollar-for-dollar basis, local expenditures up to some particular maximum. Draw the budget constraint between private goods and local public goods facing a community of identical individuals. Discuss the effect of such a matching program on communities that do not go to the maximum. Discuss the effect on communities that go beyond the maximum.

5. What would you expect to be the effects on spending on education if the federal income tax deduction for state and local taxes were eliminated? Show diagrammatically why you might expect such a change to increase the relative importance of private education.

6. Consider a community in which everyone is at the 36 percent marginal tax bracket. By how much would educational expenditures be reduced by the elimination of the tax deductibility of state and local taxes, if the price elasticity of demand for education is 1?

7. On the basis of the discussions in Chapters 6 and 10, discuss the relative merits of regulation versus matching grants as devices to elicit desired behavior on the part of state and local governments.

8. Recall from Chapter 9 the median voter theory. Consider a state that imposed a proportional income tax on everyone, but assume that the median voter did not itemize his or her deductions, so this individual's federal income tax payments did not depend on state taxes. How, according to the theory, would expenditures in such a state differ from those in a state with similar average incomes but in which the median voter did itemize his or her deductions, so that increased state and local taxes reduced the individual's federal tax payments?

9. The tax reform of 1986 eliminated tax deductibility of state sales taxes, but retained it for state income taxes. What implications should this have had for how states raised revenues? In fact, the share of individual and corporate income taxes in total state revenues today is lower than it was in 1980. How might you explain this?

10. Concern about *fungibility* of funds—of states not using money in the way intended—has led Congress in some instances to impose *maintenance-of-effort* requirements. Thus, in the 1996 welfare reform, to be eligible for a grant, states would have to continue to spend at least 75 percent of the amount that they had previously spent.[17]

    a. Show what this does to the budget constraint of the community.

    b. How effective are such restrictions likely to be over time, as incomes grow?

    c. How might a state attempt to get around this requirement by reclassifying expenditures?

11. In the past, the federal government provided general revenue sharing. The argument for general revenue sharing was that the federal government was in a better position to collect tax revenues. States may, however, simply impose a tax that is based on the individual's federal income tax; that is, they could impose a tax that is, say, 20 percent of the federal tax payment. Provided that the federal government shares its information about where individuals live and what taxes they have paid, there would then be little incremental cost of tax collection, either to individuals or the state. Why might the federal government nonetheless be in a better position to collect tax revenues than the states?

---

[17] If a state fails to meet work participation rates, the spending level rises to 80 percent. (Source: House Committee on Ways and Means, *Overview of Entitlement Programs, 1996 Green Book*, p. 1333.)

27

<div style="border:2px solid #8B2252; display:inline-block;">

SUBNATIONAL TAXES AND EXPENDITURES

</div>

Chapter 2 noted the changing pattern of taxation at the state and local levels: the decreased importance of property taxes and the increased importance of sales and income taxes. Does this changing pattern of financing result in a change in who bears the burden of state and local taxes, or only a change in the manner in which they are collected? The first part of this chapter considers the incidence of state and local taxes. The parallel issue of who benefits from the goods and services provided by local governments—the incidence of these expenditures—is then addressed. Finally, having analyzed these incidences, we discuss how the level and composition of public expenditures are determined locally.

TAX INCIDENCE APPLIED TO LOCAL PUBLIC FINANCE

In Chapter 18, we developed the basic principles of incidence analysis. There, we showed that the incidence of a tax on a commodity or a factor depends on the elasticity of demand and supply for that commodity or factor.

The limiting case occurs when the supply schedule has an infinite elasticity; that is, the supply schedule is completely horizontal. Then the incidence of the tax lies entirely on the buyer. The price the buyer pays goes up by the amount of the tax. The amount received by the seller is unaffected.

The implications of this for local taxes can be derived easily. In the long run, most factors are mobile; that is, they can move easily from community to community. This is particularly true of capital. Investors will invest in a community only if they can obtain at least the same return that they could obtain elsewhere. Mobility translates to elasticity; as a result, mobile factors will not bear the burden of local taxes.

LOCAL CAPITAL TAXES

A community that increases the taxes it imposes on capital will find that it attracts fewer investors. Even though it may not be possible for those with fixed capital equipment—such as steel mills—to remove their capital, new investments will be reduced until the before-tax return to capital is driven up. The process will continue until the after-tax return is equal to what it is elsewhere.

Thus, in the long run, the burden of the local tax on capital is not felt by the owners of capital—it is felt by the landowners and by the labor that remains. Because there is less capital, the productivity of land and labor (and hence their income) will be reduced.

If, as a result of the tax on capital, the productivity of workers is decreased, wages will be lowered. But then, in the long run, workers will emigrate; if labor is perfectly mobile, workers will continue to emigrate until their (after-tax) income is the same as it is elsewhere. This leaves land as the only factor that cannot emigrate. With less capital and less labor, the return to land is less: in the long run, the full burden of the tax is borne by landowners.

This assumes that labor is perfectly mobile. Of course, in the short run, workers will not instantaneously migrate in response to a small change in the wage rate. Indeed, many individuals have strong preferences for living in the community in which they grew up or in which they have formed strong ties. These laborers are only partially mobile. They will bear some part of the burden of the tax on capital. Their wages will be reduced as a result of the outflow of capital.

Dramatic consequences may result if these considerations are ignored. Occasionally, states have attempted to impose special taxes on particular industries. Some industries are especially "footloose." These industries will move out if higher taxes are imposed on them—they will shop for states and communities that offer the best deal.

PROPERTY TAX

Although the property tax is relatively small from a macroeconomic perspective, averaging in aggregate just over 2 percent of GDP in OECD countries, it is an extremely important source of discretionary revenue from a subnational level perspective: in the United States, the property tax accounts for 21 percent of combined state and local own-source general revenue, and 45 percent of just-local government own-source general revenue. In some states, such as Connecticut and New Hampshire, it accounts for more than 80 percent of local own-source general revenue.[1]

The property tax is generally a tax on land and capital—buildings and equipment. The incidence of the two parts is markedly different. Land is

---

[1] U.S. Census Bureau, "State and Local Government Finances and Employment," *2012 Statistical Abstract*, Tables 436 and 455. "Own source revenue" is total revenue generated from a state or local government's own sources, whereas "own source general revenue" is state or local revenue, excluding intergovernmental revenue, generated by state or local taxes and charges.

THE U.S. PROPERTY TAX REVOLT

In many states, voters have been worried about increased levels of expenditures at the state and local levels. They worry that politicians have strong incentives to initiate new policies (that is why they were elected). The politicians receive credit for additional programs, but the taxpayer bears the cost. The process of voting for legislators and city council members does not, in this perspective, provide an adequate check, as in most elections there is a host of other issues, such as abortion, the death penalty, and so forth, that frequently overshadow narrow budgetary issues. Accordingly, it is argued, voters must provide a direct limit on taxes and expenditures.

The question of imposing such ceilings was put on the ballot in California in 1978 as the People's Initiative to Limit Property Taxation, better known as Proposition 13. The timing could not have been better for such a ballot initiative, taking place during a spectacular real estate bubble: many property owners found that their tax bills were increasing at a much faster rate than their income as their property values skyrocketed with rising real estate prices. They were what has since come to be called "house poor." They had considerable wealth, but not the income (cash flow) to pay taxes on their houses. The property tax was forcing them to move. This seemed inequitable, particularly to the elderly. After Proposition 13 was approved by 64 percent of those voting, it became Article 13A of the Constitution of the State of California, which meant it could not be overridden by the state legislature. Proposition 13 capped the property tax at 1 percent of assessed market value, rolled back property values to their 1975 assessment, restricted increases in assessed value to a maximum of 2 percent per year, and prohibited reassessment of a new base year value except for new construction or change of ownership.

California's experience was subsequently replicated in many other states, such as Proposition 2½ in Massachusetts in 1980; within two years of the passage of Proposition 13, forty-three states had implemented some kind of property tax limitation or relief.

In the years since passage, the problems posed by Proposition 13 (and similar ballot initiatives elsewhere) have become increasing apparent. Although it addressed one inequity, it increased others. Owners of identical homes now frequently pay markedly different taxes: the owner of a house that was sold last year may pay ten or more times the tax paid by the owner of an identical house that has not been sold for fifteen years. Because some were paying only low taxes, the taxes on others had to be increased. These inequities were sufficiently great that many thought that fundamental constitutional issues were raised. However, in *Amador Valley Joint Union High School et al. v. State Board of Equalization et al.*, 22 Cal.3d 208 (1978), the Supreme Court of California sustained Proposition 13, and in *Nordlinger v. Hahn*, 505 U.S. 1 (1992), the U.S. Supreme Court declared Proposition 13 constitutional.*

Proposition 13 also induced some important economic inefficiencies. It significantly increased the cost of moving. An elderly couple living in a large house—well beyond the size required for their family—would decide to stay in the house; were they to move to a smaller house, their tax payments would actually go up. Thus, the housing stock was inefficiently allocated.

Finally, Proposition 13 contributed to a marked deterioration in the quality of public services in California. In some sense, this was its intention: to limit the growth of public expenditures. However, it was a blunt instrument. Proposition 13 also made it very difficult to enact new state and local taxes to replace lost property tax revenue, compelling local governments to resort to less efficient fees and charges to make up at least part of the shortfall—in 1977, property taxes contributed 24 percent and fees and charges contributed 21 percent to California city general revenues, but twenty-five years later, the shares were 17 and 36 percent, respectively.

*In 1989, the U.S. Supreme Court ruled in *Allegheny-Pittsburgh Coal Co. v. County Commission of Webster County, West Virginia*, 109 SCt. 633 (1989), that nonuniform assessment of property violates the equal protection clause of the Fourteenth Amendment. Many suspected that Proposition 13 was vulnerable on similar grounds, but it remains in effect today.

SOURCE: J. K. Rosengard, "The Tax Everyone Loves to Hate: Principles of Property Tax Reform," in *A Primer on the Property Tax,* ed. G. Cornia, W. McCluskey, and L. Walters, (Malden, MA: Wiley Blackwell, 2012); and papers presented at the symposium, "Tax and Expenditure Limitations: A Quarter Century after Proposition 13," *Public Budgeting & Finance* 24, no. 4 (Winter 2004).

inelastically supplied, so the incidence of the tax is on landowners; the tax is fully reflected in a decrease in land values. In the long run, the supply of buildings is elastic: investors can decide where to invest, and do so on the basis of after-tax returns, taking any property taxes into account. Thus, the before-tax return must increase. The property tax thus distorts investment decisions. If the tax were levied uniformly on all forms of capital, it would affect only the use of capital relative to other factors of production (such as labor). In practice, however, different kinds of capital are taxed differentially; some kinds of capital, such as working capital and the value of a firm's reputation, are essentially untaxed.

In the nineteenth century, the property tax got a terrible reputation for being administered in an arbitrary and capricious manner. For many types of property, well-developed markets do not exist and transactions occur infrequently enough that assessing precisely a property's current value is difficult. County assessors have the task of judging the market value. In the past, businesses (and other friends of the assessor) often succeeded

in persuading the assessor to give them low assessments. Today, although controversies over assessments remain, there are assessment processes (such as computer-assisted mass appraisal[2]) and taxpayer procedures (such as appeal of assessment and request for payment relief) that seem to have reduced inequities considerably. Assessment appeals are usually based on a comparison of assessments of properties with similar market value, something that is relatively easy to document today, as both government assessments and sales histories of properties are now available to the public. In the case of appeal for relief, the taxpayer accepts the tax assessment but claims not to have the means to pay the tax liability, a situation commonly referred to as being "asset rich but cash poor."

Failure to adjust the property tax to changes in economic circumstances has been blamed for contributing to urban blight. In inner-city areas, many buildings have been abandoned, as property owners claim that they do not generate enough revenue to pay taxes. However, this implies that the property tax exceeds 100 percent of the market value of the asset, or the property presumably would not be abandoned, a highly unlikely occurrence given that most annual property tax rates are between 1 and 2 percent of a property's market value. A much more significant factor is usually mortgages that are "under water," whereby the unpaid loan balance is greater than the property's current market value.

INCOME, WAGE, AND SALES TAXES

Similar principles apply to the taxation of labor. In a small community, if individuals had no particular attachment to the community, the long-run labor supply schedule would be perfectly elastic. A tax on labor would simply increase the before-tax wage and leave the after-tax wage unchanged. Again, the incidence of a wage tax is borne not by workers but by landowners. It is just an indirect—and inefficient—tax on land.

Uniform sales taxes on consumption and investment goods are, as we noted in Chapter 18, simply equivalent to proportional income taxes. They have effects that are analogous to wage taxes.[3] They are borne by land (and by workers with limited labor mobility).

---

[2] Computer-assisted mass appraisal (CAMA) is a methodology to value properties by processing standardized property data with a computer valuation model. CAMA utilizes technology and statistics to try to ensure consistency and equity in the determination of a large number of property values at a relatively low cost per property.

[3] Income taxes are taxes on wages plus income from capital. These taxes imposed on a local level can have a particularly distortionary effect on location decisions of wealthy individuals: they may choose not to live and work in a location where their productivity is highest because the net return (taking into account the additional taxes they must pay on their capital income) is lower. Of course, many rich individuals have multiple homes and claim as their principal residence the state with the lowest tax rate. This is particularly relevant for retired individuals for whom there is no work record against which to check such claims.

DISTORTIONS

The fact that all local taxes are borne by the same, immobile factors does not mean that the taxes all have the same consequences. Although a direct tax on land is nondistortionary, all the other ways of raising revenue induce distortions. The property tax (which is partly a tax on land, partly a tax on capital) raises the cost of capital to the community and induces a bias against capital. A wage tax or a sales tax raises the cost of labor and thus induces a bias against the use of labor; a sales tax may induce individuals to do their shopping across state borders. The higher sales tax in New York City, for example, induces many individuals to do their shopping in New Jersey; and when Washington, DC, imposed a higher gasoline tax than neighboring Virginia and Maryland, drivers were induced to buy their gasoline outside the city. High state income taxes may induce individuals to live in one state and commute to another to avoid the high taxes that would be levied on their income.[4] For example, because New Hampshire imposes no income tax, many people find it advantageous to live there and commute to Massachusetts, which has an income tax. The tax is inefficient, both because it raises the cost of using labor in Massachusetts and because it induces unnecessary expenditures on commuting costs.

LIMITATIONS ON THE ABILITY TO REDISTRIBUTE INCOME

The fact that taxes are borne by immobile factors means that the extent of redistribution that is feasible at the local level is very limited. Assume, for instance, that some community decides that doctors are too wealthy. The local government, accordingly, imposes a licensing tax on doctors in an attempt to redistribute income from this wealthy class of individuals to others. Doctors, in making their decision about where to set up practice, will then look at their prospects in different communities. When they discover that after-tax income is lower in this community than elsewhere, they will be discouraged from setting up practice in this community. If the tax is not too high, doctors who are already established will not leave; the costs of moving exceed the losses from the tax. The fact that its doctors do not leave may fool the community into thinking that it has been successful in extracting some additional tax out of doctors; in the short run, it may

---

[4] In such situations, they may have to pay a tax on the wages they receive where they work, although they typically must pay taxes on any interest income they receive only to the state in which they reside.

be right. Gradually, however, as fewer doctors move into the community, the scarcity of doctors will become felt, and their wages will be bid up. Wages will continue to be bid up until the after-tax wages of doctors is equal to what they could have earned elsewhere. In the long run, doctors do not bear the burden of the tax (although they do in the short run). In the long run, the community as a whole bears the burden of the tax, in the form of fewer medical services and higher prices for doctors.

The same principle holds for any factor that is mobile in the long run. A number of states have suffered under the false impression that they could, in fact, succeed in taxing capital at higher rates than it is taxed elsewhere without either offering better public services or ultimately seeing an erosion of their capital base. Some states have attempted to include income of international enterprises operating outside the state (or country) in the tax base on which they levy a corporate profits tax. If the preceding analysis is correct, such attempts cannot, in the long run, be successful. These communities are often misled into believing that they can do this, because capital does not emigrate instantaneously.[5]

RENT CONTROL

Similarly, a number of communities have been under the impression that they could reduce the return to landlords, who were viewed to be exploiting the poorer renters. They have imposed **rent control** laws, whose effect has been to lower the rents paid by renters below what they otherwise would be. Again, in the short run, such measures may indeed be successful. In the long run, however, landlords make decisions about the construction of additional apartments and the renovation and maintenance of existing apartments. If the return is lowered below the return they can obtain on capital invested in other sectors of the economy, there is no reason for them to continue to invest in housing. Consequently, the rental market will dry up. In the long run, renters will be worse off than if the government had not imposed rent control; some renters will not be able to obtain a rental apartment at any price. (Not surprisingly, this will result in a demand for the public provision of housing for those dependent on rental markets, necessitating that the community subsidize the renters through general revenues rather than having the owners of rental apartments bear the burden.)

---

[5] On the other hand, it may not be the case that, in equilibrium, the after-tax return in all communities will be exactly the same. Notions of loyalty may lead individuals to invest in their own country or community, even when they could obtain a higher return elsewhere.

Differences in information may also lead individuals to prefer investing in their own country and this reduces the mobility of capital.

CAPITALIZATION

Consider two communities that are identical in every respect except that the taxes are higher in one than in the other (say, because of less efficiency in the provision of public services). Clearly, if the price of housing in the two communities were the same, everyone would prefer to live in the community with the lower tax rate. This cannot, of course, be an equilibrium. Individuals care about the total cost of living in the community. In equilibrium, the total cost of living—the sum of taxes plus annual housing costs—must be the same in the two communities. This means that the community with the higher tax rates will find that the prices of its land is reduced proportionately (recalling our assumption that no extra services are provided with the taxes). We say that the taxes are **capitalized** in land prices.

The term "capitalized" is used here to refer to the fact that the price will reflect not only current taxes but also all future taxes. To calculate the effect of a constant tax of, say, $1000 per year on the house price, recall that a dollar next year is worth less than a dollar this year. If we received a dollar this year we could have invested it in a money market fund or bank and obtained a return of, say, 10 percent, so at the end of the year we could have $1.10. Thus, getting a dollar today is worth—or is equivalent to—$1.10 tomorrow. More generally, a dollar today is worth $1 + r$ next year, where r is the rate of interest; in other words, a dollar next year is worth $1/(1 + r)$ today.[6] Therefore, the value of T taxes this year, next year, the year after, and so on is

$$T + \frac{T}{1 + r} + \frac{T}{(1 + r)^2} + \frac{T}{(1 + r)^3} + \cdots .$$

This is the present discounted value of the tax liabilities. If the amount by which a house's price is reduced is given by the present discounted value of these tax liabilities, we say that the tax liabilities are fully capitalized in the value of the house. If two houses are identical except for their tax liabilities, and if the house prices differ by less than this amount, we say that the taxes are **partially capitalized** in the less expensive house.

INCIDENCE OF STATE AND LOCAL TAXES

- Immobile factors bear all local taxes.
  - Capital is likely to be very mobile.
  - Skilled labor is likely to be relatively mobile.
  - Land is immobile.
- Application of general principle—factors in elastic supply do not bear much of the brunt of taxation; with perfect mobility, supply elasticity is infinite.
- Limits ability to redistribute income locally.

---

[6] By the same reasoning, a dollar the year after next is equivalent to $1/(1 + r) = $1/1.10 = $0.91 next year, but this means that, because a dollar next year is worth $1/(1 + r) = $1/1.10 today, a dollar the year after next is worth $1/(1.10 × 1.10) = $0.83 this year—that is, $1(1 + r)^2$. See Chapter 11 for a more extensive discussion of present discounted value.

