

research because that is the least appropriable kind of information. The inappropriability and high social return on research lead most governments to subsidize basic research in the fields of health and science and to provide special incentives for other creative activities.

Intellectual Property Rights

Governments have long recognized that creative activities need special support because the rewards for producing valuable information are reduced by imitation. The U.S. Constitution authorizes Congress “to promote the Progress of Science and useful Arts, by securing, for limited Times, to Authors and Inventors, the exclusive Right to their respective Writings and Discoveries.” Thus special laws governing patents, copyrights, business and trade secrets, and electronic media create **intellectual property rights**. The purpose is to give the owner special protection against the material’s being copied and used by others without compensation to the owner or original creator.

The earliest intellectual property right was the **patent**, under which the U.S. government creates an exclusive use (in effect, a limited monopoly) over a “novel, nonobvious, and useful” invention for a limited period, currently set at 20 years. Similarly, copyright laws provide legal protection against unauthorized copying of original works in different media such as text, music, video, art, software, and other information goods.

Why would governments actually encourage monopolies? In effect, patents and copyrights grant property rights to inventors over books, music, and ideas. By allowing inventors to have exclusive use of their intellectual property, the government increases the degree of appropriability and thereby increases the incentives for people to invent useful new products, write books, compose songs, and write computer software. A patent also requires disclosure of the technological details of the invention, which encourages further invention and lawful imitation. Examples of successful patents include those on the cotton gin, the telephone, the Xerox machine, and many profitable drugs.

The Dilemma of the Internet

Inventions that improve communications are hardly limited to the modern age. But the rapid growth

of electronic storage, access, and transmission of information highlights the dilemma of providing incentives for creating new information. Many new information technologies have large up-front or sunk costs but virtually zero marginal costs. With the low cost of electronic information systems like the Internet, it is technologically possible to make large amounts of information available to everyone, everywhere, at close to zero marginal cost. Perfect competition cannot survive here because a price equal to a zero marginal cost will yield zero revenues and therefore no viable firms.

The economics of the information economy highlights the conflict between efficiency and incentives. On the one hand, all information might be provided free of charge—free economics textbooks, free movies, free songs. Free provision of information looks economically efficient because the price would thereby be equal to the marginal cost, which is zero. But a zero price on intellectual property would destroy the profits and therefore reduce the incentives to produce new books, movies, and songs because creators would reap little return from their creative activity. Society has struggled with this dilemma in the past. But with the costs of reproduction and transmission so much lower for electronic information than for traditional information, finding sensible public policies and enforcing intellectual property rights is becoming ever more difficult.

Experts emphasize that intellectual property laws are often hard to enforce, especially when they apply across national borders. The United States has a long-running trade dispute charging that China condones the illegal copying of American movies, musical recordings, and software. A DVD movie that sells for \$25 in the United States can be purchased for 50 cents in China. The U.S. copyright industries estimate that 85 to 95 percent of all their members’ copyrighted works sold in 2007 in China were pirated.

In a world increasingly devoted to developing new knowledge—much of it intangible, like music, movies, patents, new drugs, and software—governments must find a middle ground in intellectual property rights. If intellectual property rights are too strong, this will lead to high prices and monopoly losses, while too weak intellectual property laws will discourage invention and innovation.



SUMMARY

A. Economics of Risk and Uncertainty

1. Economic life is full of uncertainty. Consumers face uncertain incomes and employment patterns as well as the threat of catastrophic losses; businesses have uncertain costs, and their revenues contain uncertainties about price and production.
2. In well-functioning markets, arbitrage, speculation, and insurance help smooth out the unavoidable risks. Speculators are people who buy and sell assets or commodities with an eye to making profits on price differentials across markets. They move goods across regions from low-price to high-price markets, across time from periods of abundance to periods of scarcity, and even across uncertain states of nature to periods when chance makes goods scarce.
3. The profit-seeking action of speculators and arbitragers tends to create certain equilibrium patterns of price over space, time, and risks. These market equilibria are zero-profit outcomes where the marginal costs and marginal utilities in different regions, times, or uncertain states of nature are in balance. To the extent that speculators moderate price and consumption instability, they are part of the invisible-hand mechanism that performs the socially useful function of reallocating goods from feast times (when prices are low) to famine times (when prices are high).
4. Speculative markets allow individuals to hedge against unwelcome risks. The economic principle of risk aversion, which derives from diminishing marginal utility, implies that individuals will not accept risky situations with zero expected value. Risk aversion implies that people will buy insurance to reduce the potentially disastrous declines in utility from fire, death, or other calamities.

B. The Economics of Insurance

5. Insurance and risk spreading tend to stabilize consumption in different states of nature. Insurance takes large individual risks and spreads them so broadly that they become acceptable to a large number of individuals. Insurance is beneficial because, by helping to equalize consumption across different uncertain states, it raises the expected level of utility.
6. The conditions necessary for the operation of efficient insurance markets are stringent: there must be large numbers of independent events and little chance of moral hazard or adverse selection. When market failures such as adverse selection arise, prices may be distorted or markets may simply not exist.