INCENTIVES FOR PENSION SCHEMES

The fact that certain fiscal variables may not be fully capitalized has some important implications. There are incentives for communities to take advantage of this. Someone living in a community who thinks there is a reasonable chance that he or she will move out in ten years or so might vote for a large, unfunded pension scheme for public employees—that is, a pension scheme that fails to set aside the funds that will be needed to pay the promised pensions, but relies instead on future taxes. A generous pension allows the community to attract workers while paying lower current wages. In effect, future homeowners in the town will be forced to pay for current services. The future buyer of a house is being deceived in much the same way that the manufacturer of a product that does not fully disclose some important characteristics of its commodity may attempt to deceive a purchaser. An important characteristic of a house (or any piece of property) is the future tax liabilities that are associated with it; to know these, one must know the debt and unfunded pension liabilities of the community. Whether the appropriate way to deal with this is through disclosure laws (each community being required to notify all potential purchasers of a house of the debt obligations of the community prior to the completion of any sale) or through restrictions (not allowing unfunded pension schemes) is a debatable question.

CHOICE OF DEBT VERSUS TAX FINANCING

More generally, the extent of capitalization has implications for the decision whether to finance local public expenditures by debt or taxes. With full capitalization, a dollar increase in the local debt would simply decrease the net market value of the community by a dollar. Because house buyers can choose to live in this community or in some other community, their assumption of the debt of the community is a voluntary action. Therefore they will have to be compensated for it, through a corresponding decrease in the price of a house. This is true no matter how far in the future the debt is to be repaid. It does not have to be repaid during the period in which the next owner owns the house.

Assume, for example, that the debt is to be repaid in forty years, and each individual lives in the house for only ten years. The person who buys the house at the time that the debt is to be repaid clearly will pay less for the house, taking into account the increased tax liability associated with paying off the debt. However, the preceding purchaser knows that the person to whom he or she will sell the house will be willing to pay less

for it, and hence this purchaser will be willing to pay less for it (by the amount of the tax liability). Similarly, the previous purchaser knows that the price at which he or she can sell it will be lower by the amount of the increased debt, so this purchaser, too, will be willing to pay less for it, and so on.

With full capitalization, current owners pay for current services—whether directly, through taxes, or indirectly, through the expectation of a lower price on their house resulting from a higher debt used to finance the public services. Which of these two methods is preferable turns out to depend on an individual's ability to borrow and the treatment of local taxes and interest on local debt by the federal government, a question that was discussed in the previous chapter.[7]

SHORT-RUN VERSUS LONG-RUN CAPITALIZATION

Assume that taxes are increased on apartment buildings. If the amenities the community provides are unchanged, the rents will remain unchanged: the rents individuals are willing to pay depend on the services provided by the community and the landlord, rather than on the costs to the landlord of those services. In the short run, the market value of the apartment will thus decrease. However, this will make investing in apartments in the community less attractive; the supply of apartments will be reduced (as old apartments deteriorate) or, in any case, will not keep up with population growth. This will result in an increase in rents. Eventually, rents will increase to the point at which the after-tax return on the apartment is the same as investors could obtain from investments in any other community. Thus, although the tax is imposed legally on buildings, in the long run it is again land and immobile individuals (who must pay higher rents) who bear the tax.

WHO BENEFITS FROM LOCAL PUBLIC GOODS? THE CAPITALIZATION HYPOTHESIS

The same reasoning that leads us to conclude that the incidence of any tax resides with the owners of land (or other partially immobile factors)

---

[7] In the presence of credit rationing (limited availability of mortgages), the lower price of a house may increase its salability. The fact that communities may borrow more easily than individuals provides an argument for communities to borrow as much as they can.

implies that the incidence of any benefits resides with the owners of land (or partially immobile factors). Any public good that makes a community more desirable to live in drives up the rents and hence increases the value of property in the community. In the short run, some of the benefits may be enjoyed by owners of buildings, but the increased rent on their buildings leads to increased investment in housing (new apartment buildings; replacing small, old apartment buildings with larger ones; and so on), and this drives down their return.[8] Ultimately, the value of the public good is reflected in the price of land.

Similarly, some public goods make it more attractive to work in a given community. This will reduce the wage a firm must pay to recruit a worker. Again, though, the ultimate beneficiaries are the landowners.

To see the link between the wages individuals receive and the level of public services provided, consider what happens if a city decides to spend more on its symphony orchestra, which provides free concerts in the parks in the summer. This makes the city a more attractive place in which to live and work. A worker who enjoys the symphony, contemplating a job offer in this city, will accept the job at a slightly lower wage than he or she would accept in a community that is identical in every respect except for its level of expenditure on its symphony orchestra. Thus, to the extent that workers in the city (regardless of where they choose to live) value the amenities it provides, wages in the city will be lower; firms will find it attractive to locate there. As they move into this city, the price of land will be bid up. Equilibrium is attained when the price of land is bid up just enough to compensate for the lower wages, so investors receive the same return to their capital that they receive from investing it elsewhere. The ultimate beneficiaries of the provision of better public goods are not the residents in the city, but the landowners.

This analysis assumes, of course, that labor is highly mobile, so when the city provides a more attractive public good, there is sufficient migration to decrease wages and increase rents. If labor is not very mobile (and, in the short run, it may well not be), wages will not fall to fully reflect the increased amenities, and some of the benefits of the increased provision of local public goods will accrue to the current residents. Note that some current residents may be hurt by the provision of the symphony—those who do not enjoy music may find that their rents are increased or wages reduced nonetheless.

The city's provision of the symphony orchestra does have important spillovers to other communities. In particular, firms located in the

---

[8] Current owners of buildings have an incentive to maintain their higher return by restricting further investments by zoning. The higher returns they enjoy should be viewed not as a return to capital but as a return to the property rights the zoning board has created.

suburbs will find that they, too, can hire workers at a lower wage than they previously could, as music lovers will find accepting a job in a community near the city more attractive than accepting a job in an otherwise identical community without easy access to a symphony. This increases the value of the firms' land as well. Bedroom suburbs will also find that the demand for their housing has increased.

ABSOLUTE VERSUS RELATIVE CAPITALIZATION

We have discussed how if a community increases its level of expenditure on a public good, the differential expenditure will be reflected in the prices of the land in the community. There is, however, an important difference between the effects of a single community's increasing its expenditure on a public good and *all* communities' increasing their expenditures on that public good. If all communities increase their expenditures on a public good, the relative attractiveness of living in one community versus another is, of course, unchanged. Thus, in general, rentals will remain unchanged.[9]

This is an example of a phenomenon we noted in earlier chapters. The effects of a change in one community (a change in a tax on one commodity) may be quite different from the effects of a change in all communities (a change in the tax rate on all commodities).

THE USE OF CHANGES IN LAND RENTS TO MEASURE BENEFITS

Changes in rents have often been used to measure the value of certain public services. In studies of the economic effects of American railroads in the nineteenth century, one commonly employed way of measuring the benefits is to measure the change in land rents after construction of the railroad. Again, one must be careful to distinguish partial from general equilibrium effects. Making one small plot of land more accessible will increase the demand for that plot of land, and the change in the rent will provide an accurate estimate of the reduction of the transportation costs of getting to that piece of land. However, changing the

---

[9] There are exceptions. If the communities provide a public good that makes land more desirable, all individuals will attempt to rent or purchase more land, and this will increase the value of land. The opposite will be true if communities provide a public good that makes owning land less desirable. Thus, as public parks are, in part at least, a substitute for backyards, it is conceivable that if all communities spend more on providing public parks, rents and land prices would actually decrease.

CAPITALIZATION OF TAXES AND LOCAL PUBLIC GOODS

Future benefits and taxes are reflected in today's price of land.

- Implies that unfunded local public pension schemes, such as for police, depress land values.
- Implies that there is no difference between debt and tax financing—current owners pay cost of current consumption expenditures.
- Investment values can be affected in the short run, but in the long run, returns on investment must be equalized across communities.

A tax in one community on capital has a markedly different effect from a tax in all communities.

accessibility of a mass of land—as the railroad, in fact, did—has general equilibrium effects; the change in land rents will not correctly assess the value of such a change.[10]

Land values reflect the valuation of marginal individuals—those who are indifferent about choosing between living in this community and living somewhere else. When there is sufficiently large number of communities, the valuation of these marginal individuals provides a good measure of the valuation of the entire community, but not otherwise.[11]

TESTING THE CAPITALIZATION HYPOTHESIS

The question of the degree to which the benefits provided by public goods and taxes are reflected in property values has been studied extensively.

If some communities are more efficient in providing public goods than others, so they can provide the same level of public goods with lower taxes, property values in the low-tax communities should be higher. If all communities are equally efficient and maximize their property values, differences in taxes will be matched with differences in benefits. In this case, there will be no systematic relationship between property values and expenditures. This is the result obtained by Jan Brueckner of the University of Illinois, in his study based on fifty-four Massachusetts communities.[12]

On the other hand, there is evidence that the value of amenities such as clean air, for which there are not corresponding taxes, is capitalized in property values.

---

[10] There is a second limitation on the use of land rents to measure the value of such changes: they provide a good measure only in the case in which there are no inframarginal individuals—those who are enjoying a consumer surplus from living in the community.

[11] R. Arnott and J. E. Stiglitz, "Aggregate Land Rents, Expenditure on Public Goods and Optimal City Size," *Quarterly Journal of Economics* 93 (1979): 472–500; R. Arnott and J. E. Stiglitz, "Aggregate Land Rents and Aggregate Transport Costs," *Economic Journal* 91 (1981): 331–347; and D. Starrett, "Principles of Optimal Location in a Large Homogeneous Area," *Journal of Economic Theory* 9 (1974): 418–448.

[12] J. K. Brueckner, "A Test for the Allocative Efficiency in the Local Public Sector," *Journal of Public Economics* 19 (1982): 311–331.

PUBLIC CHOICE AT THE LOCAL LEVEL

In Chapter 9, which described how public choices are made, we showed that with majority voting, the allocation of public goods reflects the preferences of the median voter.[13] This voter assesses the costs and benefits to him or her of the expenditure of an extra dollar on public goods. We then assessed the efficiency of the majority voting equilibrium.

The issues at the local level are identical; in both cases, we need to focus on the incidence of the benefits and costs associated with any increase in expenditure and taxation. We need to distinguish between the effects on renters and on landowners, under assumptions of perfect and imperfect mobility (with a large or small number of competing communities). It is useful to distinguish between two types of communities: one in which everyone is a renter and there is a separate group of landlords; and the other, in which everyone is a homeowner. In the real world, of course, communities are mixed; in a few cities like New York, most people are renters, whereas in most of America, majority are homeowners. Because renters and homeowners may be affected quite differently by local taxes and expenditures, they may vote for quite different policies.

We begin with a discussion of a world in which (essentially) everyone is a renter. With perfect mobility and a large number of competing communities, any improvement in the amenities provided by a community will be fully reflected in rents; hence, marginal renters will be indifferent with respect to the public services provided. Moreover, as their rents are affected only by the services that are provided and not by the tax rates, renters will be completely unconcerned about the efficiency with which public services are provided—that is, with their cost. Thus, if there is perfect mobility and a large number of competing communities, there will be little concern about efficiency or about the public services provided in a community in which voters are mostly renters.

Under these same assumptions, landowners as a group will want public services to be increased as long as they lead to increases in rents exceeding increases in taxes. Thus, in a landowner-controlled community, in equilibrium an extra $1000 spent on public goods should just increase aggregate rents by $1000. However, the increased rents represent renters' marginal evaluation of the services provided by the community. As a result, a landowner-controlled community will provide an efficient level of public services. Moreover, because if the community can provide the same services at less cost, the after-tax receipts

---

[13] Assuming, of course, that a majority voting equilibrium exists.

PUBLIC CHOICE AT THE LOCAL LEVEL

- Renters and landowners can be affected very differently by changes in local taxes and expenditures, which will be reflected in how they vote.
- In a world with perfect mobility and strong competition among many communities, rents will reflect the value of services provided and taxes will be borne by landowners (the immobile factor). Hence, renters will be indifferent to the quality of publicly provided services and the efficiency with which they are delivered.
- In such a world, landowners' property values will fully reflect increases in efficiency and services provided, and hence landowners will have an incentive to provide an efficient level of services and to have those services provided efficiently.
- In a world with imperfect competition among communities, if communities are controlled by renters, there will be excessive diversity of benefits (e.g., in levels of expenditures), whereas if communities are controlled by landowners, there will be insufficient diversity.

of landowners will be increased, landowner-controlled communities have every incentive to ensure that public services are provided in an efficient manner. Thus, if the level of expenditures is chosen to maximize property values, and if there is effective competition among communities, the resulting allocation of resources will be Pareto efficient.

All this changes if there are relatively few communities competing against each other. Consider a metropolitan region in which there are two towns: A and B. Town A has high taxes and a high level of local public goods, whereas Town B has low taxes and a low level of public goods. Those who have a strong preference for public goods (relative to private goods) live in A, and those who have a strong preference for private goods live in B. An individual who is indifferent with respect to living in the two communities is called the *marginal* individual; the extra public goods this individual receives in A just compensate him or her for the extra taxes or rent he or she has to pay. All other individuals are called **inframarginal**. For those who live in A, for instance, the extra benefits more than offset the extra taxes they have to pay. Were A to increase its taxes slightly without altering its benefits, these individuals would still not wish to move to B.

Assume that there are houses for half the region's population in A and half in B. All housing is rented. If B decides to provide fewer public goods, rents in B will have to adjust so that the marginal individual is still indifferent with respect to living in A or B. Of all the individuals who live in B, the marginal individual is the one with the strongest preference for public goods, so the rents will fall in B to just compensate this individual for the lower level of public goods. If the rents decrease enough to make the marginal individual remain indifferent, the other individuals in B are actually better off. Rents fall by more than enough to keep the inframarginal persons, including the median voter, satisfied. Therefore, the median renter in B will have an incentive to vote for a very low level of expenditure on public goods, lower than is Pareto efficient. The same reasoning shows that the median renter in A will have an incentive to vote for a very high level of expenditure on public goods, higher than is Pareto efficient.

Landowners have exactly the opposite bias. They are concerned only with the effect of increased expenditure on land values (rents). If the increased rents exceed the increased expenditures, they are worth undertaking. In A, the increased rents from an increased expenditure reflect the marginal individual's evaluation—this is the individual who has the weakest preference for public goods. Thus, the gain to others (the inframarginal renters) exceeds the gain to the marginal renter, but the landowners will pay no attention to this. As a result, they will vote for too little expenditure on public goods. By the same reasoning, in B, landowners will vote for too high an expenditure on public goods.

Thus, just as we saw in Chapter 9 that the majority voting equilibrium did not provide a Pareto efficient level of expenditures on public goods in an isolated community, such is also the case when there is only a small number of communities. When there is limited competition, there may be systematic biases in the patterns of allocation, and marked differences between communities in which renters dominate (where differences among communities may be excessive) and those in which landowners dominate (where diversification among communities may be insufficient). However, if the number of communities is very large, and if they all recognize that capital and labor (of different skills) are perfectly mobile, then they will compete effectively against each other, providing an efficient supply of public goods corresponding to the preferences of the individuals in the different communities, and providing them in an efficient manner. Although, under these circumstances, renters will be indifferent to what the government does, landowners will not be.[14]

PROBLEMS OF MULTI-JURISDICTIONAL TAXATION

In the previous chapter we discussed the problems of fiscal federalism, of the relationship between the federal government and the states. There is a second, related set of problems posed by federalism: the relationship among the states themselves.

---

[14] There are clearly a number of communities within most metropolitan areas. However, are there enough to ensure that the resulting equilibrium is "close" to Pareto efficient? There is no agreement among economists about the answer to this question.

The case in which there is a sufficiently large number of communities to ensure efficiency has some further peculiar implications: there would be unanimity among all voters about what the local government should do, and all individuals of any skill type within the community would be identical. These implications are sufficiently counter to what is observed in most situations to suggest that models assuming limited competition are closer to the mark.

One aspect of that relationship has been stressed in this and the previous chapter: competition. We saw that whereas Tiebout competition helps ensure economic efficiency, tax competition sometimes ensures that businesses escape much of the burden of taxation. The issue has become a source of acute concern, as some states have offered large tax breaks for businesses that establish new plants in their jurisdiction. As states bid for the plants, the tax breaks get larger. Plants may, in fact, be built where they would have been built anyway—but the shareholders gain by paying lower taxes regardless of where the businesses settle, and other taxpayers are left bearing a larger tax burden.

The fact that America is such an integrated economy, with free mobility of goods and people among jurisdictions, raises a host of practical problems in taxation. Just as businesses make decisions about where to locate based partly on tax considerations, individuals who can, also make choices on where to live based partly on tax considerations. A retired person with a large capital income or a writer with large royalties might choose to make New Hampshire his or her "official" residence, because there is no income tax in that state. If this individual's income were $500,000, California, for instance, would impose a charge in excess of $50,000 a year for the right to make that state his or her official residence. Tax considerations might also affect whether a high-income individual working in New York City would choose to live in New York, New Jersey, or Connecticut.

State sales taxes clearly induce some cross-border shopping and shipping, even though, in many states with sales taxes, one is supposed to pay taxes on goods purchased in other states.

The corporation tax, imposed by many states, represents the greatest challenge. There is a fundamental problem in levying such a tax: determining how much of the corporation's income should be attributed to economic activity occurring within the state. States increasingly have come to use what is called the **unitary tax system**, relying on simple formulae. A fraction of the corporation's worldwide income is attributed to activities within the state, depending on the fraction of the corporation's assets, sales, and employment within the jurisdiction.

As long as differences among states in the levels and forms of taxation remain limited, the multijurisdictional tax problems, though conceptually important, will not be too severe. There remain, however, incentives for some states to use tax policy more actively, to increase their tax base and to attract people and businesses. To the extent that those states most inclined to do so are poorer states—such as New Hampshire—more in need of the revenue, the resulting distortions are offset by the distributional gains. Still, the same distributional outcomes could be obtained more efficiently with a more active federal policy of redistribution.

REVIEW AND PRACTICE

SUMMARY

1. If capital and labor are mobile, the incidence of any tax lies on land, the immobile factor. If labor is only partially mobile, some of the burden may lie on it.

2. Local taxes that are imposed on mobile factors—sales taxes (which are equivalent to income taxes), wage taxes, corporation income taxes, and property taxes on buildings—induce distortions.

3. Improved public services provided by the government are reflected in rents paid. In a perfectly competitive environment (with a large number of communities with similar individuals), the benefits of improved government services accrue solely to landowners.

4. Future benefits and taxes may be capitalized in current land values.

5. The effect of an increase in benefits on land rents (values) will depend on whether one community alone increases its benefits or all communities increase their benefits.

6. If the level of expenditures is chosen to maximize property values, and there is effective competition among communities, the resulting allocation of resources is Pareto efficient.

7. If there is limited competition, however, the resulting equilibrium is not Pareto efficient; there is a tendency for too little diversification in the services provided by the different landowner-controlled communities.

8. In contrast, when renters control the community, there is a tendency (under the same circumstances) for excessive diversification. Communities that spend a great deal on public goods spend too much; communities that spend little spend too little.

9. Moreover, there is no incentive for renters to be concerned with the efficiency with which the government delivers its services.

10. In an economy with many tax jurisdictions, two further problems arise. First, tax competition sometimes results in businesses' escaping much of the burden of taxation, as some states use tax policy to increase their tax base and attract people and businesses. Second, there are often difficult problems of deciding how much income should be attributed to economic activity occurring within the jurisdiction; differential tax rates among jurisdictions may result in distorted locational decisions.

KEY CONCEPTS

Capitalized
Inframarginal
Partially capitalized
Rent control
Unitary tax system

QUESTIONS AND PROBLEMS

1. Explain why the property tax may lead to lower expenditures on capital (buildings) per unit of land.

2. Many firms have employees, plants, and sales in more than one state. In imposing state corporation income taxes, states use a rule for allocating a fraction of a firm's total profits to their state. Does it make a difference what rule is used? Discuss the consequences of alternative rules.

3. Many of the issues of state and local taxation are similar to issues that arise in international contexts. Many countries, for instance, have imposed taxes on capital owned by foreigners. Discuss the incidence of such taxes. Does it pay to subsidize capital owned by foreigners?