7. If private insurance markets fail, the government may step in to provide social insurance. Social insurance is provided by governments when private insurance markets cannot function effectively and society believes that individuals should have a social safety net for major risks such as unemployment, illness, and low incomes. Even in the most laissez-faire of advanced market economies today, governments insure against unemployment and health risks in old age.

C. Health Care: The Problem That Won't Go Away

8. Health care is the largest social insurance program. The health-care market is characterized by multiple market failures that lead governments to intervene. Health-care systems have major externalities. Additionally, the asymmetric information between doctors and patients leads to uncertainties about the appropriate treatment and level of care, and the asymmetry between patients and insurance companies leads to adverse selection in the purchase of insurance. Finally, because health care is so important to human welfare and to labor productivity, most governments strive to provide a minimum standard of health care to the population.
9. When the government subsidizes health care and attempts to provide universal coverage, there will be excess demand for medical services. One of the challenges is to develop efficient and equitable mechanisms of nonprice rationing.

D. Innovation and Information

10. Schumpeter emphasized the importance of the innovator, who introduces "new combinations" in the form of new products and new methods of organization and is rewarded by temporary entrepreneurial profits.
11. Today, the economics of information emphasizes the difficulties involved in the efficient production and distribution of new and improved knowledge. Information is different from ordinary goods because it is expensive to produce but cheap to reproduce. The inability of firms to capture the full monetary value of their inventions is called inappropriability. To increase appropriability, governments create intellectual property rights governing patents, copyrights, trade secrets, and electronic media. The growth of electronic information systems like the Internet has increased the dilemma of how to efficiently price information services.

CONCEPTS FOR REVIEW

Risk, Uncertainty, and Insurance

arbitrage leading to regional equalization of prices
ideal seasonal price pattern
speculation, arbitrage, hedging
risk aversion and diminishing marginal utility

consumption stability vs. instability
insurance and risk spreading
moral hazard, adverse selection
social insurance
nonprice rationing

Economics of Information

information economics
inappropriability, protection of intellectual property rights, dilemma of efficient production of knowledge
market failure in information

FURTHER READING AND INTERNET WEBSITES

Further Reading

The concept of social insurance was described by Martin Feldstein in "Rethinking Social Insurance," *American Economic Review*, March 2005 and available at www.nber.org/feldstein/aeajan8.pdf.

For an analysis of gambling, see William R. Eadington, "The Economics of Casino Gambling," *Journal of Economic Perspectives*, Summer 1999.

The Schumpeterian hypothesis was developed in Joseph Schumpeter, *Capitalism, Socialism, and Democracy* (Harper & Row, New York, 1942).

Many of the economic, business, and policy issues involved in the new information economy are covered in a nontechnical book by two eminent economists, Carl Shapiro and Hal R. Varian, *Information Rules* (Harvard Business School Press, Cambridge, Mass., 1998). A discussion of the economics of the Internet is contained in Jeffrey K. MacKie-Mason

and Hal Varian, "Economic FAQs about the Internet," *Journal of Economic Perspectives*, Summer 1994, p. 92.

A discussion by the U.S. government of Chinese infringement of intellectual property rights can be found at www.usitc.gov/Document_Library/Reports_Publications/Section_Index.html.

Websites

One of the most interesting websites about the Internet and intellectual property rights is compiled by Hal R. Varian, chief economist of Google and former dean of the School of Information Management and Systems at the University of California at Berkeley. This site, called "The Economics of the Internet, Information Goods, Intellectual Property and Related Issues," is at www.sims.berkeley.edu/resources/infoecon.

Information on the American health-care system is usefully compiled by the National Center on Health Statistics at www.cdc.gov/nchs/.

QUESTIONS FOR DISCUSSION

1. Suppose a friend offers to flip a fair coin, with you paying your friend \$100 if it comes up heads and your friend paying you \$100 if it comes up tails. Explain why the expected dollar value is \$0. Then explain why the expected utility value is negative if you are risk-averse.
2. Consider the example of grade insurance (see page 218). Suppose that with a grade-insurance policy, students would be compensated \$5000 a year for each point that their grade point average fell below the top grade (the resulting number might be an estimate of the impact of grades on future earnings). Explain why

the presence of grade insurance would produce moral hazard and adverse selection. Why would moral hazard and adverse selection make insurance companies reluctant to sell grade insurance? Are you surprised that you cannot buy grade insurance?

3. After the terrorist attacks of September 11, 2001, most insurance companies canceled their insurance coverage for terrorism. According to President Bush, "More than \$15 billion in real estate transactions have been canceled or put on hold because owners and investors could not obtain the insurance protection they need."

As a result, the federal government stepped in to provide coverage for up to \$90 billion in claims. Using the principles of insurance, explain why insurance companies might decline to insure property against terrorist attacks. Explain whether or not you think the federal program is an appropriate form of social insurance.