4. Discuss the incidence of a city wage tax.

5. Many cities have passed rent control legislation. Discuss carefully who benefits and loses from such legislation, in the short run and in the long run. Discuss the political economy of such legislation.

6. Henry George, a famous nineteenth-century American economist, proposed that only land (not buildings) be taxed. Would this be unfair to landowners? Would it distort resource allocations, making land more expensive relative to buildings?

7. Who are the main beneficiaries of tax-exempt industrial development bonds (which enable communities to borrow funds to re-lend to firms constructing new plants within the city):

   a. Workers in the town?

   b. Landowners in the town?

   c. The industries that move into the town?

   Give your assumptions. Does it make a difference whether only one community provides these bonds or all communities provide them?

8. Who benefited from the construction of the subway system in Washington, DC:

   a. Owners of land near the subway line at the time the route was announced?

   b. Owners of land near the subway line at the time the subway was completed?

   c. Renters of apartments near the subway line at the time the route was announced?

   d. Renters of apartments near the subway line after the subway was completed?

   e. Renters of apartments not near the subway line?

   In each case, give your assumptions.

9. In 1993, a major dispute broke out between California and the United Kingdom involving taxation of the income of Barclays Bank, one of the major British banks. California imposed a tax that allocated a fraction of Barclays's income to its activities in California, and taxed that accordingly; Barclays said that the formula used resulted in far too high a fraction. The U.K. government supported Barclays's position and threatened economic retaliation. The United States, in its international negotiations over the years, had criticized other countries using what are called *unitary* tax systems, in which income is allocated on the basis of a simple formula, as described in the text. It argued for using the transfer price system, in which an attempt is made to estimate the value of the goods and services that a company's plant in one country receives from and delivers to the company's facilities in other countries. (It attempts to "price" these transfers.) With the worldwide integration of economic production, such systems appear to be increasingly cumbersome; no state within the United States attempts to use a transfer price system. What do you think the U.S. government should have done in the Barclays's case?

28

FISCAL DEFICITS AND GOVERNMENT DEBT

FOCUS QUESTIONS

1. What has given rise to fiscal deficits?

2. Why have fiscal deficits been so difficult to reduce? What are the central economic issues in the debate about how deficits should be reduced?

3. What are the procedural reforms that have been proposed to ensure that deficits do not emerge in the future, and what are the problems with these reforms?

4. What are the long-term problems facing the United States and other high-income countries that, if not adequately addressed, are likely to result in a deficit problem emerging in the future, even if the deficit is reduced now?

5. What are the long-term consequences of chronic government deficits?

When the federal government spends more than it receives in taxes and other revenues in any given year, it has a budget deficit, commonly referred to as the **fiscal deficit**. The U.S. fiscal deficit ballooned during the 1980s and emerged from being a problem that interested mainly economists and political pundits into the national spotlight. By the 1992 presidential election, public opinion polls were persistently ranking the huge deficits among the central problems facing the country. The increased deficits led to increasing debt, and an increasing burden on the government simply to pay interest on the debt. To many economists, the higher deficits threatened the long-term well-being of the economy; but regardless of one's view of the deficit's economic consequences, the attempt to reduce the deficit had fundamental effects on every aspect of government operations. Federal employment was cut back to levels of the early 1960s, and, as a percentage of civilian employment, to levels not seen since the early 1930s.

Two major steps to bring the deficit under control were undertaken, in 1990 and 1993, both entailing controlling expenditures and raising revenues. After reaching 4.7 percent of GDP in 1992, the deficit as a share of GDP shrank rapidly in the ensuing years, to a negligible 0.3 percent of GDP in 1997. In August 1997, Congress passed and the President signed a five-year budget agreement that promised to eliminate the deficit by 2002. By 1998, the reforms of 1990 and 1993—combined with a booming economy and strong economic growth—had turned the deficit into a $70 billion surplus, with even larger surpluses looming over the horizon.

However, although many economists worried that increasing expenditures on entitlements, especially for the aged, would almost surely lead to mounting deficits within a quarter century, President Bush argued for tax cuts in 2001 and 2003, and expanded the Medicare program by including prescription drugs, with an expensive proviso that the federal government, the largest purchaser, could not bargain with the drug companies. Alan Greenspan, chairman of the Federal Reserve, supported President Bush, arguing that the surpluses would enable the country to quickly pay back the national debt, and that without a national debt, the conduct of monetary policy would be difficult. In retrospect, the magnitude of the misjudgment was colossal: the tax cuts resulted in large deficits, which only grew in magnitude after the global financial crisis of 2008.

This chapter provides a bridge between the two branches of economics: macroeconomics, which is concerned with national aggregates such as output, employment, and inflation; and microeconomics, which is concerned with the individual decisions of households and firms. Today, courses in the economics of the public sector focus on how taxes and expenditures affect these decisions and thus the *structure* of the economy—what goods get produced, how they get produced, and for whom they get produced—whereas courses in macroeconomics focus on how government affects the *level of* economic activity. In reality, the two subjects are intimately connected; the deficit has affected both macroeconomic and microeconomic policies. It has, for instance, reduced the ability of the government to use traditional fiscal policy (reductions in taxes or increases in expenditure) to stimulate the economy, and hampered government's ability to increase public investments, such as in infrastructure like roads, human capital like education, and research to promote economic growth.

This chapter reviews the origins of the deficit problem, the solutions of the early 1990s that led to fleeting budget surpluses, the subsequent relapse into annual deficits, and the debate about the consequences of deficits.

THE U.S. DEFICIT PROBLEM SINCE THE 1980s

The U.S. deficit problem began around 1981, when taxes were cut but expenditures were not cut commensurately. Figure 28.1 shows what happened. The real deficit peaked at more than $390 billion in 1992. The deficit as a percentage of GDP rose to what was at that time a peacetime record of 6 percent in 1983, hovered between 3 and 5 percent for most of the remainder of the decade, gradually decreased from 4.7 percent in 1992, and finally turned into a surplus in 1998. Deficits recurred beginning in 2002, peaking at 10 percent of GDP in 2009 in the throes of the Great Recession.

Figure 28.2 places the budget performance of the United States in international comparative perspective. The deficits of most countries increased during the global economic downturn, and countries with budget surpluses saw these surpluses shrink considerably.

But one has to be careful about interpreting these comparisons. The situation of each country is different. Japan has had large deficits for many years, but because the Japanese government can borrow at close to zero interest rate, the deficits have not (so far) presented a problem. Greece ran into difficulties in financing its deficit even when its national debt (relative to GDP) was lower than that of Japan, or even Italy. Moreover, the U.S. deficit was a key factor in explaining the better U.S. performance in the aftermath of the crisis. Six years after the crisis, the United States was among the few countries that had a real output per working-age population that was above the pre-crisis level.

SOURCES OF THE DEFICIT PROBLEM

A good way to pose the question of what caused the soaring budget deficit is to ask: What changed between the 1970s (and earlier decades) and the 1980s (and subsequent decades)? We are especially interested in how the country could have gone from large surpluses in 2001 to persistent large deficits in a few short years, and why some forecasters (like Alan Greenspan) estimated the future so badly. (Some of the mistakes were clearly just wishful thinking: some *wanted* a tax cut, and constructed a forecast that made it seem as though the country could afford it.) There are six main answers to this question.

REDUCED FEDERAL TAXES   During the 1970s, federal receipts averaged 18 percent of GDP. In 1980, that percentage climbed to 19 percent,

**BUDGET DEFICITS IN THE
UNITED STATES, 1980–2010**
The federal budget deficit is
depicted here in three series
of data. (A) Shows the nominal
deficit. (B) Shows the real
deficit in 2005 dollars. (C) Shows
the federal deficit as a
percentage of GDP.

SOURCE: Executive Office of the
President, Office of Management and
Budget, *Budget of the United States
Government, Fiscal Year 2013,
Historical Tables*, Tables 1.1–1.3.
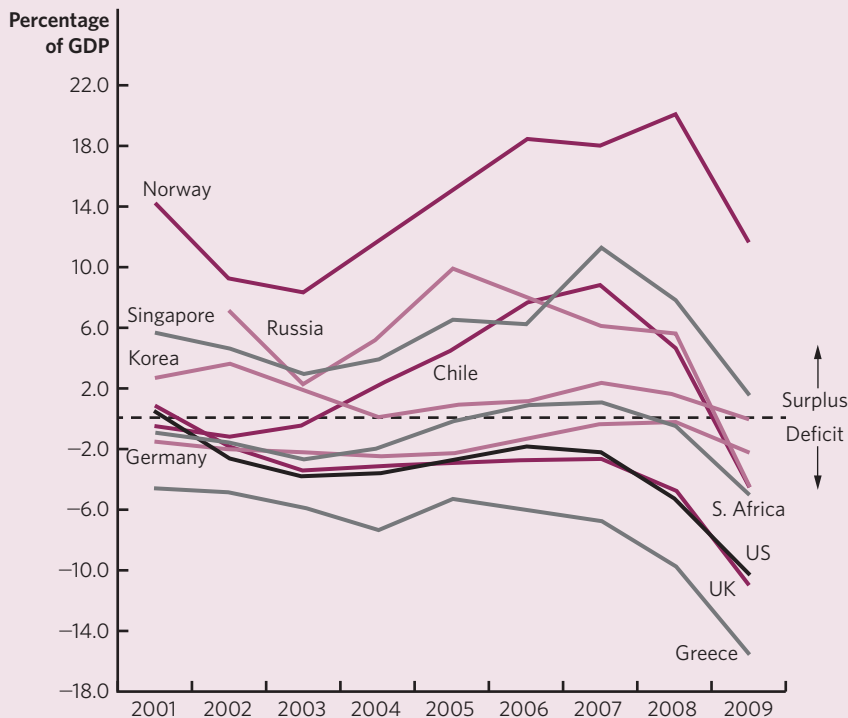
BUDGET DEFICITS AROUND THE WORLD, 2001–2009
Although the United States had one of the largest budget deficits among high-income countries in 2009, surpassed only by the United Kingdom and Greece, all countries saw their deficits rise (or surpluses shrink) during the recent global economic downturn.

SOURCE: "Cash Surplus/Deficit (% of GDP)," *World Bank World Development Indicators.*

and rose again to 20 percent in 1981. (Remember, 1 percent of a $15 trillion economy will be $150 billion.) Tax cuts enacted early in the Reagan presidency pushed the federal tax take back down to its historical average of 18 percent of GDP. The slight increase in taxes in 1993 raised the average to 19 percent, but the subsequent 2001 and 2003 tax cuts reduced this average to 17 percent, and, coupled with the impact of the Great Recession, federal tax revenue as a share of GDP dropped below 15 percent in 2009 and 2010, the lowest it has been since 1950. (With progressive tax systems, when incomes go down, the share of income collected goes down.)

HIGHER DEFENSE SPENDING   Federal defense spending fell during the 1970s, as the Vietnam War came to an end, dropping from 9.4 percent of GDP in 1968 to 4.7 percent of GDP in 1978, half the share of a decade earlier. In 1979, though, following the Soviet Union's invasion of Afghanistan, President Carter called for a large defense buildup. After Ronald Reagan was elected President in 1980, he followed through on these plans. From 1983 to 1988, defense spending was 6 percent of GDP. With the end of the Cold War, there was a gradual reduction in defense spending, reaching a

low of 3 percent of GDP from 1999 to 2001. Although the threat of small wars in various parts of the world has kept the "peace dividend" smaller than many had hoped, the reduction in defense spending as a percentage of GDP played an important part in the deficit reductions after 1993. However, the conflicts in Afghanistan and Iraq proved enormously expensive—far more expensive than the Bush administration estimated at the time those wars began, with the ultimate costs in the trillions of dollars.[1] Defense spending as a share of GDP rose even beyond that, by hundreds of billions of dollars a year, reaching 4.8 percent in 2010—far higher than it was in 1993, and comparable to its share in 1979, at the height of the Cold War. The U.S. spending now is roughly equal to that of the rest of the world combined.

HIGHER SOCIAL SPENDING ON THE ELDERLY   As the elderly population in the United States has grown, not only absolutely but as a proportion of the population, federal expenditures on programs such as Social Security and Medicare have expanded dramatically. These programs averaged 5 percent of GDP in the 1970s, but increased to over 6 percent of GDP by 1980 and close to 7 percent by 1982. They have now risen to more than 8 percent of GDP, and are expected to increase substantially over the next several decades. However, even though these increases have been occurring steadily over the past thirty years, they do not explain much of the changed fiscal circumstances after 2001.

INCREASING HEALTH CARE EXPENDITURES   Through Medicare (program providing health care to the aged) and Medicaid (program providing health care to the poor), the government has assumed an increasing share of total health care expenditures to the point at which today the federal government alone pays 36 percent of all health care expenditures. Health care expenditures themselves have soared. Through the 1980s and early 1990s, these expenditures were increasing at close to 12 percent per year, doubling every six years. A number of initiatives undertaken during the Bush and Clinton administrations brought the rate of increase down to between 8 and 10 percent, a number still far higher than the rate of income growth. By the late 1990s, health care costs were beginning once again to rise at a more rapid rate, but there is some evidence that since 2010 the rate of increase has come down.

Some of the increases reflect the aging of the population; some are the consequence of important innovations, which improve the length and quality of life. The real concern with the U.S. health care system, as we noted in

---

[1] See L. Bilmes and J. E. Stiglitz, *The Three Trillion Dollar War: The True Costs of the Iraq Conflict* (New York: W. W. Norton & Company, 2008).

Chapter 13, is that it is very inefficient—delivering arguably poorer outcomes with much greater expenditures than other advanced industrial countries. By some calculations, if the United States had as efficient a health care system as the best four or five in Europe, the nation would have no deficit.

All this was known at the time of the 2001 and 2003 tax cuts, and so it does not fully explain the seeming unexpected change in the country's fiscal outlook. However, the provision that the government could not bargain with the drug companies (an example of what critics call *corporate welfare*) is estimated to cost the federal government a half trillion dollars over ten years.

HIGHER INTEREST PAYMENTS   Like other borrowers, the federal government pays interest. During the 1970s, federal net interest payments averaged 1.5 percent of GDP. From 1985 to 1992, however, their relative size more than doubled, exceeding 3 percent of GDP. The main reason was the increasing debt. If the debt in 1995 had been the same as it had been at the start of the Reagan era in 1981 (after adjusting for inflation), the government would have been running a balanced budget, rather than a deficit of $160 billion. Deficits, through interest payments, feed on themselves. The lower interest rates in the late 1990s contributed to eliminating the deficit. Ironically, despite a substantial increase in federal government borrowing during the Great Recession, near-zero interest rates have lowered government interest payments as a share of GDP to the 1970s level of about 1.5 percent. Again, though, this does not explain the turnaround since 2001.

THE ECONOMIC DOWNTURN   We have noted that tax revenues (as a share of GDP) automatically decrease when the economy goes into a downturn. At the same time, expenditures on unemployment and welfare benefits, like food stamps, and even social insurance, goes up. Older people who cannot find jobs start collecting Social Security. Disability payments also go up. About half the deficit in the years immediately following the recession was caused by the economic downturn. Of course, no one in 2001 expected a downturn of this magnitude.

FACTORS NOT CONTRIBUTING TO THE DEFICIT PROBLEM

In a sense, any expenditure can be thought of as contributing to the deficit problem—reducing that expenditure, other things equal, would reduce the deficit. However, some factors are blamed for the deficit problem

undeservedly. For instance, polls suggest that many Americans believe that welfare payments and foreign assistance are at fault. General welfare payments are less than 3 percent of the federal budget, though, and their share of the federal budget has decreased in recent decades. Under the AFDC/TANF program, real benefits per family fell by 57 percent from 1970 to 2009. Similarly, foreign assistance is minuscule—about 1 percent of the federal budget.

SUCCESS IN TAMING THE DEFICIT: THE EXPERIENCE OF THE 1990s

The deficit was brought under control beginning around 1993. Why was this so difficult, and how was success finally attained?

MEASURING BUDGET DEFICITS: WHAT'S LARGE, WHAT'S REAL, AND WHAT'S RIGHT?

The deficits of the 1980s and early 1990s were large. At its peak in 1992, the dollar figure was $290 billion. But are dollar amounts the right way to measure the size and importance of deficits? Shouldn't we take into consideration the effects of inflation and the growth of the overall economy?

The late Robert Eisner of Northwestern University argued for focusing on the overall increase in the *real* debt—that is, the debt adjusted for changes in the price level. With total debt outstanding to the public of approximately $3.6 trillion, an inflation rate of 3 percent implies a *reduction* of the real value of the debt of $108 billion per year. This decrease in the real value of the outstanding debt should, in Eisner's view, be subtracted from any increase in debt due to the deficit. In fiscal 1996, this inflation-adjusted deficit was only $9 billion, compared to the measured deficit of $117 billion.

According to this definition, the Carter administration actually ran an inflation-adjusted surplus due to the effects of inflation on the value of the debt. By contrast, the relatively low rates of inflation during the Reagan and Bush administrations were not enough to offset the large deficits due to expenditures far exceeding revenues.

The *full-employment* or *structural* deficit takes into account the level of economic activity in the economy. It asks: What would the deficit have been if the economy had been operating at full employment? For instance, in 1991, as the economy was in recession, the deficit was officially $269 billion—the largest in history up to that point—but the structural deficit was $191 billion, still large by any standards, but $78 billion less than the actual deficit. The Reagan years look particularly bad from this perspective, because, except during wars, the economy had never previously run large structural deficits.

In recent years, as well as in the years immediately following World Wars I and II, the U.S. government was saddled with a huge debt. The interest payments on this inherited debt made it particularly difficult to attain a balanced budget. The *primary deficit* takes into account what the deficit would have been had there been no inherited debt; that is, it subtracts interest payments from the deficit. If the government is running a primary surplus, it means that revenues more than cover current expenditures. Since 1970, the government ran a modest primary deficit for most years until it generated a significant primary surplus from 1995 to 2002. Since then, the government has run a primary deficit every year except 2007, and the primary deficit is now more than $1 trillion (see figure below).
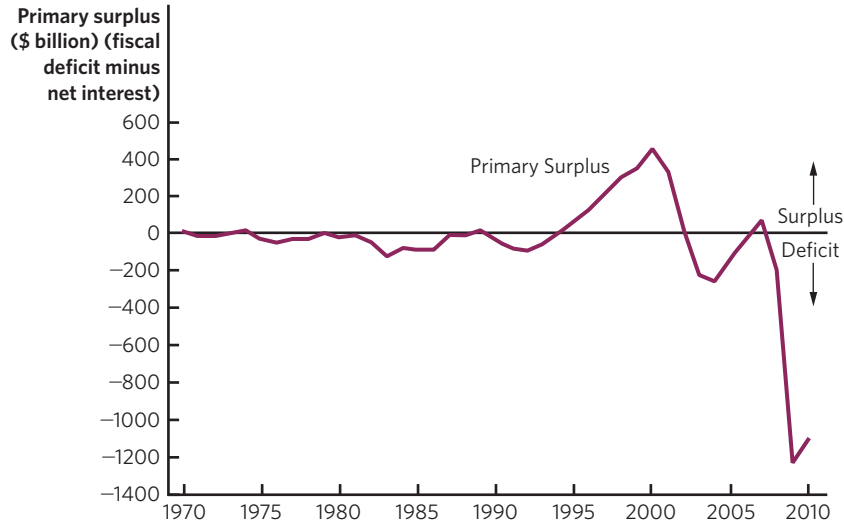
SOURCE: Executive Office of the President, Office of Management and Budget, *Budget of the United States Government, Fiscal Year 2013, Historical Tables*, Tables 1.1 and 3.1.

The answer to the first question can be seen by looking once again at the composition of government expenditures (Figure 28.3). Expenditures typically are divided into three categories: defense, nondefense discretionary, and entitlements. **Entitlement programs** are programs such as Social Security and Medicare, in which the government defines eligibility criteria for certain benefits ("entitlements"); actual expenditures then depend on how many people meet those criteria and what those benefits cost. The government does not, on an annual basis, actually control expenditures; it can change expenditure levels only by changing the criteria for eligibility or the level of benefits. By contrast, for the **discretionary programs**, government sets the expenditure levels on an annual basis.

Thus, on an annual basis, government does not have direct control over two-thirds of its expenditures (interest, Social Security, Medicare,

Health (2.8%)
Other (2.6%)
Veterans benefits and services (3.1%)
Physical resources (2.6%)
Education, training, employment, and social services (3.7%)
National defense (20.1%)
Net interest (5.7%)
Social Security (20.4%)
Medicare (13.1%)
Medicaid (7.9%)
Income security (18.0%)
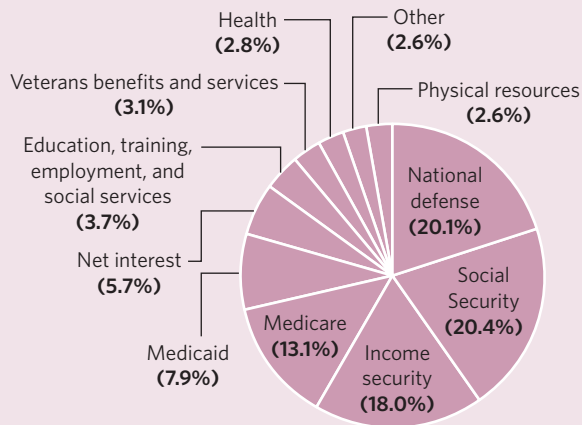
**U.S. Federal Expenditures, 2010:
($3.46 trillion)**

Medicaid, and other entitlements). Of the remainder, almost two-thirds goes to defense. This leaves only 12 percent of the budget in the "non-defense discretionary" category. Given President Reagan's commitments not to raise taxes, to increase defense expenditures, and not to cut Social Security, there simply was insufficient scope for cutting other expenditures to bring the budget into balance. Eliminating the deficit would have required cutting other categories of expenditure in half—beyond the level that seemed acceptable.

Three factors played a role in eventually taming the deficit in the 1990s: increased taxes on high-income individuals, limiting the growth of expenditures, and a strong economy. Each factor is estimated to have contributed approximately one-third of the overall deficit reduction, but the factors are, in fact, intertwined.[2]

The link between growth and deficits when the economy is at full employment is a simple one: if the government has to borrow more money, interest rates rise. Figure 28.4 shows the demand and supply for funds (capital).

---

[2] Economic recovery—the reduction of the unemployment rate from 7.8 to 4.9 percent—by itself would have generated additional revenues equal to approximately 1.2 to 1.8 percent of GDP. By what is known as *Okun's law*—named after Arthur Melvin Okun, the chairman of the Council of Economic Advisers under President Johnson, who first "discovered" this empirical regularity—each percentage reduction in the unemployment rate gives rise to a 2 to 3 percent increase in output. Each percent increase in output in turn gives rise to an increase in tax revenues of approximately 0.2 percent of GDP.
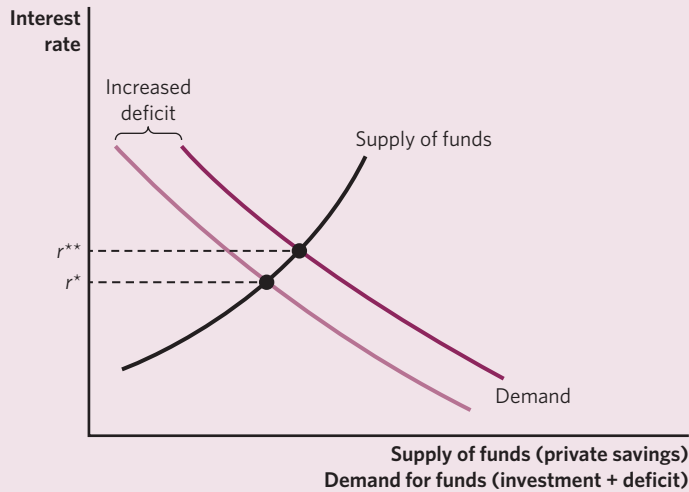
Demand and Supply Curves for Funds (Capital)
An increased deficit increases the demand for funds, leading to a higher interest rate and lower investment.

Increased government borrowing shifts the demand curve to the right, leading to higher interest rates. At higher interest rates, firms are willing to invest less. With less investment, there is less growth. We say that government borrowing is **crowding out** private investment. When there are high levels of unemployment, however, as there were in the years after the global financial crisis of 2008, matters are different, as we shall see.

The budget debates of the 1990s were complicated by two factors. In 1993, when President Clinton assumed office, the economy was still in a downturn, and it was felt that actions to tame the deficit immediately would make matters worse. This meant that there had to be great care in the timing of expenditure cuts. Additionally, there was a concern that if the deficit were cut in the wrong way, not only could poverty increase, but long-term growth could be stymied—and that would be counterproductive even to the ultimate goal of deficit reduction. Conservatives claimed that raising taxes to reduce the deficit would have even more adverse effects on growth: deficits might be bad, but increased taxes would be even worse. They argued that taxes on capital gains and interest reduce savings, lower savings lead to higher interest rates, and higher interest rates lead to less investment. As we saw in Chapter 18, though, the econometric evidence does not provide much support for this position, as savings do not seem to be very responsive to interest rates.

On the other hand, some economists claimed that reducing public investments would slow economic growth, especially in the long run. They contended that public investments are **complements** to private

investment—that is, they increase the productivity of private investment so, at any interest rate, the level of private investment will be higher. In this view, for instance, a more educated labor force makes investments in the United States more attractive; government-supported research leads to findings that provide the basis of profitable innovations by the private sector. Although those who saw a limited role for government agreed, they argued that there was plenty of room for expenditure cuts—for instance, in entitlement programs—that would not adversely affect investments, and that the productivity of many public investments was so low that cutting back on them would have little effect on growth. As evidence, they cited studies showing low returns on many job training programs.

The thrust of the Clinton administration's budgets was to redirect more public expenditures toward investment and increase the effectiveness of public expenditures by spending more money on programs that had shown high returns. By 1993, there was real concern about the economy's infrastructure such as roads and airports. During the previous twelve years, public expenditures on infrastructure as a percentage of GDP had sunk to half what they were in the 1960s. The ratio of public capital to private capital had dropped by almost a third from 1965 to 1988. By 1993, the Department of Transportation estimated that almost 20 percent of the nation's highways had poor or mediocre pavement and almost 20 percent of the bridges were structurally deficient. The percentage of the federal budget devoted to all types of public investment, including research and education, fell from 35 percent in 1963 to 17 percent in 1992.

The Clinton administration made significant inroads in increasing public investment only in one area: education. Given the concern about the size of the deficit, an unwillingness to raise taxes, and an inability or unwillingness to cut back the major areas of noninvestment expenditures like the entitlement programs—although the rate of increase of Medicare and Medicaid was reduced—there was little scope for increasing investments. To achieve its deficit reduction targets, it decided to maintain, and even slightly increase, expenditures on education, but to cut back in real terms most other areas of nondefense nonentitlement expenditures, including investments. Those who believe that the long-run growth prospects of the American economy depend on innovation, and that innovation must be based on a strong science and engineering foundation, were particularly disturbed at proposed cutbacks in support of research by between 20 and 30 percent. They argued that the marginal return from these expenditures is not only high, but also higher than many other forms of investment, including some education and training programs.

Within the research budget, there was also controversy. Most scientists questioned the returns of the high levels of investment in space—they

argued that money spent on the space station could be far better spent elsewhere. The National Aeronautics and Space Administration (NASA) had done a fantastic job of ensuring strong political support for its projects—for instance, by ensuring that contractors and subcontractors were located in a large number of congressional districts—far better than the basic research community had done, although there was bipartisan agreement that basic research should be supported. Basic research is like a public good: the marginal cost of an additional individual's enjoying the benefits of the knowledge is low, and the cost of exclusion is often high. However, the Clinton administration argued that the government should also support the development of new technology. There were substantial spillovers from many new technologies that were not captured by the inventor, as evidenced by the laser and the transistor, two innovations that had a huge range of benefits to the economy, extending well beyond what had been originally contemplated. Conservatives argued that these areas should be left to the private sector; the government should not be in the business of deciding which industries to promote, even by supporting research. They contended that the government had a bad track record of picking winners, and that such support was prone to influence by special interest groups.

The reality probably lay between the two extreme positions: even though there were many examples of misguided government research programs, government had long played a central, and successful, role in promoting technology. The first telegraph line, between Baltimore and Washington, was financed by the federal government in 1842, as was the Internet. The rapid increase in productivity in agriculture during the nineteenth and twentieth centuries was based on government-funded research, and on the government-run extension services that widely disseminated the newfound knowledge. In one way or another, much of modern technology has been supported by the government, often as part of defense efforts. In the 1990s, with the slowing down of defense expenditures, key questions facing policy makers were: Would more explicit governmental support of technology be required for the country to maintain its technological leadership? Furthermore, were there ways, such as requiring more equity participation by private firms in government-supported technology research programs, to increase the success rate of government-supported projects? Under the exigency of the budget constraints, expanding investments in new technology, which had been a major initiative, was pared back—although the programs survived at lower levels of support.

This debate has been revived today. The arguments are similar to those of the 1990s, with little resolved in the interim. For example, President

Obama has tried to direct much of the economic stimulus spending to investments in the nation's physical infrastructure and human resources to make the country more competitive, as reflected by the name of the stimulus package: the American Recovery and Reinvestment Act of 2009. His administration believed that the short-term costs are justified by the long-term returns—the focus is on increasing GDP to grow the nation out of the deficit and debt crises through the generation of more economic activity, and thus tax revenue. This approach is complemented by anticipated defense spending cuts as the wars in Iraq and Afghanistan wind down, together with increased efficiency in health care spending under the Patient Protection and Affordable Care Act of 2010.

When there is high and prolonged unemployment, government spending can actually **crowd in** (that is, increase) private spending. Earlier, we noted that the reason that government borrowing crowded out private investment was that it forced interest rates to rise. However, in the short run at least, the Fed sets the interest rate, and with a deep recession, the Fed does not respond by increasing interest rates, simply because it too is trying to restore the economy to full employment. For years after the 2008 crisis, it kept interest rates near zero, regardless of the level of government deficits.

Indeed, as output increases as a result of stimulation, firms can be induced to invest more. This is especially true if public and private investments are *complements*—that is, public investments increase the returns to private investments. Government investments in roads in the 1930s increased the returns to investing in railroads, because it meant that the areas that could be reached easily by the nation's railroads *directly or indirectly* were vastly increased. Government investments in technology that led to the Internet have been the spur to much of the investments in recent years—so, too, for government research that led to breakthroughs in medicine.

Households also may be induced to spend more if they see the economy beginning to recover—they will not have to save as much for the rainy days that otherwise would appear almost inevitable.

Critics of the 2009 stimulus package say that it did not work, because, despite the stimulus, unemployment reached 10 percent. The relevant question, however, is what economists call the *counterfactual*—what would it have been in the absence of the stimulus. Most economists believe that unemployment would have reached 12 to 13 percent.

Of course, the extent to which a particular program stimulates the economy depends on its design. Spending on foreign contractors working for the defense department in Afghanistan does not stimulate the U.S. economy much—far less than spending, say, on teachers in the United States.

Similarly, tax cuts, especially for upper-income Americans, do not stimulate the economy as much as providing unemployment benefits to the long-term unemployed. What matters is the amount of money spent and respent inside the United States. The Obama administration made a crucial mistake: it underestimated the severity of the downturn, and thus was overly optimistic about the unemployment that would result. Its stimulus was too small, too short, and not designed as well as it could have been to stimulate the economy, with too large a proportion of the stimulus coming from tax cuts.

It was important to get money quickly into the economy's bloodstream, which is why the stimulus program focused on what it called "shovel ready" investment projects. Alternatively, though, it could have provided more money for states and localities, so that they would not have been forced to cut back on teachers and universities.

Providing unemployment insurance and making up for shortfalls in state and local revenues that are a result of an economic downturn have one further advantage: they are **automatic stabilizers**—that is, they provide money to stimulate the economy only to the extent needed. If the economy recovers, state revenues recover too and unemployment falls, so the stimulus is automatically discontinued.

The Clinton administration tried to square the circle by stimulating the economy in the short run, but making commitments to cut the deficit in the longer run—commitments that it fulfilled. However, making such credible commitments is difficult, more difficult today than it was perhaps twenty years ago. Back then, there was more scope for cutting discretionary government spending. In the subsequent two decades, these have been pared back, except for the military. The focus, then, is on entitlements, which have strong popular support among voters. The most serious problems are posed by health care expenditures, but cutting back government spending without fixing the underlying problems in the health care sector will simply force households to pay more and/or result in de facto rationing of health care.

There is one more way that government can stimulate the economy without increasing the deficit—this takes advantage of a principle called the **balanced budget multiplier**, which holds that if the government simultaneously increases taxes and expenditure *by the same amount*, the economy is stimulated. In effect, the contractionary effect of the increase in taxes is less than the expansionary effect of the expenditure increases. This is especially so if the tax and expenditure changes are well designed—for instance, tax increases on upper-income Americans (who spend a smaller fraction of their income than do those at the bottom) and the expenditure increases are

on investments that are complements to private investment, and hence crowd in private expenditures. Unfortunately, much of the political debate in the United States has been going in the opposite direction—cutting taxes and cutting back expenditures, not just in tandem, but even more so, to bring down the deficit. These changes weaken macroeconomic performance.

CONSEQUENCES OF GOVERNMENT DEFICITS

When the government runs a deficit, it must borrow to pay the difference between its expenditures and its revenues. When it runs a deficit year after year, it must borrow year after year. The cumulated value of these borrowings is the *federal debt*—what the government owes. Figure 28.5 shows the real federal debt held by the public over the past 70 years. It totaled $8 trillion in 2010, equal to 63 percent of GDP. The immediate consequence of rising federal debt is that the government must pay out more and more in interest (or would unless the interest rate declines)—one of the factors that we identified earlier as itself contributing to the deficit. Net interest payments totaled 5.7 percent of total federal expenditures in 2010 (see Figure 28.3).

Economists have traditionally argued that government borrowing, just like individual borrowing, may be justified relative to the purpose for which the money is used. It makes sense to borrow to buy a house that you will live in for many years, or a car that you will drive for several years. In that way, you spread out paying for the item as you use it. It makes economic sense to borrow money for an educational degree that will lead to a higher-paying job in the future. However, if you are paying this year for your vacation from two years ago, maybe you should cut up your credit cards. Companies have a balance sheet that shows their overall financial strength. On one side of the balance sheet are assets (what the company owns), on the other side are liabilities (what it owes). If the company borrows to buy a valuable asset, its financial position improves, because the value of the assets increases more than the value of the liabilities.

Countries are in a similar situation. Borrowing to finance a road, school, or industrial project that will be used for many years may be quite appropriate. Again, the nation's balance sheet improves: although it owes more, the value of its assets (its infrastructure, its human capital) increases. Borrowing to pay for projects that are never completed (or perhaps are never even started), or borrowing to finance this year's

A

Billions of 2005 dollars

Real federal debt held by the public

B

Percentage of GDP

Federal debt held by the public
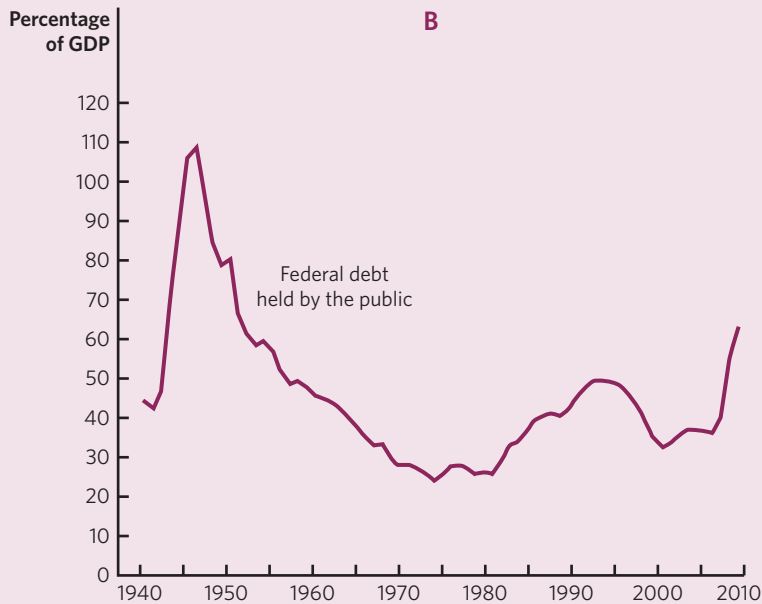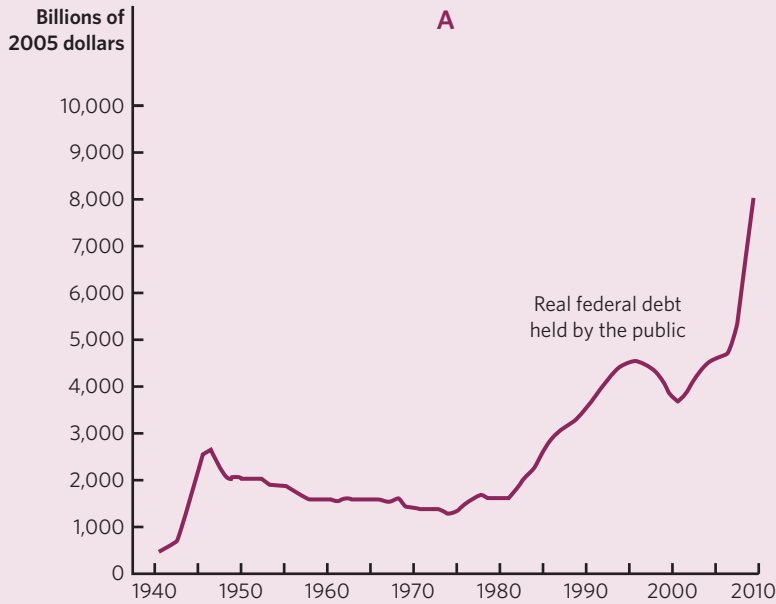
THE FEDERAL DEBT
The debt represents the accumulation of previous deficits. (A) Shows the federal debt (in real dollars) soaring enormously in the 1980s. By the early 1990s, it had far exceeded the previous record set in World War II. (B) Shows the debt as a percentage of GDP. Although the real debt has increased continuously, as a percentage of GDP it had finally started to stabilize until it began to rise sharply again in 2008.

CENTRAL GOVERNMENT DEBT AROUND THE WORLD, 2003–2010
Although the United States federal debt has risen sharply over the past few years, it is still relatively modest when compared with the central government debt of other high-income countries.
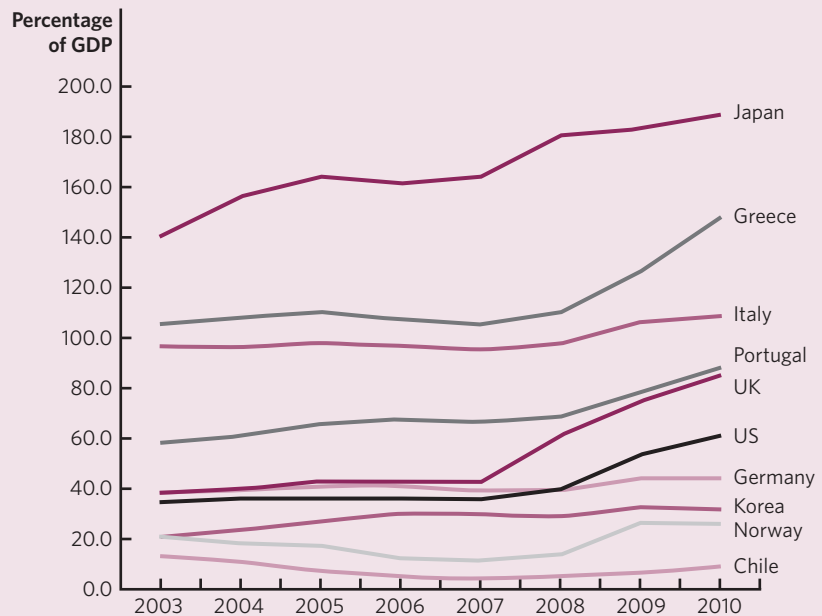
government salaries, however, poses real problems. Many governments have taken on more debt than they could comfortably pay off, forcing them to raise taxes sharply and reduce living standards. Others have simply failed to repay, jeopardizing their ability to borrow in the future, although this risk is often exaggerated.