4. In the early nineteenth century, little of the nation's agricultural output was sold in markets, and transportation costs were very high. What would you expect to have been the degree of price variation across regions as compared with that of today?
5. Assume that a firm is making a risky investment (say, spending \$2 billion developing a competitor to Windows). Can you see how the diversified ownership of this firm could allow near-perfect risk spreading on the software investment?
6. Health insurance companies sometimes do not allow new participants to be covered on "existing conditions," or preexisting illnesses. Explain why this policy might alleviate problems of adverse selection.
7. Joseph Schumpeter wrote as follows:

The modern standard of life of the masses evolved during the period of relatively unfettered "big business." If we list the items that enter the modern workman's budget and, from 1899 on, observe the course of their prices, we cannot fail to be struck by the rate of the advance which,

considering the spectacular improvement in qualities, seems to have been greater and not smaller than it ever was before. Nor is this all. As soon as we inquire into the individual items in which progress was most conspicuous, the trail leads not to the doors of those firms that work under conditions of comparatively free competition but precisely to the doors of the large concerns—which, as in the case of agricultural machinery, also account for much of the progress in the competitive sector—and a shocking suspicion dawns upon us that big business may have had more to do with creating that standard of life than keeping it down. (*Capitalism, Socialism, and Democracy*)

Use this passage to describe the tradeoff between "static" monopoly inefficiencies and "dynamic" efficiencies of technological change.

8. Long-term care for the elderly involves helping individuals with activities (such as bathing, dressing, and toileting) that they cannot perform for themselves. How were these needs taken care of a century ago? Explain why moral hazard and adverse selection make long-term-care insurance so expensive today that few people choose to buy it.
9. Economic studies have found that the private rate of return on inventions is typically as low as one-third of the social return. Explain this finding in terms of the economics of innovation.

PART THREE

Factor Markets: Labor, Land, and Capital

How Markets Determine Incomes

12



You know, Ernest, the rich are different from us.

F. Scott Fitzgerald

Yes, I know. They have more money than we do.

Ernest Hemingway

A. INCOME AND WEALTH

Earlier chapters have surveyed the output and prices of goods and services produced by tiny farms and giant corporations. But the vast array of products that we enjoy do not simply gush from the earth—they are produced by workers who are equipped with machines, which are housed in factories, which are sitting on land. These inputs into the productive process earn factor incomes—wages, profits, interest, and rents. The time has come to understand the determination of factor prices along with the forces that affect the distribution of income among the population.

America is a land of extremes of income and wealth. If you are one of the 400 richest Americans, you are likely to be a 60-year-old white male with a degree from a top university and a net worth of about \$4 billion. This tiny sliver of American society owns about 3 percent of the total wealth of the country. In the past, you made your fortune in manufacturing or real estate, but recent billionaires come largely from information technology and finance. Your voyage to the top was as much the product of birth as of brains, for your family probably gave you a head

start with an expensive education, but there are more self-made men and women today than there were a decade ago.

At the other extreme are forgotten people who never make the cover of *Forbes* or *People* magazine. Listen to the story of Robert Clark, homeless and unemployed. A roofer and Vietnam veteran, he came to Miami from Detroit looking for work. He slept on the city streets on a piece of cardboard covered by a stolen sheet. Every day he and other homeless men crept out of the culverts into the daylight to work for temporary-employment firms. These firms charged clients \$8 to \$10 an hour, paid the men the minimum wage, and then took most of the money back for transportation and tools. Clark's pay stub showed earnings of \$31.28 for 31 hours of work.

How can we understand these extremes of income and wealth? Why are some people paid \$10 million a year, while others net only \$1 an hour? Why is real estate in Tokyo or Manhattan worth thousands of dollars a square foot, while land in the desert may sell for but a few dollars an acre? And what is the source of the billions of dollars of profits earned by giant enterprises like Microsoft and General Electric?

Questions about the distribution of income are among the most controversial in all economics.

Some people argue that high incomes are the unfair result of past inheritance and luck while poverty stems from discrimination and lack of opportunity. Others believe that people get what they deserve and that interfering with the market distribution of income would injure an economy's efficiency and make everyone worse off. Government programs in America today reflect an uneasy consensus that incomes should be largely determined by market earnings but the government should provide a social safety net to catch the deserving poor who fall below some minimum standard of living.

INCOME

In measuring the economic status of a person or a nation, the two yardsticks most often used are income and wealth. **Income** refers to the flow of wages, interest payments, dividends, and other things of value accruing during a period of time (usually a year). The aggregate of all incomes is *national income*, the components of which are shown in Table 12-1. The biggest share of national income goes to labor, either as wages or salaries or as fringe benefits. The remainder

goes to the different types of *property income*: rent, net interest, corporate profits, and proprietors' income. This last category basically includes the returns to the owners of small businesses.¹

The earnings in a market economy are distributed to the owners of the economy's factors of production in the form of wages, profits, rent, and interest.

Factor Incomes vs. Personal Incomes

It is important to understand the distinction between factor incomes and personal incomes. Table 12-1 reports the distribution of factor incomes—the division between labor and property incomes. But the same person may own many different factors of production. For example, someone might receive a salary, earn interest on money in a savings account, get dividends from shares in a mutual fund, and collect rent on a real-estate investment. In economic language, we observe that a person's market income is

¹ Economists and accountants often measure "income" in different ways. We studied accounting measures of income and wealth in Chapter 7.

Type of income	Amount (\$, billion)	Share of total (%)	Examples
Labor income:			
Wages and salaries	6,356	51.8	Autoworker's wages; teacher's salary
Benefits and other labor income	1,457	11.9	Company contribution to pension fund
Property income:			
Proprietors' income	1,056	8.6	Barber's earnings; lawyer's share of partnership net income
Rental income	40	0.3	Landlord's rent from apartments after expenses and depreciation
Corporate profits	1,642	13.4	Microsoft's profits
Net interest	664	5.4	Interest paid on savings account
Taxes on production and other		<u>1,056</u>	<u>8.6</u>
Total	12,271	100.0	

TABLE 12-1. Division of National Income, 2007

National income includes all the incomes paid to factors of production. Almost three-quarters consists of wages and other kinds of compensation of labor, while the rest is divided among rents, corporate profits, and the incomes of proprietors.

Source: U.S. Department of Commerce, Bureau of Economic Analysis, at the Web page www.bea.gov.

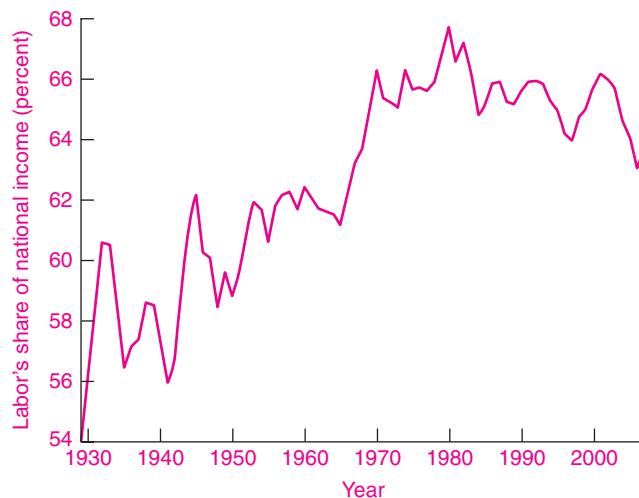


FIGURE 12-1. The Share of Labor in National Income

The share of labor income increased gradually until 1970. Since then, it has been remarkably stable at around two-thirds of national income. The remainder of income is distributed among rents, interest, corporate profits, and proprietors' income and miscellaneous items like production taxes.

simply the quantities of factors of production sold by that person times the wage or price of each factor.

About two-thirds of national income goes to labor, while the rest is distributed as some form of returns to property. The last quarter-century has been a turbulent one. What has been the impact of energy shortages, the computer revolution, globalization, corporate downsizing, and the financial turmoil of recent years on labor's share of the total income pie? Looking at Figure 12-1, we can see that the share of national income going to labor has changed very little since 1970. This is one of the remarkable features of the income distribution in the United States.

Role of Government

How does government fit into this picture? Governments at every level form the largest source of wages, rents, and interest payments. The results of government purchases are included in the payments to factors of production shown in Table 12-1.

Yet government also has a direct role in incomes that does not show up in Table 12-1. To begin with, the government collects a sizable share of national income through taxation and other levies. In 2008 about 30 percent of gross domestic product was collected by federal, state, and local governments as various types of taxes, including personal income taxes, corporate-profit taxes, and social security taxes.

But what governments tax, they also spend or give away. Governments at all levels provide incomes in the form of **transfer payments**, which are payments by governments to individuals that are not made in return for current goods or services. The biggest single category of transfer payments is social security for older Americans, but transfer payments also include unemployment insurance, farm subsidies, and welfare payments. Whereas Americans derived almost none of their incomes from governments in 1929, fully 15 percent of personal incomes in 2008 came from government transfer payments.

Personal income equals market income plus transfer payments. Most market income comes from wages and salaries; a small, affluent minority derives its market income from earnings on property. The major component of government transfers is social security payments to the elderly.