Figure 28.6 shows central government debt around the world since 2003. Japan has the largest debt as a share of GDP among high-income countries at almost twice the size of the nation's economy—roughly half its annual budget in recent years has been financed by new borrowing. Other high-income countries with high levels of debt include Greece, Italy, Portugal, and the United Kingdom, which ranged from 86 to 148 percent of GDP in 2010. Despite current criticism, the U.S. debt to GDP is relatively moderate when compared with those of other high-income countries.

Financing government expenditures by borrowing rather than by raising taxes results in higher levels of consumption in the short run (because disposable income is higher). When the economy is at full employment, higher consumption implies that there is less room for investment. To maintain the economy at full employment without inflation, the Federal Reserve Board must increase interest rates. Deficit financing leads to lower investment, and thus, in the long run, to lower output and consumption.

Reducing the deficit has the opposite effect: it allows interest rates to fall, stimulating investment, and thus promoting economic growth and better future living standards.

However, this reasoning gets completely reversed in the midst of an economic downturn. Then, more spending can lead to a higher level of output today. There is a fundamental difference between a household that has to live within its means, and a government, like that of the United States. As the U.S. government spends more (and borrows more), more jobs are created and national income rises. This will induce more investment, especially so if the government spends its money well on investments that are anticipated to increase future growth and are complements to private investments. Thus, rather than the trade-off described earlier, increased government spending can lead to increased incomes now *and* in the future.

HOW DEFICITS AFFECT FUTURE GENERATIONS WHEN THE ECONOMY IS AT FULL EMPLOYMENT

This is, of course, markedly different from the situation when the economy is at full employment. By borrowing, the government places the burden of reduced consumption on future generations. It does this in two ways. Future output is lowered as a result of lower investment, as we have seen. Beyond this, though, some of the burden of current expenditures is put onto future generations. To see how this is done, consider the financing of World War II. This was done partly through borrowing rather than raising taxes. Suppose that the bonds the government issued were purchased by 40-year-old workers. Then, thirty years later, as these workers entered retirement, the government paid off the bonds by raising taxes on those who were then in the labor force. In effect, the government was transferring funds from these younger workers to those who were the workers during the war, those now 70 years old and retired. Thus, part of the cost of the war was borne by the generation who entered the work force after the war. The lifetime consumption of those who were 40 years old during the war was affected by the taxes they paid during the war. But that part of the cost to the war that was debt financed was shifted. They might otherwise have put their savings into stocks or bonds issued by firms; the war (to the extent it was financed by debt, or bonds) affected the form of their savings, but not the total amount they had to spend over their lifetime.

ALTERNATIVE PERSPECTIVES ON THE BURDEN OF THE DEBT

The discussion so far represents the current dominant views. Some economists believe that these views overstate the burden of the debt, even when the economy is at full employment. Four different reasons are given.

THE "DEBT DOES NOT MATTER BECAUSE WE OWE IT TO OURSELVES" ARGUMENT   It used to be argued in the United States that the fiscal deficit does not matter because we simply owe the money to ourselves. The budget deficit was compared to the effect on the total welfare of the family when one brother borrows from another. One member of the family may be better off, another worse off, but the indebtedness does not matter much to the family as a whole. Financing government expenditures by debt, it was argued, could lead to a transfer of resources between generations, but this transfer would still keep all the buying power in the hands of U.S. citizens.

We now recognize that this argument is wrong on three counts. First, even if we owe the money to ourselves, the debt affects investment and thus future wages and productivity, as noted. Second, today we do not, in fact, owe the money to ourselves. The United States is borrowing abroad and becoming indebted to foreigners. When there is full employment, the consequences of the country's spending beyond its means are no different from the consequences of a family's spending beyond its means: eventually it has to pay the price of its consumption binge. In the case of a national consumption binge, it is future generations who will have to pay the price. Third, simply to pay interest on the debt requires high levels of taxes, and taxes introduce distortions into the economy. (There is some disagreement among economists about the quantitative significance of these effects.)

OPEN ECONOMY: AVOIDING CROWDING OUT   The fact that the United States today is part of a global economy does, however, reduce the extent to which deficits crowd out private investment. For a small open economy in a world in which capital flows freely, the interest rate that its firms have to pay is determined internationally, and will be little affected by the size of its deficit. Therefore, the level of investment, and thus the rate of growth in its GDP—the output produced within the country—may not be adversely affected by deficits. To finance the investment plus an increased deficit, however, the country will have to borrow more abroad than it would with a smaller deficit. Its citizens will have to pay interest to foreigners on these borrowings, and thus the *net income* of its citizens—what they have to spend, after paying foreigners for the use of their capital; its gross national product (GNP)—will be lower. The country will, in this sense, be worse off, even though the actual level of output is unaffected.

AUSTERITY IN A RECESSION: EXPANSIONARY OR CONTRACTIONARY?

There is today a broad consensus that cutbacks in government spending when unemployment is high and interest rates are very low further weakens the economy.

Europe has instituted such cutbacks called *austerity*, or **fiscal consolidation**, with the predictable (and predicted) result that many of their economies sank back into recession; in 2012, after an anemic recovery, and five years after the onset of recession, Europe still had lower incomes than before the downturn. In the case of Greece and Spain, austerity is widely blamed for bringing on a depression, with youth unemployment exceeding 50 percent.

Historically, there are few, if any, instances of austerity bringing a country back to growth; in the few cases in which growth was restored in spite of the austerity, it was because increased exports replaced reduced government spending. When the IMF examined 173 episodes over the past thirty years during which seventeen advanced economies undertook austerity programs, it concluded that these policies are contractionary, not expansionary—they reduce incomes and raise unemployment: an austerity program of 1 percent of GDP reduces inflation-adjusted incomes by about 0.6 percent and raises the unemployment rate by almost 0.5 percent within two years. Spending by households and firms also declines.

Critics claim that the costs of increased expenditures undermine confidence in the economy because of the large budget deficits they generate, which, in turn, deter private sector investment and discourage job creation. Even though there is little empirical evidence supporting this hypothesis and considerable evidence to the contrary (expenditure cuts lead systematically to lower GDP), it is a belief that has persisted.

SOURCE: J. Guajardo, D. Leigh, and A. Pescatori, "Expansionary Austerity: New International Evidence," IMF Working Paper 11/158, Washington, DC, 2011.

The United States is a large economy; when it borrows more, there is an effect on international interest rates. It is estimated that if the United States runs a $100 billion additional deficit, slightly more than a third of that is financed by increased foreign borrowing, somewhat less than a third from increased domestic savings responding to the higher interest rates, and the rest by reduced investment—the deficits do partially crowd out private investment.

RICARDIAN EQUIVALENCE: BEQUESTS OFFSET THE DEBT   Another argument says that in the face of increased deficits, individuals save more. Robert Barro of Harvard University, developing an argument made (and later rejected) by David Ricardo, one of the nineteenth century's greatest economists, believes that individuals' concern for their children is so great that they increase their bequests when they see government deficits

threatening their offspring with future indebtedness.[3] To be able to be more generous in their bequests, they increase their household savings by exactly the amount of the increase in the deficit: national savings does not change. The increased government dissaving is fully offset. This view is called **Ricardian equivalence** because it contends that taxation and deficits are equivalent means of financing expenditures.

The evidence does not support Barro's theoretical contention. Increased government deficits may lead to slightly higher household savings, but far less than necessary to fully offset the deficit increases. In the late 1980s and early 1990s, when the deficit was running at more than 5 percent of GDP, the household savings rate was only 3 to 4 percent. If Barro's theory were correct, then in the absence of the government deficit, household savings rates would have been an implausibly low number, between –1 and –2 percent. What happened after the 2001 and 2003 tax cuts is even more telling: rather than household savings increasing, it fell dramatically, to near zero levels.

Statistical evidence from a variety of countries confirms the experience in the United States that private savings does not fully offset government borrowing. This is not surprising: individuals are not as rational as Barro assumes (fully taking into account public liabilities in current decision making), nor as altruistic (setting aside an additional dollar of bequests for their heirs every time public indebtedness increases by a dollar). In many cases, they face borrowing constraints; they would like to consume more than they do, but cannot finance their additional consumption. If the government increases their disposable income, by lowering taxes, they spend most of this increased income, even if the government has increased public indebtedness at the same time.[4]

UNDERUTILIZATION OF RESOURCES: DEFICITS MAY ACTUALLY HELP   This is the most important caveat to the view that deficits lead to lower growth, and the one most relevant to the debates that occurred as the deficit soared after the 2008 crisis. If the economy is operating at less than full employment, then under traditional macroeconomic theories, deficits, either as a result of increased government expenditures or decreased taxes, can stimulate the economy. Interest rates may not rise, or rise very much, so investment (and hence growth) will not be adversely affected. In effect, when the economy is not fully utilizing its resources, output today *and* output in the future can both be increased.

---

[3] R. Barro, "Are Government Bonds Net Wealth?" *Journal of Political Economy* 82 (1974): 1095–1117. Subsequent research has both generalized his result and shown how stringent the assumptions are under which it is valid. See, for instance, J. E. Stiglitz, "On the Relevance or Irrelevance of Public Financial Policy," in *The Economics of Public Debt,* ed. K. J. Arrow and M. J. Boskin (New York: Macmillan Press, 1988), pp. 41–76; and O. Röhn, "New Evidence on the Private Saving Offset and Ricardian Equivalence," OECD Economics Department Working Papers, OECD Publishing 762, Paris, 2010.

[4] For further discussions, see the Symposium in the *Journal of Economic Perspectives* 3, no. 2 (Spring 1989).

This is especially true, as we have seen, if the increased deficit arises from increased government expenditures on investments.

As advocates of such deficit spending emphasize, the deficit did not cause the economic downturn; it was, to a large extent, the other way around. The best way to reduce the deficit is to "put the country back to work," and the best way to do that—given the limited effectiveness of monetary policy, especially when interest rates are already near zero—is to provide for fiscal stimulus. Countries that have tried the opposite tactic such as those in East Asia, in the East Asian crisis at the end of the 1990s; Argentina, in the beginning of the century; or those in Europe more recently, cutting expenditures first, have been sorely disappointed. Not only did they go into recessions or depressions, but, as a result, the fiscal position even worsened, as measured by the debt/GDP ratio. (See case study, "Austerity in a Recession.")

> ## CONSEQUENCES OF GOVERNMENT DEFICITS
>
> 1. Some of the burden of current expenditures is shifted to future generations directly.
> 2. Issuing bonds may decrease investment and thus make future generations worse off indirectly.
> 3. Foreign indebtedness may increase, reducing future standards of living.
> 4. Government dissaving is not offset by private savings.

IMPROVING THE BUDGETARY PROCESS

The persistence of the deficit led many to argue that there should be a change in the way budgets are adopted. In this section, we consider two such changes. One of these has already been implemented; the other remains under discussion.

BUDGET ENFORCEMENT ACT AND SCORING

In 1990, Congress passed the Budget Enforcement Act (BEA), designed to reduce the likelihood of runaway deficits. Under BEA, whenever Congress passed a new program, it was required to levy new taxes to pay for it or find offsetting expenditure cuts, a principle commonly referred to as PAYGO ("pay-as-you-go"). The Congressional Budget Office (CBO), ostensibly a nonpartisan group of professionals, "scored" the program—that is, it calculated how much it would cost. Inevitably, issues of judgment were involved. Thus, in debates on health care reform in the late 1990s, the CBO argued that managed care would, in the short run, result in only limited savings. Had it scored managed care as generating large savings, it might have been far easier to adopt health care reforms. BEA was extended several times,

but finally expired in 2002. Since then, milder versions of BEA have been passed, most recently the Statutory Pay-As-You-Go Act of 2010.[5]

Further undermining fiscal discipline was the practice developed under the Bush administration of funding the wars in Afghanistan and Iraq through supplementary budgets. These made it especially hard to ascertain the total amounts that were being spent on the war, and the magnitude of the fiscal deficit.

CAPITAL BUDGETS

Businesses emphasize the distinction between expenditures to buy machines—capital expenditures—and other expenditures. Investments enhance the firm's ability to produce in the future. We argued earlier that there was a marked difference between borrowing to make an investment and borrowing to buy a vacation. Government accounting systems do not make a distinction, although businesses do. Not only do businesses distinguish between the kinds of expenditure in their annual reports, but they also maintain capital accounts that show their assets (including machines) and liabilities.

Some have argued that government should have a capital budget, which would identify expenditures on investments. Although some countries have done so, the United States has not, partially because of the political problems of defining investments. Most capital budgets include only investments in physical objects like roads, and education advocates have worried that such a capital budget would divert resources away from human capital formation, such as education and research. Similarly, health care advocates argue that providing good health care for children is an investment in their future, raising their productivity and lowering future expenditures on medical care.

OTHER STRATEGIES

Two other suggestions for improving the budgetary process have been widely discussed. One, the **line item veto** gives the President the right to veto particular expenditures within a budget bill. Without a line item veto, Presidents can accept or reject only the entire bill. They are loath to veto a large bill for, say, national defense, simply because of some small wasteful expenditure that went to benefit the constituents of some congressional representative. The Line Item Veto Act of 1996 gave President Clinton this power briefly, but

---

[5] Like BEA, the Statutory Pay-As-You-Go Act of 2010 includes PAYGO for increases in entitlement spending and decreases in taxes; but unlike BEA, it does not include discretionary spending caps, so is less effective in controlling discretionary spending.

in *Clinton v. City of New York*, 524 U.S. 417 (1998), the Supreme Court ruled that this law violated the Presentment clause of the U.S. Constitution—the clause that describes legislative procedure by which bills originating in Congress become federal laws. Despite several attempts, Congress has failed to enact alternative versions of the line item veto since this ruling.

The other suggestion is a constitutional amendment called the **balanced budget amendment**, requiring Congress to adopt a balanced budget. Such an amendment has come within a few votes of passing several times. Advocates say that it would (by definition) force fiscal responsibility; the government would have to come up with a balanced budget. Critics point out the severe problems of implementation: What would happen if Congress failed to comply, for instance because revenues fell short because of an economic recession? Balancing the budget on the basis of prospective expenditures and revenues would require someone to make estimates. Who should do that, and how would one ensure that it was done in a nonpolitical way? Many economists criticize the balanced budget amendment because it takes away one of the main tools of economic stabilization. It would make it more difficult to maintain the economy at full employment, putting almost all the responsibility on monetary policy. The crisis of 2008 showed the limits of monetary policy: even setting interest rates near zero was insufficient to restore the economy to full employment. Therefore, some of the proposals have included escape clauses in the event of a recession—although not the ones that have come close to passage by Congress.

THE LONG-TERM PROBLEM: ENTITLEMENTS AND THE AGED

Even though the United States was able to eliminate its deficit at the turn of the century, this accomplishment was short-lived. Reversing the measures that account for most of the turnaround of the U.S. fiscal position between 2001 and 2010—from large surplus to large deficit—would go a long way to ending the deficit: (a) bringing the United States back to full employment, (b) taming the defense budget, (c) reversing the tax decreases, and (d) reducing the special benefits received by drug companies.

A few "easy" measures go beyond this: (a) imposing environmental taxes (described in Chapter 6), which could raise very substantial revenues at the same time that they curb pollution and other forms of environmental degradation; (b) charging market prices for the use of publicly owned natural resources like minerals, rather than selling them at prices that are substantially below competitive levels—a giveaway worth billions to the

corporations that are the beneficiaries; (c) using more competitive procurement processes, not just for drugs, but especially for defense; (d) the reduction, or elimination, of corporate welfare, especially the special provisions hidden within the tax code that cost the government tens of billions of dollars a year and simultaneously distort the economy; and (e) enacting a "fair" tax system that would tax capital gains and dividends at the same rate that ordinary income is taxed.[6] These measures would more than eliminate the deficit, once the economy was restored to full employment.

Looking forward, though, a long-term problem persists: entitlement expenditures for Medicare and Social Security are likely to increase in the coming decades. The reasons are twofold. First, the number of elderly in the U.S. population will increase dramatically, vastly increasing the number of eligible recipients for Social Security and Medicare. Second, health care costs for the elderly may continue to increase, though at a much more moderate pace. Chapters 13 and 16 have described both these trends, and outlined some of the policies that might address them. Of the two problems, Medicare is far more serious; slight adjustments in the Social Security program—slightly higher taxes, a higher income threshold over which Social Security contributions are made, slight modifications in the cost of living adjustments, and slight adjustments in the age of retirement—would bring it back into balance. As noted in Chapter 13, the cost curve of medical care seems to have bent down; if this continues, it will reduce fiscal pressure facing Medicare. If costs start to increase again, we will be faced with a daunting challenge: fixing Medicare without addressing the underlying problems in the U.S. health care system. As we have noted, this will be extremely difficult.

At its root, the problem is that the political process tends to focus on pressing concerns, not those that might or even probably will occur in the future. There is a chance that the low rate of increase of health care costs continues or that productivity might increase faster than anticipated, thus increasing tax revenues without increasing tax rates. However, there is also a chance that health care costs might rise more rapidly than is currently anticipated, or that productivity might increase more slowly than anticipated. Thus, most economists believe that we should base our actions today on the best available information.

How the government responds to this challenge, how it brings the entitlement programs under control, and what the consequences will be for other programs if these programs are not dramatically changed are among the most important issues facing the economics of the public sector in the coming decades.

---

[6] Indeed, a CBO study showed that eliminating the preferential treatment of capital gains and dividends (the benefits of which go almost entirely to those at the top of the income distribution) would generate almost $2 trillion over ten years, beginning in 2014.

REVIEW AND PRACTICE

SUMMARY

1. The early 1980s were marked by a surge in the size of federal budget deficits.

2. The first decade of the twenty-first century saw a marked deterioration in the U.S. fiscal position. Four changes were central to this turnaround: large tax cuts, aimed particularly at upper-income Americans; two very expensive wars, and substantial increases in defense expenditures beyond that; a new drug benefit under Medicare, with a proviso that the government could not bargain with the drug companies over price; and the deep economic downturn that began in 2008.

3. The deficit was finally brought under control in the late 1990s, through increased taxes, reduced expenditures—especially on defense—and a strong economy.

4. The fiscal deficit could be brought under control in a variety of ways, going beyond reversing the measures just described. These include imposing environmental taxes, taxing capital gains and dividends at the same rate as ordinary income, and making those who get access to public resources (e.g., oil and mining companies) pay market value for these resources, reducing corporate welfare, and especially the special provisions within the tax code that lead corporations to have an effective tax rate that is far lower than the legislated tax rate.

5. The aging of the population has meant that an increasing share of government expenditures has gone to the aged in the form of entitlements, and this trend is likely to continue. These are long-run programs, and it is difficult to make forecasts a few years into the future, let alone decades. Based on current projections, the Social Security program could be put into balance with minor adjustments. The magnitude of the problems posed by health care programs depends on the rate of increase of medical care costs. The dramatic slowing down of this pace in recent years holds out the possibility that minor reforms will be able to significantly reduce, if not eliminate, the drain of the federal budget. If medical care cost inflation returns to previous high levels, restoring fiscal balance will likely be achieved only through still further reforms in the country's health care system.

6. Government borrowing can burden future generations economically in several ways. First, future generations may have to bear the burden of paying off the borrowing; there is a transfer from one generation to another. Second, government borrowing can crowd out investment, which will reduce future output and wages. Third, when money is borrowed from foreign investors, then Americans as a whole must pay some of their national income each year to foreigners just for interest, resulting in lower standards of living.