WEALTH

We see that some income comes from interest or dividends on holdings of bonds or stocks. This brings us to the second important economic concept: **Wealth** consists of the net dollar value of assets owned at a given point in time. Note that wealth is a *stock* (like the volume of a lake) while income is a *flow* per unit of time (like the flow of a stream). A household's wealth

Distribution of Assets of All Families as Percentage of All Assets, 1989–2004			
	Percentage of Total Assets		
	1989	1995	2004
Financial:			
Bank deposits and similar	9.4	7.7	6.2
Bonds	3.1	2.3	1.9
Stocks	6.2	10.4	11.5
Retirement accounts	6.6	10.3	11.4
Other	5.3	6.0	4.7
Tangible and other assets:			
Own home	31.9	30.0	32.3
Other real estate and property	13.4	10.0	11.1
Vehicles	3.9	4.5	3.3
Business equity	18.6	17.2	16.7
Other	1.7	1.5	1.0
Thousands of 2004 Dollars			
Family net worth:			
Median	68.9	70.8	93.1
Average	272.3	260.8	448.2

TABLE 12-2. Trends in Wealth of American Households

Households own tangible assets (such as houses and cars) as well as financial assets (such as savings accounts and stocks). The largest single asset for most Americans continues to be the family home. The median wealth is much smaller than the average, reflecting the great inequality of wealth holding.

Source: Federal Reserve Board, Survey of Consumer Finances, available in *Federal Reserve Bulletin* or at www.federalreserve.gov/Pubs/oss/oss2/2004/bull0206.pdf.

includes its tangible items (houses, cars and other consumer durable goods, and land) and its financial holdings (such as cash, savings accounts, bonds, and stocks). All items that are of value are called *assets*, while those that are owed are called *liabilities*. The difference between total assets and total liabilities is called wealth or *net worth*.

Table 12-2 presents a breakdown of the asset holdings of Americans from 1989 to 2004. The single most important asset of most households is the family home: 68 percent of families own houses, as compared with 55 percent a generation ago. Most households own a modest amount of financial wealth in savings accounts, and about one-fifth directly own corporate stocks. But it turns out that a large proportion of the nation's financial wealth is concentrated in the hands of a small fraction of the population.

About one-third of all wealth is owned by the richest 1 percent of American households.

B. INPUT PRICING BY MARGINAL PRODUCTIVITY

The **theory of income distribution** (or **distribution theory**) studies how incomes are determined in a market economy. People are often puzzled by the vast differences in incomes of different families. Are they caused by differences in talents? By monopoly power? By government intervention? Why is Bill Gates worth \$60 billion while half of American black families have net worth less than \$20 thousand? Why

are land prices so much higher in the city than in the desert?

Our first answer to these questions is that the distribution theory is a special case of the theory of prices. Wages are the price of labor; rents are the price for using land; and so forth. Moreover, the prices of factors of production are primarily set by the interaction between supply and demand for different factors—just as the prices of goods are largely determined by the supply and demand for goods.

But pointing to supply and demand is just the first step on the road to understanding income distribution in a competitive market economy. We will see that the key to incomes lies in the *marginal products* of different factors of production. In this section, we will see that wages are determined by the value of the *marginal product of labor*, or what is known as the marginal revenue product of labor. The same holds for other factors of production as well. We first discuss this new concept and then show how it solves the puzzle of how incomes are determined.

THE NATURE OF FACTOR DEMANDS

The demand for factors differs from that for consumption goods in two important respects: (1) Factor demands are derived demands, and (2) factor demands are interdependent demands.

Demands for Factors Are Derived Demands

Let's consider the demand for office space by a firm which produces computer software. A software company will rent office space for its programmers, customer service representatives, and other workers. Similarly, other companies like pizza shops or banks will need space for their activities. In each region, there will be a downward-sloping demand curve for office space linking the rental being charged by landlords to the amount of office space desired by companies—the lower the price, the more space companies will want to rent.

But there is an essential difference between ordinary demands by consumers and the demand by firms for inputs. Consumers demand final goods like computer games or pizzas because of the direct enjoyment or utility these consumption goods provide. By contrast, a business does not pay for inputs like office space because they yield direct

satisfaction. Rather, it buys inputs because of the production and revenue that it can gain from employment of those factors.

Satisfactions are in the picture for inputs—but at one stage removed. The satisfaction that consumers get from playing computer games determines how many games the software company can sell, how many clerks it needs, and how much office space it must rent. The more successful its software, the greater its demand for office space. An accurate analysis of the demand for inputs must, therefore, recognize that consumer demands do *ultimately* determine business demands for office space.

This analysis is not limited to office space. Consumer demands determine the demand for all inputs, including farmland, oil, and pizza ovens. Can you see how the demand for professors of economics is ultimately determined by the demand for economics courses by students?

The firm's demand for inputs is derived indirectly from the consumer demand for its final product.

Economists therefore speak of the demand for productive factors as a **derived demand**. This means that when firms demand an input, they do so because that input permits them to produce a good which consumers desire now or in the future. Figure 12-2 on page 234 shows how the demand for a given input, such as fertile cornland, must be regarded as being derived from the consumer demand curve for corn. In the same way, the demand for office space is derived from the consumer demand for software and all the other products and services provided by the companies that rent office space.

Demands for Factors Are Interdependent

Production is a team effort. A chain saw by itself is useless for cutting down a tree. A worker with empty hands is equally worthless. Together, the worker and the saw can cut the tree very nicely. In other words, the productivity of one factor, such as labor, depends upon the amount of other factors available to work with.

Therefore, it is generally impossible to say how much output has been created by a single input taken by itself. Asking which factor is more important is like asking whether a mother or a father is more essential in producing a baby.

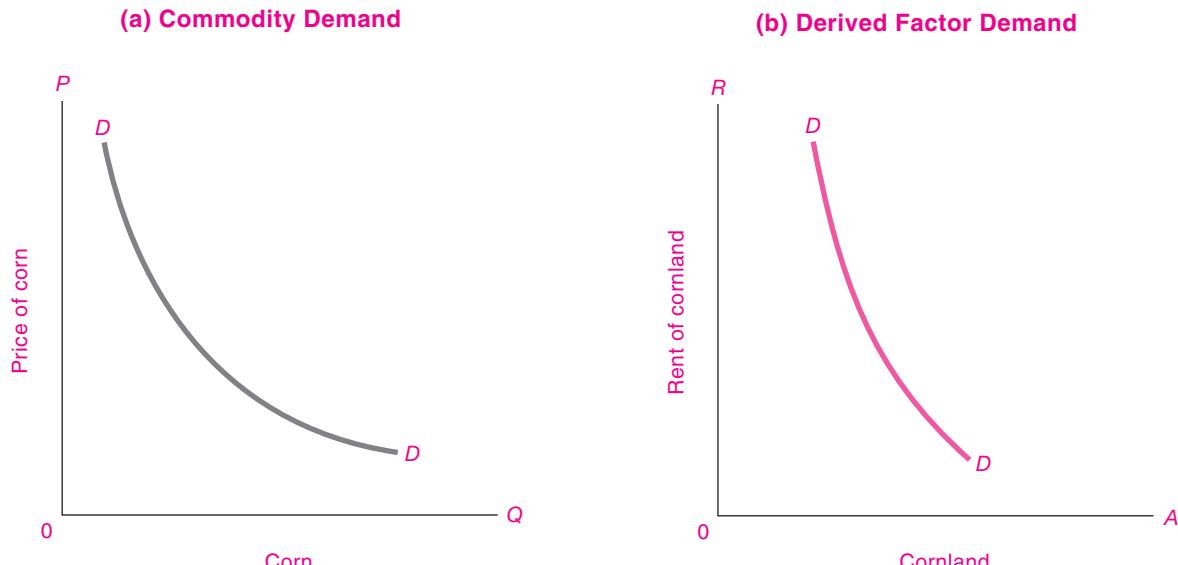


FIGURE 12-2. Demand for Factors Is Derived from Demand for Goods They Produce

The green curve of derived demand for cornland comes from the blue curve of commodity demand for corn. Shift the blue curve out, and out goes the green curve. If the blue commodity curve becomes more inelastic, the same tends to happen to the green input demand curve.

It is the *interdependence* of productivities of land, labor, and capital that makes the distribution of income a complex topic. Suppose that you were in charge of determining the income distribution of a country. If land had by itself produced so much, and labor had by itself produced so much, and machinery had by itself produced the rest, distribution would be easy. Moreover, under supply and demand, if each factor produced a certain amount by itself, it could enjoy the undivided fruits of its own work.

But reread the above paragraph and underline such words as “by itself.” They refer to a fantasy world of independent productivities which simply does not exist in reality. When an omelette is produced by chef’s labor and chicken’s eggs and cow’s butter and land’s natural gas, how can you unscramble the separate contributions of each input?

To find the answer, we must look to the interaction of marginal productivities and factor supplies—both of which determine the competitive prices and quantities of factors of production.



Review of Production Theory

Before showing the relationship between factor prices and marginal products, we will review the essentials of Chapter 6’s production theory.

The theory of production begins with the notion of the *production function*. The production function indicates the maximum amount of output that can be produced, with a given state of technical knowledge, for each combination of factor inputs. The production-function concept provides a rigorous definition of marginal product. Recall that the *marginal product* of an input is the extra product or output added by 1 extra unit of that input while other inputs are held constant.² The first three columns

² Note that the marginal product of a factor is expressed in *physical* units of product per unit of additional input. So economists sometimes use the term “marginal physical product” rather than “marginal product,” particularly when they want to avoid any possible confusion with a concept we will soon encounter called “marginal revenue product.” For brevity, we will skip the word “physical” and abbreviate marginal product as *MP*.

Marginal Revenue Product				
(1) Unit of labor (workers)	(2) Total product (bushels)	(3) Marginal product of labor (bushels per worker)	(4) Price of output (\$ per bushel)	(5) Marginal revenue product of labor (\$ per worker)
0	0	20,000	3	60,000
1	20,000	10,000	3	30,000
2	30,000	5,000	3	15,000
3	35,000	3,000	3	9,000
4	38,000	1,000	3	3,000
5	39,000			

TABLE 12-3. Calculation of Marginal Revenue Product for Perfectly Competitive Firm

The marginal product of labor is shown in column (3). Marginal revenue product of labor shows how much additional revenue the firm receives when an additional unit of labor is employed. It equals the marginal product in column (3) times the competitive output price in column (4).

of Table 12-3 provide a review of the way marginal products are calculated.