7. Government dissavings (deficits) have not been offset by private savings, as Ricardian equivalence would have suggested, and it is not true that "the debt does not matter, because we only owe it to ourselves."

8. How the deficit is reduced may have an important effect on economic growth. Some economists worry that reducing the deficit by increasing taxes will reduce savings, but most econometric evidence suggests that this effect is small. Other economists worry that reducing the deficit by decreasing public investments will hurt economic growth, both because such investments have high returns and because they lead to increased private investment.

9. When the economy is in a deep downturn, more government spending can lead to increased output in both the short run and the long run, even if it causes an increase in the deficit in the short run. Increased government spending on investments (e.g., infrastructure, education, technology) can improve the country's overall fiscal position, as its assets can increase more than its liabilities.

10. Government can stimulate the economy through a balanced budget multiplier, increasing taxes in tandem with increased spending.

KEY CONCEPTS

Automatic stabilizers

Balanced budget amendment

Balanced budget multiplier

Complements

Crowding in

Crowding out

Discretionary programs

Entitlement programs

Fiscal consolidation

Fiscal deficit

Line item veto

Ricardian equivalence

QUESTIONS AND PROBLEMS

1. True or false: "Government borrowing can transfer resources from future generations to the present, but it cannot affect the overall wealth of the country." Discuss.

2. In an open economy, there are three sources of funds to finance investment: private savings (S_p), government savings (the difference between government revenues and expenditures, S_g), and foreign borrowing, B:

$$I = S_p + S_g + B.$$

a. Suppose a certain country has private savings of 6 percent of GDP, foreign borrowing of 1 percent of GDP, and a balanced budget. What is its level of investment?

b. Suppose now that the government runs a deficit of 3 percent of GDP, and investment and private savings remain unchanged. What must happen to foreign borrowing?

c. Suppose now that the government runs a deficit of 3 percent of GDP, and this increases

the interest rate. If this, in turn, leads to increased private savings to 7 percent of GDP and reduced investment from 7 to 6 percent of GDP, what happens to foreign borrowing?

3. In a closed economy, private savings plus government savings must equal investment:

$$I = S_p + S_g.$$

a. Draw a savings–investment diagram to show the effect of an increased budget deficit.

b. Assume that the increased government deficit is the result of increased government investment, and that the government investment increases the returns on private investment, so at each interest rate there is increased private investment. Show the effect of the deficit on investment and interest rates now.

c. Assume that Ricardian equivalence holds. What happens to national savings when the government increases the deficit? What happens to private savings?

d. Ricardian equivalence focuses on the impact of future tax liabilities associated with increased deficits on private savings. What implications does the logic of Ricardian equivalence have for the consequences of additional public investment? In particular, what will be the impact on private savings of an increased deficit used to finance public investment which has a very high return, which will be realized only in the lives of the next generation?

4. Consider a small open economy with a perfectly elastic supply of savings at interest rate r^*, which is below the interest rate that would have prevailed in the economy if it could not have borrowed, and a normal investment curve. Show the market equilibrium, including the magnitude of foreign borrowing.

a. Now show the effect of an increased budget deficit.

b. Show the effect of an increased budget deficit that originates as a result of increased public investment, in which the public investment is a complement to private investment.

5. The real value of the debt (in 1995 dollars) in 1995 was $3.6 trillion and in 1980 was $1.3 trillion. Assume that 50 percent of this buildup in the debt would have gone into investment, so the capital stock would have been $1.15 trillion larger in 1995. Assume that the capital stock in 1995 was $15 trillion. By how much lower was GDP in 1995 as a result of the displaced investment, if an increase in the capital stock by 1 percent leads to a 0.2 percent increase in GDP?

6. One proposal to contain the deficits is a balanced budget amendment to the Constitution, that would require the government to maintain a balanced budget every year.

   a. Explain why, if the economy goes into a downturn, this would either force the government to raise taxes or reduce expenditures.

   b. What would be the consequences of raising taxes or reducing expenditures under these circumstances?

   c. How might you design a balanced budget amendment that would not force the government to cut back expenditures or raise taxes in an economic downturn?

   d. Explain why, if interest rates are increased to dampen the economy, the balanced budget amendment might force the government to cut back other expenditures (or raise taxes). More generally, describe the interaction between monetary and fiscal policy under a balanced budget amendment.

7. Should expenditures on health care for children be considered an "investment" in a capital budget? What other expenditures might legitimately be classified as "investment"?

8. Assume the government sells some of the land it currently rents out for grazing. Should the revenues received be counted as reducing the deficit? How would the transaction be treated under a capital budget?

9. Why does the way government achieves deficit reduction make a difference for economic growth?

   a. Assuming that the government raises taxes on interest income, what is the total impact on national savings, taking into account both deficit reduction and changes in private savings? What is the impact on investment? (Hint: How does your answer depend on the interest elasticity of savings?)

   b. Assuming that the government raises taxes on wage income, what is the impact on GDP in the short run? (Hint: How does your answer depend on the elasticity of labor supply?) What is the impact on the deficit and on investment?

   c. Assume that the government reduces expenditures on investments in technology and education, which yield a return of 15 percent, whereas the marginal return on private investment is 7 percent. What is the impact on growth?

10. Throughout the text, we have emphasized how different government actions affect various groups in the population differently.

   a. Assume the government is considering whether to finance a war by borrowing or by taxes. Describe how different groups are affected by the two alternatives.

   b. Assume the government is considering whether to reduce the deficit by increasing capital gains taxes or by increasing payroll taxes. Describe how different groups are affected by the alternatives.

REFERENCES

Aaron, H. J., "Lessons for Tax Reform," in J. Slemrod (ed.), *Do Taxes Matter: The Impact of the Tax Reform Act of 1986* (Cambridge, MA: MIT Press, 1990), pp. 321–31.

Aber, L., and Rawlings, B., "North–South Knowledge Sharing on Incentive-Based Conditional Cash Transfer Programs," Special Protection & Labor Discussion Paper No. 1101, Washington, DC, World Bank, January 2011.

Adema, W., Fron, P., and Ladaique, M., "Is the European Welfare State Really More Expensive? Indicators on Social Spending, 1980–2012," OECD Social Employment and Migration Working Paper, France, November 2011.

Afonso, A., Schuknecht, L., and Tanzi, V., "Public Sector Efficiency: An International Comparison," European Central Bank Working Paper No. 242, Frankfurt am Main, 2003.

Akerhielm, K., "Does Class Size Matter?" *Economics of Education Review* 13, no. 3 (September 1995): 229–41.

Akerlof, G., "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* 84 (1970): 488–500.

Akerlof, G., Yellen, J., and Katz, M., "An Analysis of Out-of-Wedlock Childbearing in the United States," *Quarterly Journal of Economics* 111, no. 2 (May 1996): 277–317.

Allen, F., "Optimal Linear Income Taxation with General Equilibrium Effects on Wages," *Journal of Public Economics* 17 (1982): 135–43.

Allen, J. E., and Simms, M. C., "Is a New Yardstick Needed to Measure Poverty?" *Focus* (February 1990): 6–8.

American Civil Liberties Union, "Banking on Bondage: Private Prisons and Mass Incarceration," November 2, 2011.

Arnott, R., and Stiglitz, J. E., "Aggregate Land Rents, Expenditure on Public Goods and Optimal City Size," *Quarterly Journal of Economics* 93 (1979): 472–500.

Arnott, R. and Stiglitz, J. E., "Aggregate Land Rents and Aggregate Transport Costs," *Economic Journal* 91 (1981): 331–47.

Arriagada, R., and Perrings, C., "Paying for International Environmental Public Goods," Ecosystem Services Economics Working Paper No. 17, United Nations Environment Program, October 2011.

Arrow, K. J., *Social Choice and Individual Values* (New York: Wiley, 1963).

Arrow, K. J., "Uncertainty and Welfare Economics of Medical Care," *American Economic Review* 53 (1963): 941–73.

Arrow, K. J., "Higher Education as a Filter," *Journal of Public Economics* 2 (1973): 193–216.

Arrow, K. J., et al., "Is There a Role for Benefit-Cost Analysis in Environmental, Health, and Safety Regulation?" *Environment and Development Economics* 2, no. 2 (May 1997): 196–201.

Arrow, K. J., et al., "How Should Benefits and Costs Be Discounted in an Intergenerational Context? The Views of an Expert Panel," Resources for the Future Discussion Paper No. 12–53, December 2012.

Ashenfelter, O., "Measuring the Value of a Statistical Life: Problems and Prospects," Institute for the Study of Labor (IZA) Discussion Paper No. 1911, January 2006.

Asquith, P., and Mullins, D., "Equity Issues and Offering Dilution," *Journal of Financial Economics,* 15, no. 1–2 (January–February 1986): 61–89.

Atkinson, A. B., and Stiglitz, J. E., "The Structure of Indirect Taxation and Economic Efficiency," *Journal of Public Economics* (1972): 97–119.

Atkinson, A. B., and Stiglitz, J. E., "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics* 6 (1976): 55–75.

Atkinson, A. B., and Stiglitz, J. E., *Lectures in Public Economics* (New York: McGraw-Hill, 1980).

Auerbach, A. J., "Wealth Maximization and the Cost of Capital," *Quarterly Journal of Economics* (August 1979): 433–66.

Auerbach, A. J., "The Theory of Excess Burden and Optimal Taxation," in Alan J. Auerbach and Martin Feldstein (eds.), *Handbook of Public Economics,* vol. 1 (New York: Elsevier, 1987), pp. 100–101.

Auten, G. E., Cilke, J. M., and Randolph, W. C., "The Effects of Tax Reform on Charitable Contributions," *National Tax Journal* 45, no. 3 (September 1992): 267–90.

Bagwell, L. S., and Shoven, J. B., "Cash Distributions to Shareholders," *Journal of Economic Perspectives* (Summer 1989): 129–40.

Baicker, K., and Chandra, A., "Defensive Medicine and Disappearing Doctors," *Regulation* (Fall 2005).

Baker, L., "HMOs and Fee-for-Service Health Care Expenditures: Evidence from Medicare," NBER Working Paper 5360, November 1995, p. 27.

Barr, N., Diamond, P., and Engel, E., "Reforming Pensions: Lessons from Economic Theory and Some Policy Directions, with Comment," *Economía* 11 (Fall 2010): 1–23.

Barro, R., "Are Government Bonds Net Wealth?" *Journal of Political Economy* 82 (1974): 1095–1117.

Barro, R. J., *The Impact of Social Security on Private Savings* (Washington, DC: American Enterprise Institute, 1978).

Baumol, W. J., "Contestable Markets: An Uprising in the Theory of Industry Structure," *American Economic Review* 72 (1982): 1–15.

Baumol, W. J., Panzar, J., and Willig, R., *Contestable Markets and the Theory of Industrial Organization* (New York: Harcourt Brace Jovanovich, 1982).

Becker, G., *Human Capital: A Theoretical and Empirical Analysis with Special References to Education* (New York: National Bureau of Economic Research; Columbia University Press, 2d ed., 1975).

Behn, R. D., "Designing PerformanceStat," *Public Performance and Management Review* 32, no. 2 (December 2008): 206–235.

Belfield, C. R., and Levin, H. M. (eds.), *The Price We Pay: Economic and Social Consequences of Inadequate Education* (Washington, DC: Brookings Institution, 2007).

Berends, M., and Walberg, H., *Charter School Outcomes* (Hillsdale, NJ: Lawrence Erlbaum Associates, 2008).

Bewley, T., "A Critique of Tiebout's Theory of Local Public Expenditures," *Econometrica* 49 (1981): 713–40.

Blackorby, C., and Donaldson, D., "Cash versus In-Kind, Self-Selection and Efficient Transfers," *American Economic Review* (September 1988): 691–700.

Blank, R. M., "When Can Public Policy Makers Rely on Private Markets? The Effective Provision of Social Services," *The Economic Journal* 110 (2000): 34–49.

Blouin, J., Raedy, J., and Shackelford, D., "Did Firms Substitute Dividends for Share Repurchases after the 2003 Reductions in Shareholder Tax Rates?" National Bureau of Economic Research Working Paper 13601, November 2007.

Blumenthal, M., and Slemrod, J., "The Compliance Cost of the U.S. Individual Income Tax System: A Second Look after Tax Reform," *National Tax Journal* 45, no. 2 (June 1992): 185–202.

Bondonio, D., and Greenbaum, R. T., "Do Local Tax Incentives Affect Economic Growth? What Mean Impacts Miss in the Analysis of Enterprise Zone Policies,"

Regional Science and Urban Economics 37 (2007): 121–136.

Borzutzky, S., "Reforming the Reform: Attempting Social Solidarity and Equity in Chile's Privatized Social Security System," *Journal of Policy Practice* 11 (2012): 77–91.

Bosworth, B., and Burtless, G., "Effects of Tax Reform on Labor Supply, Investment, and Saving," *Journal of Economic Perspectives* (Winter 1992): 3–25.

Boteux, M., "On the Management of Public Monopolies Subject to Budgetary Constraints," *Journal of Economic Theory* 3 (1971): 219–40.

Bovenberg, A. L., and Goulder, L. H., "Optimal Environmental Taxation in the Presence of Other Taxes: General-Equilibrium Analyses," *American Economic Review* 86, no. 4 (September 1996): 985–1000.

Boyle, M. H., Torrance, G. W., Sinclair, J. C., and Horwood, S. P., "Economic Evaluation of Neonatal Intensive Care of Very-Low-Birth-Weight Infants," *New England Journal of Medicine* 308 (1983): 1330–37.

Bradford, D., "The Incidence and Allocation Effect of a Tax on Corporate Distributions," *Journal of Public Economics* 15 (1981): 1–22.

Break, G., "Income Taxes and Incentives to Work: An Empirical Study," *American Economic Review* 47 (1957): 529–49.

Brenzel, L., "Brazil—Bolsa Familia Project: Results-Based Financing (RBF) for Health—Country Snapshot," Washington, DC, World Bank, November 2009.

Breyer, S. G., *Breaking the Vicious Circle: Toward Effective Risk Regulation* (Cambridge, MA: Harvard University Press, 1993), p. 12.

Bridge, G., "Citizen Choice in Public Services: Voucher System," in E. S. Savas (ed.), *Alternatives for Delivering Public Services* (Boulder, CO: Westview Press, 1977).

Brito, D. L., et al., "Randomization in Optimal Tax Schedules," *Journal of Public Economics* 56, no. 2 (February 1995): 189–223.

Bruce, J., Hoesung, L., and Haites, E. (eds.), *Climate Change, 1995: Economic and Social Dimensions of Climate Change*, Contribution of Working Group III to the Second Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge, UK, and New York: Cambridge University Press, 1996).

Brueckner, J. K., "A Test for the Allocative Efficiency in the Local Public Sector," *Journal of Public Economics* 19 (1982): 311–31.

Buchanan, J. M., "Tax Reform as Political Choice," *Journal of Economic Perspectives* 1 (Summer 1987): 29–35.

Burch, P., "After the Fall: Educational Contracting in the USA and the Global Financial Crisis," *Journal of Education Policy* 25 (November 2010): 757–66.

Burman, L., and Randolph, W., "Measuring Permanent Responses to Capital Gains Tax Changes in Panel Data," *American Economic Review* 84, no. l (September 1994): 794—809.

Busso, M., and Kline, P., "Do Local Economic Development Programs Work? Evidence from the Federal Empowerment Zone Program," Cowles Foundation Discussion Paper No. 1638, February 2008.

Card, D., and Krueger, A., "Does School Quality Matter? Returns to Education and the Characteristics of Public School in the United States," *Journal of Political Economy* 100 (February 1992): 1–40.

Card, D., and Krueger, A., "Labor Market Effects of School Quality: Theory and Evidence," in Gary Burtless (ed.), *Does Money Matter: The Effect of School Resource on Student Achievement and Adult Success* (Washington, DC: Brookings Institution, 1996), pp. 97–140.

Card, D., and Krueger, A., "School Resources and Student Outcomes," *Annals of the American Academy of Political and Social Science* 559 (September 1998): 39–53.

Card, D., and Krueger, A. B., *Wages, School Quality, and Employment Demand* (New York: Oxford University Press, 2011).

Carraro, C., Galeotti, M., and Gallo, M., "Environmental Taxation and Unemployment: Some Evidence on the 'Double Dividend Hypothesis' in Europe," *Journal of Public Economics* 62, no. 1–2 (October 1996): 141–81.

Carson, R., "Contingent Valuation: A Practical Alternative When Prices Aren't Available," *Journal of Economic Perspectives* 26 (Fall 2012): 3–56.

Chaikind, H., Copeland, C. W., Redhead, C. S., and Staman, J., *PPACA: A Brief Overview of the Law, Implementation, and Legal Challenges*, Congressional Research Service, March 2011.

Chapman, B., "Income Contingent Loans for Higher Education: International Reforms," Chapter 25 in E. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, vol. 2 (Amsterdam: Elsevier, 2006), pp. 1435–503.

Chetty, R., and Saez, E., "Dividend and Corporate Taxation in an Agency Model of the Firm," *American Economic Journal: Economic Policy* (August 2010): 1–31.

Choe, Y. S., and Jeong, J., "Charitable Contributions by Low- and Middle-Income Taxpayers: Further Evidence with a New Method," *National Tax Journal* 46, no. 1 (March 1993): 33–39.

Cliff, S., "Our Health Care Spending, in One Map," *Washington Post Wonkblog*, December 8, 2011.

Clotfelter, C. T., and Schmalbeck, R. L., "The Impact of Fundamental Tax Reform on Nonprofit Organizations," in Henry J. Aaron and William G. Gale (eds.), *Economic Effects of Fundamental Tax Reform* (Washington, DC: Brookings Institution Press, 1996), pp. 211–43.

Coase, R. H., "The Problem of Social Cost," *Journal of Law and Economics* 3 (1960): 1–44.

Cohen, T. H., and Hughes, K. A., *Medical Malpractice Insurance Claims in Seven States, 2000–2004*, NCJ Report 216339, Bureau of Justice Statistics, Department of Justice, March 2007.

Coleman, J., et al., *Equality of Educational Opportunity* (Washington, DC: Department of Health, Education, and Welfare, 1966).

Coleman, J., Hoffer, T., and Kilgore, S., *Public and Private Schools: An Analysis of High School and Beyond* (Washington, DC: U.S. Department of Education, National Center for Educational Statistics, November 1981).

Common, M., Reid, I., and Blaney, R., "Do Existence Values for Cost-Benefit Analysis Exist?" *Journal of Environmental and Resource Economics* 9, no. 2 (1997): 225–38.

Corlett, W. J., and Hague, D. C., "Complementarity and the Excess Burden of Taxation," *Review of Economic Studies* 21 (1953): 21–30.

Courant, P. N., Gramlich, E. M., and Rubinfeld, D. L., "The Stimulative Effect of Intergovernmental Grants: Or Why Money Sticks Where It Hits," in P. Mieszkowski and W. Oakland (eds.), *Fiscal Federalism and Grants-in-Aid* (Washington, DC: Urban Institute, 1979), pp. 5–21.

Cropper, M. L., and Oates, W. E., "Environmental Economics: A Survey," *Journal of Economic Literature* 3 (June 1992): 675–740.

Cummings, S. R., et al., "The Cost-Effectiveness of Counseling Smokers to Quit," *Journal of the American Medical Association* 261 (1989): 75–79.

Currie, J., "Welfare and the Well-Being of Children: The Relative Effectiveness of Cash and In-Kind Transfers," in James M. Poterba (ed.), *Tax Policy and the Economy,* vol. 8 (Cambridge, MA: MIT Press for the National Bureau of Economic Research, 1994), pp. 1–43.

Cutler, D., "A Guide to Health Care Reform," *Journal of Economic Perspectives* 8, no. 3 (Summer 1994): 13–29.

Danzinger, S., Haveman, R., and Plotnick, R., "How Income Transfers Affect Work, Savings, and Income Distribution," *Journal of Economic Literature* (September 1981).

Dasgupta, P., and Stiglitz, J. E., "Differential Taxation, Public Goods and Economics Efficiency," *Review of Economics Studies* 39 (1971): 151–74.

Dasgupta, P., and Stiglitz, J. E., "Benefit-Cost Analysis and Trade Policies," *Journal of Political Economy* 82 (January–February 1974): 1–33.

Daves, D. W., and Christensen, L. R., "The Relative Efficiency of Public and Private Firms in a Competitive Environment: The Case of Canadian Railroads," *Journal of Political Economy* 88 (1980): 958–76.

Dávila, J., Eeckhout, J., and Martinelli, C., "Bargaining over Public Goods," *Journal of Public Economic Theory* 11 (December 2009): 927–45.

Deaton, A., and Muellbauer, J., *Economics and Consumer Behavior* (London: Cambridge University Press, 1980), pp. 37–41.

de Brauw, A., et al., "The Impact of Bolsa Família on Education and Health Outcomes in Brazil," paper by the International Food Policy Research Institute presented at the Second Generation of CCTs Evaluations Conference, World Bank, October 2011.

DeBrock, L., and Arnold, R. J., "Utilization Control in HMOs," *Quarterly Review of Economics and Finance* 32, no. 3 (Autumn 1992).

Dee, T., and Jacob, B., "The Impact of No Child Left Behind on Student Achievement," National Bureau of Economic Research Working Paper 15531, November 2009.

De Neve, J. E., "The Median Voter Data Set: Voter Preferences across 50 Democracies," *Electoral Studies* 30 (December 2011): 865–87.

Devaney, B., and Fraker, T., "The Effect of Food Stamps on Food Expenditures," *American Journal of Agricultural Economics* (February 1989): 99–104.

Diamond, P., "A Many-Person Ramsey Tax Rule," *Journal of Public Economics* 4 (1975): 335–42.

Diamond, P., and Mirrlees, J., "Optimal Taxation and Public Production," *American Economic Review* 61 (1971): 261–78.

Diamond, P., and Saez, E., "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives* 25 (Fall 2011): 165–90.

Dilger, R. J., and Beth, R. S., *Unfunded Mandates Reform Act: History, Impact, and Issues* (Washington, DC: Congressional Research Service, June 2014).

Domar, E. D., and Musgrave, R. A., "Proportional Income Taxation and Risk-Taking," *Quarterly Journal of Economics* 58 (1944): 388–422.

Dorman, P., *Markets and Mortality: Economics, Dangerous Work, and the Value of Human Life* (Cambridge, UK: Cambridge University Press, 1996).

Duffield, J., "Nonmarket Valuation and the Courts: The Case of Exxon *Valdez*," *Contemporary Economic Policy* 15, no. 4 (October 1997): 98–110.

Dumez, H., and Jeunemaitre, A., "Privatization in France: 1983–1993," in Vincent Wright (ed.), *Industrial Privatization in Western Europe: Pressures, Problems, and Paradoxes* (London and New York: Pinter Publishers, 1994), pp. 83–105, 194.

Eberts, R. W., and Gronberg, T. J., "Jurisdictional Homogeneity and the Tiebout Hypothesis," *Journal of Urban Economics* 10 (1981): 227–39.

"Economic Valuation of the Environment: A Special Issue," *Environmental Science & Technology* 34, no. 8 (April 2000): 1381–461.

Eddy, D. M., "Screening for Breast Cancer," *Annals of Internal Medicine* 111 (1989): 389–99.

Eddy, D. M., "Screening for Cervical Cancer," *Annals of Internal Medicine* 113 (1990): 214–26.

Edwards, J., et al. (eds.), *Recent Developments in Corporate Finance* (New York and Melbourne: Cambridge University Press, 1986).

Eissa, N., "Labor Supply and the Economic Recovery Tax Act of 1981" in Martin Feldstein and James Poterba (eds.), *Empirical Foundations of Household Taxation* (Chicago and London: University of Chicago Press, 1996), pp. 5–32.

Eissa, N., and Liebman, J., "Labor Supply Responses to the Earned Income Tax Credit," *Quarterly Journal of Economics* 111, no. 2 (May 1996): 603–37.

Eissa, N., and Hoynes, H. W., "Behavioral Responses to Taxes: Lessons from the EITC and Labor Supply," in J. M. Poterba (ed.), *Tax Policy and the Economy*, vol. 20 (Cambridge, MA: MIT Press, 2006), pp. 73–110.

Ellwood, D., and Summers, L., "Poverty in America: Is Welfare the Answer or the Problem?" in S. Danziger and D. Weinberg (eds.), *Fighting Poverty: What Works and What Doesn't* (Cambridge, MA: Harvard University Press, 1986), pp. 78–105.

Elvery, J. A., "The Impact of Enterprise Zones on Resident Employment," *Economic Development Quarterly* 23 (February 2009): 44–59.

Erard, B., "Taxation with Representation: An Analysis of the Role of Tax Practitioners in Tax Compliance," *Journal of Public Economics* 52, no. 2 (September 1993): 163–97.

Estrin, S., Hanousek, J., Kocenda, E., and Svejnar, J., "The Effects of Privatization and Ownership in Transition Economies," *Journal of Economic Literature* 47 (2009): 699–728.

Fazzari, S., Hubbard, G., and Peterson, B., "Investment Financing Decisions and Tax Policy," *American Economic Review* 78, no. 2 (May 1988): 200–205.

Feenberg, D. R., and Rosen, H. S., "Recent Developments in the Marriage Tax," National Bureau of Economic Research Working Paper 4705, April 1994.

Feige, E., and Cebula, R., *America's Unreported Economy: Measuring the Size, Growth and Determinants of Income Tax Evasion in the U.S.*, MPRA Paper No. 34781, September 2011.

Feldstein, M., "Distributional Equity and the Optimal Structure of Public Prices," *American Economic Review* 62 (1973): 32–36.

Feldstein, M., "On the Optimal Progressivity of the Income Tax," *Journal of Public Economics* 2 (1973): 357–76.

Feldstein, M., "Social Security, Induced Retirement, and Aggregate Capital Accumulation," *Journal of Political Economy* 82, no. 5 (1974): 905–26.

Feldstein, M., "The Effect of Marginal Tax Rates on Table Income: A Panel Study of the 1986 Tax Reform Act," *Journal of Political Economy* 1–3, no. 3 (1995): 551–72.

Feldstein, M., "Social Security and Saving: New Time Series Evidence," *National Tax Journal* 49, no. 2 (June 1996): 151–64.

Feldstein, M., Slemrod, J., and Yitzhake, S., "The Effects of Taxing on Selling and Switching of Common Stock and the Realization of Capital Gains," *Quarterly Journal of Economics* 94 (1980): 777–91.

Fisher, E., et al., "Slowing the Growth of Health Care Costs—Lessons from Regional Variation," *New England Journal of Medicine* 360 (February 2009): 849–52.

Fisher, G. M., "The Development and History of the U.S. Poverty Thresholds—A Brief Overview," *GSS/ SSS Newsletter* (Winter 1997): 6–7.

Foley, D. K., "Resource Allocation and the Public Sector," *Yale Economic Essays* 7 (1967): 45–98.

Fox, M. K., Hamilton, W., and Lin, B. (eds.), "Literature Review," in *Effects of Food Assistance and Nutrition Programs on Nutrition and Health*, vol. 3 (Washington, DC: U.S. Department of Agriculture, Economic Research Service, Food Assistance and Nutrition Research Program, 2004).

Fraker, T., *The Effects of Food Stamps on Food Consumption: A Review of the Literature* (Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Center, Office of Analysis and Evaluation, 1990).

Fraker, T., Martini, A., and Ols, J., "The Effect of Food Stamp Cashout on Food Expenditures: An Assessment of the Findings from Four Demonstrations," *Journal of Human Resources* 30, no. 4 (Fall 1995): 633–49.

Friedman, M., *Capitalism and Freedom* (Chicago: University of Chicago Press, 1962).

Frydman, R., Gray, C., Hessel, M., and Rapaczynski, A., "When Does Privatization Work? The Impact of Private Ownership on Corporate Performance in the Transitional Economies," *The Quarterly Journal of Economics* 114 (1999): 1153–91.

Fuchs, V., *Who Shall Live? Health Economics and Social Choice* (New York: Basic Books, 1975).

Fuchs, V., "The Supply of Surgeons and the Demand for Operations," in V. Fuchs (ed.), *The Health Economy* (London and Cambridge: Harvard University Press, 1986), pp. 126–47.

Fullerton, D., and Henderson, Y. K., "Incentive Effects of Taxes on Income from Capital: Alternative Policies in the 1980s," in C. R. Hulten and I. V. Sawhill (eds.), *The Legacy of Reaganomics: Prospects for Long-Term Growth* (Washington, DC: The Urban Institute Press, 1984).

Garber, A. M., and Phelps, C. E., "Economic Foundations of Cost-Effectiveness Analysis," *Journal of Health Economics* 16, no. 1 (1997): 1–31.

Garber, A. M., and Phelps, C. E., "Future Costs and the Future of Cost-Effectiveness Analysis," *Journal of Health Economics* 27, no. 4 (2008): 819–21.

Gaynor, M., and Town, R., "Competition in Health Care Markets," Chapter 9 in M. Pauly, T. McGuire, and P. Barros (eds.), *Handbook of Health Economics*, vol. 2 (Waltham, MA: North Holland, 2012), pp. 499–638.

Glaeser, E. L., and Gottlieb, J. D., "The Economics of Place-Making Policies," *Brookings Papers on Economic Activity* (2008): 155–239.

Glass, G. V., *School Class Size: Research and Policy* (Beverly Hills, CA: Sage Publications, 1982), p. 47.

Glaze, L. E., "Correctional Populations in the United States, 2009," *Bureau of Justice Statistics Bulletin* (December 2010).

Goldman, L., et al., "Cost-Effectiveness of HMG-CoA Reductase Inhibition for Primary and Secondary Prevention of Coronary Heart Disease," *Journal of the American Medical Association* 265 (1991): 1145–51.

Gomez-Ibanez, J. A., "Alternatives to Infrastructure Privatization Revisited: Public Enterprise Reform from the 1960s to the 1980s," Policy Research Working Paper 4391, Washington, DC, World Bank, 1997.

Goodsell, C. T., *The Case for Bureaucracy* (Chatham, NJ: Chatham House Publishers, 1983).

Goulder, L. H., "Environmental Taxation and the Double Dividend: A Reader's Guide," *International Tax and Public Finance* 2 (August 1995): 157–83.

Gramlich, E. M., "Intergovernmental Grants: A Review of the Empirical Literature," in W. E. Oates (ed.), *The Political Economy of Fiscal Federalism* (Lexington, MA: Lexington Books, 1977).

Gravelle, J., *The Economic Effects of Taxing Capital Income* (Cambridge and London: MIT Press, 1994).

Gravelle, J. G., "International Corporate Tax Rate Comparisons and Policy Implications," Congressional Research Service Report R41743, December 28, 2012.

Greenberg, D., and Halsey, H., "Systematic Misreporting and Effects of Income Maintenance Experiments on Work Effort: Evidence from the Seattle–Denver Experiments," *Journal of Labor Economics* 1, no. 4 (October 1983): 380–407.

Greene, J., et al., *Expanding Choice in Elementary and Secondary Education: A Report on Rethinking the Federal Role in Education* (Washington, DC: Brown Center on Education Policy at Brookings, February 2010).

Greenfield, H. I., *Invisible, Outlawed and Untaxed: America's Underground Economy* (Westport, CT: Praeger, 1993).

Greenwald, B., and Stiglitz, J. E., "Externalities in Economies with Imperfect Information and Incomplete Markets," *Quarterly Journal of Economics* (May 1986): 229–64.

Greenwald, B., Stiglitz, J. E., and Weiss, A., "Informational Imperfections in the Capital Market and Macroeconomic Fluctuations," *American Economic Review* 74, no. 2: 194–99.

Grimmett, R. F., "Conventional Arms Transfers to Developing Nations, 2003–2010," Congressional Research Service, September 2011, Table 31.

Grimsey, D., and Lewis, M. K., *Public Private Partnerships: The Worldwide Revolution in Infrastructure Provision and Project Finance* (Cheltenham, UK: Edward Elgar, 2004).

Grinols, E. L., "The Link between Domestic Investment and Domestic Savings in Open Economies: Evidence from Balanced Stochastic Growth," *Review of International Economics* 4, no. 2 (June 1996): 119–40.

Grossman, S. J., and Stiglitz, J. E., "On the Impossibility of Informationally Efficient Markets," *The American Economic Review* 70, no. 3 (June 1980): 393–408.

Gruber, J., "The Tax Exclusion for Employer-Sponsored Health Insurance," National Bureau of Economic Research Working Paper 15766, February 2010.

Guajardo, J., Leigh, D., and Pescatori, A., "Expansionary Austerity: New International Evidence," IMF Working Paper 11/158, Washington, DC, 2011.

Guislain, P. *The Privatization Challenge: A Strategic, Legal, and Institutional Analysis of International Experience* (Washington, DC: World Bank, 1997).

Hannaway, J., and Carnoy, M. (eds.), *Decentralization and School Improvement* (San Francisco: Jossey-Bass, 1993).

Hanushek, E. A., "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature* 24, no. 3 (September 1986): 1141–77.

Hanushek, E. A., "School Resources and Student Performance," in Gary Burtless (ed.), *Does Money Matter: The Effect of School Resource on Student Achievement and Adult Success* (Washington, DC: Brookings Institution, 1996), pp. 43–73.

Hanushek, E., and Jorgenson, D. (eds.), *Improving America's Schools: The Role of Incentives* (Washington, DC: National Research Council, 1996).

Hanushek, E., and Woessmann, L., "Symposium on Performance Pay for Teachers," *Economics of Education Review* 30 (2011).

Harberger, A., "Taxation, Resource Allocation and Welfare," in J. Due (ed.), *The Role of Direct and Indirect Taxes in the Federal Revenue System* (Princeton, NJ: Princeton University Press, 1964); reprinted in A. Harberger, *Taxation and Welfare* (Chicago: University of Chicago Press, 1974).

Hausman, J., "Labor Supply," in H. Aaron and J. Pechman (eds.), *How Taxes Affect Economic Behavior* (Washington, DC: Brookings Institution, 1981), pp. 27–72.

Hausman, J., "Exact Consumer's Surplus and Deadweight Loss," *American Economic Review* 77 (1981): 662–76.

Hausman, J., and Poterba, J., "Household Behavior and the Tax Reform Act of 1986," *Journal of Economic Perspectives* (Summer 1987): 101–20.

Hausman, J., "Contingent Valuation: From Dubious to Hopeless," *Journal of Economic Perspectives* 26 (Fall 2012): 3–56.

Hayward, N., and Kuper, G., "The National Economy and Productivity in Government," *Public Administration Review* 38 (1978).

Heckman, J. J., "What Has Been Learned about Labor Supply in the Past Twenty Years?" *American Economic Association Papers and Proceedings* (May 1993): 116–21.

Heckman, J. J., "Comment on Labour Supply and the Economic Recovery Tax Act of 1981," in M. Feldstein and J. Poterba (eds.), *Empirical Foundations of Household Taxation* (Chicago and London: University of Chicago Press, 1996), pp. 32–38.

Hellman, T., and Stiglitz, J. E., "A Unifying Theory of Credit and Equity Rationing in Markets with Adverse Selection," *European Economic Review* 44, no. 2 (February 2000): 281–304.

Hill, P., Pierce, L., and Guthrie, J., *Reinventing Public Education: How Contracting Can Transform America's Schools* (Chicago: University of Chicago Press, 1997).

Hirschman, A. O., *Shifting Involvements: Private Interest and Public Action* (Princeton, NJ: Princeton University Press, 1982).

Hitch, C. J., *Decision Making for Defense* (Berkeley: University of California Press, 1966), pp. 50–51.

Hodge, G., and Greve, C., "PPPs: The Passage of Time Permits a Sober Reflection," *Economic Affairs* 29 (2009): 33–39.

Holland, D. M., "The Effect of Taxation on Effort: Some Results for Business Executives," National Tax Association Proceedings of the Sixty-Second Annual Conference, 1969.

Hotelling, H., "Stability in Competition," *Economic Journal* 39 (March 1929): 41–57.

Hubbard, G. (ed.), *Asymmetric Information, Corporate Finance, and Investment: A National Bureau of Economic Research Project Report* (Chicago and London: University of Chicago Press, 1990).

Imberman, S., "Achievement and Behavior in Charter Schools: Drawing a More Complete Picture," *The Review of Economics and Statistics* 93 (May 2011): 416–35.

Institute of Medicine, *To Err Is Human: Building a Safer Health System* (November 1999).

Jencks, C., *Rethinking Social Policy* (Cambridge, MA: Harvard University Press, 1992).

Jenkins, G., Jenkins, H., and Kuo, C. Y., "Is the Value Added Tax Naturally Progressive?" Queen's Economics Department Working Paper No. 1059, April 2006.

Jensen, M., "Takeovers: Their Causes and Consequences," *Journal of Economic Perspectives* 2, no. 1 (Winter 1988): 21–48.

Kessler, D., and McClellan, M., "Do Doctors Practice Defensive Medicine?" *Quarterly Journal of Economics* 111, no. 2 (1996).

King, J., Mullainathan, S., and Congdon, W., *Policy and Choice: Public Finance through the Lens of Behavioral Economics* (Washington, DC: Brookings Institution Press, 2011).

King, M., *Public Policy and the Corporation* (London: Chapman and Hall, 1977).

King, M., and Fullerton, D., *The Taxation of Income from Capital* (Chicago: University of Chicago Press, 1984).

Kling, C., et al., "From Exxon to BP: Has Some Number Become Better than No Number?" *Journal of Economic Perspectives* 26 (Fall 2012): 3–56.

Koplow, D., "Subsidies to Uranium Enrichment," in *Nuclear Power: Still Not Viable without Subsidies* (Cambridge, MA: Union of Concerned Scientists, February 2011), pp. 62–72.

Kotlikoff, L. J., and Summers, L. H., "Tax Incidence," Chapter 16 in A. J. Auerbach and M. Feldstein (eds.), *The Handbook of Public Economics,* vol. 2 (New York: Elsevier, 1987), pp. 1043–92.

Kramer, G., "On a Class of Equilibrium Conditions for Majority Rule," *Econometrica* 41 (1973): 285–97.

Kritzer, B. E., "Note: Chile's Next Generation Pension Reform," *Social Security Bulletin* 68 (2008): 1–22.

Krueger, A. B., "Experimental Estimates of Educational Production Functions," *Quarterly Journal of Economics* 63 (May 1999).

Lau, L. J., Sheshinski, E., and Stiglitz, J. E., "Efficiency in the Optimum Supply of Public Goods," *Econometrica* 46 (1978): 269–84.

Levy, F., and Michel, R., "Work for Welfare: How Much Good Will It Do?" *American Economic Review* 76 (May 1986): 399–404.

Liebman, J., "The Optimal Design of the Earned Income Tax Credit," in B. D. Meyer and D. Holtz-Eakin (eds.), *Making Work Pay: The Earned Income Tax Credit and Its Impact on American Families* (New York: Russell Sage Foundation, 2002), pp. 196–234.

Lindahl, E., "*Positive Lösung, Die Gerechtigkeit der Besteuerung,*" translated as "Just Taxation—A Positive Solution," in R. A. Musgrave and A. T. Peacock (eds.), *Classics in the Theory of Public Finance* (New York: St. Martin's Press, 1958).

Lindsey, L., "Individual Taxpayer Response to Tax Cuts: 1982–1984, With Implications for Revenue

Maximizing Tax Rate," *Journal of Public Economics* 33 (1987): 173–206.

Lindsey, L., *The Growth Experiment: How the New Tax Policy Is Transforming the U.S. Economy* (New York: Basic Books, 1990).

Lipsey, R. G., and Lancaster, K., "The General Theory of Second Best," *Review of Economic Studies* 24 (1956–1957): 11–32.

Littenberg, B., et al., "Screening for Hypertension," *Annals of Internal Medicine* 112 (1990): 192–202.