As a final element of review, recall the *law of diminishing returns*. Column (3) of Table 12-3 shows that each successive unit of labor has a declining marginal product. “Declining marginal product” is another name for diminishing returns. Moreover, we can interchange land for labor, varying the amount of land while holding constant labor and other inputs, and we would generally observe the law of diminishing returns at work for land as well as for labor.

marginal product. Before showing this result, we begin by defining some new terms.

Marginal Revenue Product

We can use the tools of production theory to devise a key concept, *marginal revenue product (MRP)*. Suppose we are operating a giant shirt factory. We know how many shirts each additional worker produces. But the firm wants to maximize profits measured in dollars, for it pays salaries and dividends with money, not with shirts. We therefore need a concept that measures the additional *dollars* each additional unit of input produces. Economists give the name “marginal revenue product” to the money value of the additional output generated by an extra unit of input.

The **marginal revenue product** of input A is the additional revenue produced by an additional unit of input A.

DISTRIBUTION THEORY AND MARGINAL REVENUE PRODUCT

The fundamental point about distribution theory is that *the demands for the various factors of production are derived from the revenues that each factor yields on its*

Perfectly Competitive Case. It is easy to calculate marginal revenue product when product markets are perfectly competitive. In this case, each unit of the worker's marginal product (MP_L) can be sold at the competitive output price (P). Moreover, since we are considering perfect competition, the output price is unaffected by the firm's output, and price therefore equals marginal revenue (MR). If we have an MP_L of 10,000 bushels and a price and MR of \$3, the dollar value of the output produced by the last worker—the marginal revenue product of labor (MRP_L)—is \$30,000 (equal to $10,000 \times \$3$). This is shown in column (5) of Table 12-3. Hence, under perfect competition, each worker is worth to the firm the dollar value of the last worker's marginal product; the value of each acre of land is the marginal product of land times the output price; and so forth for each factor.

Table 12-3 provides the essential linkage between production theory and factor demand theory; it should be studied carefully. The first three columns show the inputs, output, and marginal product of labor. Multiplying the MP in column (3) by the price in column (4), we derive the marginal revenue product of labor (in dollars per worker) in column (5). It is this last column which is critical for determining the demand for labor, as we will see later in this chapter. Once we know the wage rate, we can calculate the demand for labor from column (5).

Imperfect Competition. What happens in the case of imperfect competition, where the individual firm's demand curve is downward-sloping? Here, the marginal revenue received from each extra unit of output sold is less than the price because the firm must lower its price on previous units to sell an additional unit. Each unit of marginal product will be worth $MR < P$ to the firm.

To continue our previous example, say that the MR is \$2 while the price is \$3. Then the MRP of the second worker in Table 12-3 would be \$20,000 (equal to the MP_L of $10,000 \times$ the MR of \$2), rather than the \$30,000 of the competitive case.

To summarize:

Marginal revenue product represents the additional revenue a firm earns from using an additional unit of an input, with other inputs held constant. It is calculated as the marginal product of the input multiplied by the marginal revenue obtained from

selling an extra unit of output. This holds for labor (L), land (A), and other inputs. In symbols:

$$\begin{aligned} \text{Marginal revenue product of labor} \\ (MRP_L) = MR \times MP_L \end{aligned}$$

$$\begin{aligned} \text{Marginal revenue product of land} \\ (MRP_A) = MR \times MP_A \end{aligned}$$

and so forth.

Under conditions of perfect competition, because $P = MR$, this implies:

$$\begin{aligned} \text{Marginal revenue product} \\ (MRP_i) = P \times MP_i \end{aligned}$$

for each input.

THE DEMAND FOR FACTORS OF PRODUCTION

Having analyzed the underlying concepts, we now show how profit-maximizing firms decide upon the optimal combination of inputs, which allows us to derive the demand for inputs.

Factor Demands for Profit-Maximizing Firms

What determines the demand for any factor of production? We can answer this question by analyzing how a profit-oriented firm chooses its optimal combination of inputs.

Imagine that you are a profit-maximizing farmer. In your area, you can hire all the farmhands you want at \$20,000 per worker. Your accountant hands you a spreadsheet with the data in Table 12-3. How would you proceed?

You could try out different possibilities. If you hire one worker, the additional revenue (the MRP) is \$60,000 while the marginal cost of the worker is \$20,000, so your extra profit is \$40,000. A second worker gives you an MRP of \$30,000 for an additional profit of \$10,000. The third worker produces extra output yielding revenue of only \$15,000 but costs \$20,000; hence, it is not profitable to hire the third worker. Table 12-3 shows that the maximum profit is earned by hiring two workers.