Maag, E., "Poverty and Income Tax Entry Threshold," Urban Brookings Tax Policy Center Notes, August 29, 2011.

MaCurdy, T., "Work Disincentive Effects of Taxes: A Reexamination of Some Evidence," *American Economic Association Papers and Proceedings* (May 1992): 243–49.

Management & Training Corporation, *Privatization in Corrections: Increased Performance and Accountability Is Leading to Expansion,*" (MTC Institute, December 2009).

McClure, W., "Buying Right: The Consequences of Glut," *Business and Health* (September 1985): 43–46.

McIntyre, R. S., Garner, M., and Phillips, R., *The Sorry State of Corporate Taxes: What Fortune 500 Firms Pay (or Don't Pay) in the USA and What They Pay Abroad—2008 to 2012* (Washington, DC: Citizens for Tax Justice and Institute on Taxation and Economic Policy, February 2014).

Meade, J., *Trade and Welfare: Mathematical Supplement* (Oxford: Oxford University Press, 1955).

Megginson, W. L., et al., "The Financial and Operating Performance of Newly Privatized Firms: An International Empirical Analysis," *Journal of Finance* 49, no. 2 (1994).

Megginson, W., and Netter, J., "From State to Market: A Survey of Empirical Studies on Privatization," *Journal of Economic Literature* 39 (2001): 321–89.

Mintz, J. A., " 'Treading Water': A Preliminary Assessment of EPA Enforcement during the Bush II Administrations," *Environmental Law Institute News & Analysis* 34 (October 2004): 10933–53.

Modigliani, F., and Miller, M. H., "The Cost of Capital, Corporation Finance, and the Theory of Investment," *American Economic Review* 48 (1958): 261–97.

Moffit, R., "Incentive Effects of the U.S. Welfare System: A Review," *Journal of Economic Literature* (March 1992): 1–61.

Moffit, R., and Rangarajan, A., "The Work Incentives of AFDC Tax Rates: Reconciling Different Estimates," *Journal of Human Resources* (Winter 1991): 165–79.

Moore, A., Nolan, J., and Segal, G. F., "Putting Out the Trash: Measuring Municipal Service Efficiency in U.S. Cities," *Urban Affairs Review* 41, no. 2 (2005): 237–59.

Morley, E., Bryant, S. P., and Hatry, H. P., *Comparative Performance Measurement* (Washington, DC: The Urban Institute Press, 2001).

Morrall III, J. R., "Exposure to Occupational Noise," in James C. Miller III and Bruce Yandle (eds.), *Benefit–Cost Analyses of Social Regulation* (Washington, DC: American Enterprise Institute for Public Policy Research, 1979).

Mroz, T. A., "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions" *Econometrica* (July 1987): 765–99.

Myers, S. C., and Majluf, N. S., "Corporate Financing and Investment Decisions When Firms Have Information that Investors Do Not Have," *Journal of Financial Economics* 13, no. 2 (June 1984): 187–221.

National Public Radio, *A Century of U.S. Campaign Finance Law,* www.npr.org.

National Research Council, *Measuring Poverty: A New Approach* (Washington, DC: National Academy of Sciences, 1996).

National Science Foundation, *Science and Engineering Indicators, 2012,* Table 4–20.

Neal, D., "The Design of Performance Pay in Education," Chapter 6 in E. Hanushek, S. Machin, and L. Woessmann (eds.), *Handbook of the Economics of Education*, vol. 4 (Waltham, MA: North-Holland, 2011).

Newhouse, J. P., *Free for All? Lessons from the RAND Health Insurance Experiment* (Cambridge and London: Harvard University Press, 1993), p. 120.

Nordhaus, W., "Some Foundational and Transformative Grand Challenges for the Social and Behavioral Sciences: The Problem of Global Public Goods," National Science Foundation SBE 2020, September 2010.

Nozick, R., *Anarchy, State, and Utopia* (Oxford: Basil Blackwell, 1974).

Oates, W. A., "Green Taxes: Can We Protect the Environment and Improve the Tax System at the Same Time?" *Southern Economic Journal* 61, no. 4 (April 1995): 915–22.

Oates, W. E., and Schwab, R. M., "The Impact of Urban Land Taxation: The Pittsburgh Experience," *National Tax Journal* 50, no. 1 (March 1997): 1–21.

Pack, H., and Pack, J., "Metropolitan Fragmentation and Local Public Expenditure," *National Tax Journal* 31 (1978): 349–62.

Pauly, M., McGuire, T., and Barros, P. (eds.), *Handbook of Health Economics*, vol. 2 (Waltham, MA: North-Holland, 2012).

Pechman, J., *Federal Tax Policy,* 5th ed. (Washington DC: Brookings Institution, 1987).

Peltzman, S., "Pricing in Public and Private Enterprises: Electric Utilities in the United States," *Journal of Law and Economics* 14, no. 1 (April 1971): 109–47.

Peterson, Peter G., *Will America Grow Up before It Grows Old?* (New York: Random House, 1996).

Phelps, C. E., and Parente, S. T., "Priority Setting for Medical Technology and Medical Practice Assessment," *Medical Care* 28, no. 8 (August 1990): 702–23.

Pigou, A. C., *The Economics of Welfare* (London: Macmillan, 1918).

Pigou, A. C., *A Study in Public Finance,* 3rd ed. (London: Macmillan, 1947).

Piketty, T., Saez, E., and Stantcheva, S., "Taxing the 1%: Why the Top Tax Rate Could Be Over 80%," *Vox* 8 (December 2011), http://www.voxeu.org/article/taxing-1-why-top-tax-rate-could-be-over-80.

Portney, P. R., "The Contingent Valuation Debate: Why Economists Should Care," *Journal of Economic Perspectives* 8, no. 4 (Fall 1994): 3–17.

Poterba, J., and Summers, L. H., "Dividend Taxes, Corporate Investment and 'Q'," *Journal of Public Economics* (1983): 135–67.

Pradhan, S., "Evaluating Public Spending: A Framework for Public Expenditure Reviews," World Bank Discussion Paper 323, Washington, DC, 1996.

Ramsey, F., "A Contribution of the Theory of Taxation," *Economic Journal* 37 (1927): 47–61.

Rizzuto, R., and Wachtel, P., "Further Evidence on the Returns to School Quality," *Journal of Human Resources* (Spring 1980): 240–54.

Robins, P. K., "A Comparison of the Labor Supply Findings from the Four Negative Income Tax Experiments," *Journal of Human Resources* 20, no. 4 (Fall 1985): 567–82.

Robinson, L., "How US Government Agencies Value Mortality Risk Reductions," *Review of Environmental Economics and Policy* 1 (Summer 2007): 283–99.

Röhn, O., "New Evidence on the Private Saving Off Set and Ricardian Equivalence," OECD Economics Department Working Papers 762, Paris, 2010.

Romer, T., "Individual Welfare, Majority Voting, and the Properties of a Linear Income Tax," *Journal of Public Economies* 4 (1975): 163–85.

Rosengard, J. K., "The Tax Everyone Loves to Hate: Principles of Property Tax Reform," in G. Cornia, W. McCluskey, and L. Walters (eds.), *A Primer on the Property Tax* (Malden, MA: Wiley Blackwell, 2012).

Rothschild, M., and Stiglitz, J. E., "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics* 90 (1976): 629–50.

Rouse, C. E., "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program," *Quarterly Journal of Economics* 113, no. 2 (May 1998): 553–602.

Rudman, W. J., et al., "Healthcare Fraud and Abuse," *Perspectives in Health Information Management* 6 (Fall 2009).

Saez, E., *Striking It Richer: The Evolution of Top Incomes in the United States (Updated with 2011 Estimates)*, UC Berkeley, January 23, 2013.

Salop, S., "Information and Monopolistic Competition," *American Economic Review* (May 1976): 240–45.

Sandmo, A., "The Effects of Taxation on Savings and Risk-Taking," in A. J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics,* vol. 1 (New York: Elsevier, 1987), pp. 293–309.

Schelling, T., "The Life You Save May Be Your Own," reprinted in T. Schelling, *Choices and Consequences* (Cambridge, MA: Harvard University Press, 1984).

Scholz, J., "Tax Policy and the Working Poor: The Earned Income Tax Credit," *Focus* (Winter 1993–1994): 1–12.

Shah, A., "Fiscal Decentralization in Developing and Transition Economies," World Bank Policy Research Working Paper 3282, Washington, DC, April 2004.

Shoepfle, G. K., Perez-Lopez, J. F., and Griego, E., "The Underground Economy in the United States," U.S. Department of Labor Occasional Paper on the Informal Sector, no. 2, September 1992.

Short, K., "The Research Supplemental Poverty Measure: 2011," *Current Population Reports*, P60–244 (U. S. Census Bureau, November 2012).

Shoven, J., "New Developments in Corporate Finance and Tax Avoidance: Some Evidence," in L. Summers (ed.), *Tax Policy and the Economy* (National Bureau of Economic Research, 1987).

Shoven, J., "The Tax Consequences of Share Repurchases and Other Non-Dividend Cash Payments to Equity Owners," in L. Summers (ed.), *Tax Policy and the Economy* (Cambridge and London: MIT Press, 1987): pp. 29–54.

Showalter, M. H., and Thurston, N. K., "Taxes and Labor Supply of High-Income Physicians," *Journal of Public Economics* 66 (1997): 73–97.

Silliman, B. R., "College Tuition Tax Credits: An Examination of the Impact of the American Opportunity, HOPE, and Lifetime Learning Tax Credits since 1998," *Proceedings of ASBBS* 18, no. 1 (February 2011): 279–88.

Silverman, E., "CompStat's Innovation," in D. Weisburd and A. Braga (eds.), *Police Innovation: Contrasting Perspectives* (Cambridge, UK: Cambridge University Press, 2006).

Sinclair, U., *The Jungle*, (New York: Doubleday, Page & Company, 1906).

Skinner, R. R., "The No Child Left Behind Act: An Overview of Reauthorization Issues for the 111th Congress," Congressional Research Service Report RL33749, May 2009.

Skoufias, E., Unar, M., and Gonzalez-Cossio, T., "The Impacts of Cash and In-Kind Transfers on Consumption and Labor Supply: Experimental Evidence from Rural Mexico," Policy Research Working Paper No. 4778, Washington, DC, World Bank, 2008.

Slavin, R., "School and Classroom Organization in Beginning Reading," in R. Slavin et al. (eds.), *Preventing Early School Failure: Research, Policy, and Practice* (Boston: Allyn and Bacon, 1994), pp. 122–30.

Slemrod, J., "Old George Orwell Got It Backward: Some Thoughts on Behavioral Tax Economics," CESifo Working Paper No. 2777, September 2009.

Slutsky, S., "A Voting Model for the Allocation of Public Goods: Existence of an Equilibrium," *Journal of Economic Theory* 11 (1975): 292–304.

Smith, A., *The Wealth of Nations* (New York: Modern Library, 1937).

Sox Jr., H. C., et al., "The Role of Exercise Testing in Screening for Coronary Artery Disease," *Annals of Internal Medicine* 110 (1989): 456–69.

Sparrow, M. K., "Testimony at Criminal Prosecution as Deterrent to Health Care Fraud," before Senate Committee on Judiciary, Subcommittee on Crime and Drugs, May 20, 2009.

Spence, A. M., "Job Market Signaling," *Quarterly Journal of Economics* 87 (1973): 355–74.

Spiegelman, R. G., and Yaeger, K. E., "The Seattle and Denver Income Maintenance Experiments: Overview," *Journal of Human Resources* 15, no. 4 (Fall 1980): 463–79.

Springer, M., et al., *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT)* (Evanston, IL: Society for Research on Educational Effectiveness, 2011).

Squires, D., "The U.S. Health System in Perspective: A Comparison of Twelve Industrialized Nations," The Commonwealth Fund Pub. 1532, *Issues in International Health Policy* 16 (July 2011).

Starrett, D., "Principles of Optimal Location in a Large Homogeneous Area," *Journal of Economic Theory* 9 (1974): 418–48.

Stern, N. H., "On the Specification of Models of Optimum Income Taxation," *Journal of Public Economics* 6 (1976): 123–62.

Stern, N. H., "Optimum Taxation with Errors in Administration," *Journal of Public Economics* 17 (1982): 181–211.

Steuerle, C. E., and Bakija, J. M., *Retooling Social Security for the 21st Century: Right and Wrong Approaches to Reform* (Washington, DC: The Urban Institute Press, 1994), pp. 41, 43, 109, 287.

Steuerle, E., "Is Income from Capital Subject to Individual Income Taxation?" *Public Finance Quarterly* (July 1982): 283–303.

Stigler, G. J., "The Theory of Economic Regulation," *Bell Journal Economics and Management Science* 2 (Spring 1971): 3–21.

Stigler, G. J., "Free Riders and Collective Action: An Appendix to Theories of Economic Regulation," *Bell Journal Economics and Management Science* 5 (Autumn 1974): 359–65.

Stiglitz, J. E., "The Effects of Income, Wealth and Capital Gains Taxation on Risk Taking," *Quarterly Journal of Economics* 83 (1969): 262–83.

Stiglitz, J. E., "Taxation, Corporate Financial Policy, and the Cost of Capital," *Journal of Public Economics* (February 1973): 1–34.

Stiglitz, J. E., "On the Irrelevance of Corporation Financial Policy," *American Economic Review* 64 (1974): 851–66.

Stiglitz, J. E., "The Theory of Screening Education and the Distribution of Income," *American Economic Review* 65 (1975): 283–300.

Stiglitz, J. E., "The Corporation Income Tax," *Journal of Public Economics* (April–May 1976): 303–11.

Stiglitz, J. E., "Theory of Local Public Goods," in M. Feldstein and R. Inman (eds.), *The Economics of Public Services* (New York: Macmillan, 1977), pp. 274–333.

Stiglitz, J. E., "Self-Selection and Pareto-Efficient Taxation," *Journal of Public Economics* 17 (1982): 213–40.

Stiglitz, J. E., "Utilitarianism and Horizontal Equity: The Case for Random Taxation," *Journal of Public Economics* 21 (1982): 257–94.

Stiglitz, J. E., "Public Goods in Open Economies with Heterogeneous Individuals," in J. F. Thisse and H. G. Zoller (eds.), *Locational Analysis of Public Facilities* (New York: North-Holland, 1983).

Stiglitz, J. E., "Pareto Efficient and Optimal Taxation and the New Welfare Economics," in Alan J. Auerbach and Martin Feldstein (eds.), *Handbook of Public Economics* (New York: Elsevier, 1987), pp. 991–1042.

Stiglitz, J. E., "On the Relevance or Irrelevance of Public Financial Policy," in K. J. Arrow and M. J. Boskin (eds.), *The Economics of Public Debt* (New York: Macmillan Press, 1988), pp. 41–76.

Stiglitz, J. E., and Weiss, A., "Credit Rationing in Markets with Imperfect Information," *American Economic Review* 71, no. 3 (June 1981): 393–410.

Stiglitz, J. E., "The Theory of International Public Goods and the Architecture of International Organizations," Background Paper No. 7, Third Meeting, High Level Group on Development Strategy and Management of the Market Economy, UNU/WIDER, Helsinki, Finland, July 8–10, 1995.

Stiglitz, J. E., "Unfair Trade Laws and Other Mischief," in *Globalization and Its Discontents* (New York: W. W. Norton & Company, 2002), pp. 166–79.

Stiglitz, J. E., and Bilmes, J., *The Trillion Dollar War: The True Cost of the Iraq Conflict* (New York: W. W. Norton & Company, 2008).

Stiglitz, J. E., *The Price of Inequality: How Today's Divided Society Endangers Our Future*, (New York: W. W. Norton & Company, 2012).

Symposium on Contingent Valuation, *Journal of Economic Perspectives* 8, no. 4 (Fall 1994).

Synder, T. D., and Dillow, S. A., *Digest of Education Statistics 2010* (Washington, DC: U.S. Department of Education, Institute of Education Sciences, April 2011), Table 391.

Szymendera, S., "Social Security Administration (SSA): Budget Issues," Congressional Research Service Report R41716, August 2011.

Thomas, C., Bolinger, D., and Badalamenti, J., "Private Adult Correctional Facility Census," Private Corrections Project, Center for Studies of Criminology and Law, University of Florida, March 1997, p. iv.

Tiebout, C., "A Pure Theory of Local Expenditure," *Journal of Political Economy* 64 (1956): 416–24.

Tobin, J., "On Limiting the Domain of Inequality," *Journal of Law and Economics* 13 (1970): 263–77.

Triest, R. K, "The Effect of Income Taxation on Labor Supply in the United States," *Journal of Human Resources* (Summer 1990): 491–516.

van den Houweland, A., and Wooders, M., "Share Equilibrium in Local Public God Economics," *Journal of Mathematical Economics* 47 (May 2011): 376–81.

Varian, H., *Microeconomic Analysis*, 3rd ed. (New York: W. W. Norton & Company, 1982), pp. 106–7.

Veras Soares, F., Perez Ribas, R., and Guerreiro Osório, R., "Evaluating the Impact of Brazil's Bolsa Família: Cash Transfer Programs in Comparative Perspective," *Latin American Research Review* 45 (2010): 173–90.

Viscusi, W. K., *Risk by Choice: Regulating Health and Safety in the Workplace* (Cambridge, MA: Harvard University Press, 1983).

Viscusi, W. K., "The Value of Life: Estimates with Risks by Occupation and Industry," *Economic Inquiry* 42, no. 1 (January 2004): 29–48.

Wainger, L., and Mazzotta, M., "Realizing the Potential of Ecosystem Services: A Framework for Relating Ecological Changes to Economic Benefits," *Environmental Management* 48 (October 2011): 710–33.

Wallin, B. A., "Symposium Tax and Expenditure Limitations: A Quarter Century after Proposition 13," *Public Budgeting & Finance* 24, no. 4 (Winter 2004).

Weinstein, M. C., "Economic Assessment of Medical Practices and Technologies," *Medical Decision Making* 1 (1981): 309–30.

Weiss, A., "Human Capital versus Signaling Explanations of Wages," *Journal of Economics Perspectives* 9, no. 4 (Fall 1995): 133–54.

Wenglinsky, H., *When Money Matters: How Educational Expenditures Improve Student Performance and How They Don't* (Princeton, NJ: Educational Testing Service, Policy Information Center, April 1997).

Wennberg, J., and Gittelsohn, A., "Small Are Variations in Health Care Delivery," *Science* 182 (1973): 1102–8.

Whitehurst, G. J., and Chingos, M. M., *Class Size: What Research Says and What It Means for State Policy* (Washington, DC: Brookings Institution, May 2011).

Wilensky, G., "Government and the Financing of Health Care," *American Economic Review* 72, no. 2 (May 1982): 205.

Willig, R., "Consumer's Surplus without Apology," *American Economic Review* 66 (1976): 589–97.

Windham, D., "Economic Analysis and the Public Support of Higher Education: The Divergence of Theory and Policy," in *Economic Dimensions of Education: A Report of a Committee of the National Academy of Education* (May 1979).

Wolfe, B. L., "Incentives, Challenges, and Dilemmas of TANF," Institute for Research on Poverty Discussion Paper No. 1209-00, University of Wisconsin—Madison, May 2000.

Wolff, E. N., Baumol, W., and Saini, A. N., "A Comparative Analysis of Education Costs and Outcomes: The United States vs. Other OECD Countries," June 26, 2013, http://ssm.com/abstract=2285597.

Word, E., et al., *Student/Teacher Achievement Ratio (STAR)—Tennessee's K-3 Class Size Study* (Tennessee Department of Education, 1994).

Zimmer, R., et al., *Charter Schools in Eight States: Effects on Achievement, Attainment, Integration, and Competition* (Santa Monica, CA: Rand Corporation, 2009).

Zodrow, G. R., and Mieszkowski, P., "Pigou, Tiebout, Property Taxation, and the Underprovision of Local Public Goods," *Journal of Urban Economics* 19 (1986): 356–70.

INDEX