By using this reasoning, we can derive the rule for choosing the optimal combination of inputs:

To maximize profits, firms should add inputs up to the point where the marginal revenue product of

the input equals the marginal cost or price of the input.

For perfectly competitive factor markets, the rule is even simpler. Recall that under perfect competition the marginal revenue product equals price times marginal product ($MRP = P \times MP$).

The profit-maximizing combination of inputs for a perfectly competitive firm comes when the marginal product times the output price equals the price of the input:

$$\begin{aligned} \text{Marginal product of labor} \times \text{output price} \\ = \text{price of labor} = \text{wage rate} \end{aligned}$$

$$\begin{aligned} \text{Marginal product of land} \times \text{output price} \\ = \text{price of land} = \text{rent} \end{aligned}$$

and so forth.

We can understand this rule by the following reasoning: Say that each kind of input is bundled into little packages each worth \$1—packages of \$1 worth of labor, \$1 worth of land, and so forth. To maximize profits, firms will purchase inputs up to that point where each little \$1 package produces output which is worth just \$1. In other words, each \$1 input package will produce MP units of corn so that the $MP \times P$ just equals \$1. The MRP of the \$1 units is then exactly \$1 under profit maximization.

Least-Cost Rule. We can restate the condition much more generally in a way that applies to both perfect and imperfect competition in product markets (as long as factor markets are competitive). Reorganizing the basic conditions shown above, profit maximization implies:

$$\begin{aligned} \frac{\text{Marginal product}}{\text{Price of labor}} &= \frac{\text{marginal product}}{\text{price of land}} = \dots \\ &= \frac{1}{\text{marginal revenue}} \end{aligned}$$

Suppose that you own a cable television monopoly. If you want to maximize profits, you will want to choose the best combination of workers, land easements for your cables, trucks, and testing equipment to minimize costs. If a month's truck rental costs \$8000 while monthly labor costs per worker are \$800, costs are minimized when the marginal

products *per dollar of input* are the same. Since trucks cost 10 times as much as labor, truck MP must be 10 times labor MP .

Least-cost rule: Costs are minimized when the marginal product per dollar of input is equalized for each input. This holds for both perfect and imperfect competitors in product markets.

Marginal Revenue Product and the Demand for Factors

Having derived the MRP for different factors, we can now understand the demand for factors of production. We just saw that a profit-maximizing firm would choose input quantities such that the price of each input equaled the MRP of that input. This means that from the MRP schedule for an input, we can immediately determine the relationship between the price of the input and the quantity demanded of that input. This relationship is what we call the input demand curve.

Glance back at Table 12-3 on page 235. This table shows in the last column the MRP of labor for our corn farm. By the profit-maximizing condition, we know that at a wage of \$60,000 the firm would choose 1 unit of labor; at a \$30,000 wage, 2 units of labor would be sought; and so forth.

The MRP schedule for each input gives the demand schedule of the firm for that input.

We have used this result in Figure 12-3 to draw a labor demand curve for our corn farm using the data shown in Table 12-3. We have in addition drawn a smooth curve through the individual points to show how the demand curve would appear if fractional units of labor could be purchased.

From Firm to Market Demand. The final step in determining the demand for labor and other factors is the aggregation of the demand curves for different firms. As with all demand curves, the competitive-market demand curve is the *horizontal summation of the demand curves of all the firms*. Hence, if there were 1000 identical firms, then the market demand for labor would be exactly like that in Figure 12-3 except the horizontal axis would have each entry multiplied by 1000. We see, then, that the competitive demand

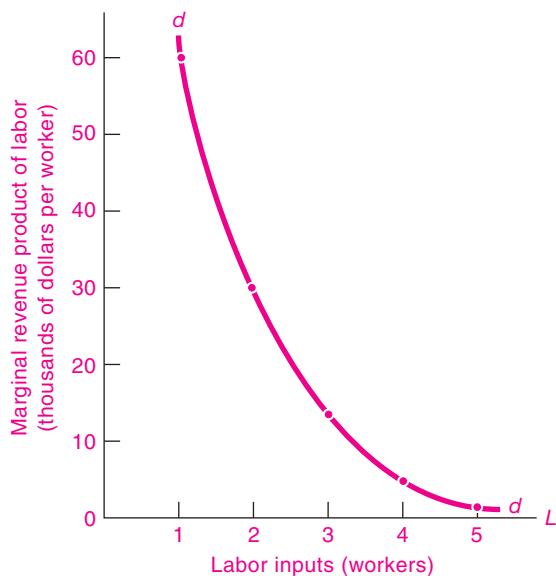


FIGURE 12-3. Demand for Inputs Derived through Marginal Revenue Products

The demand for labor is derived from the marginal revenue product of labor. This figure uses the data for the competitive firm displayed in Table 12-3.

for factors of production is determined by the sum of the demands of all the firms at each marginal revenue product.

Substitution Rule. A corollary of the least-cost rule is the **substitution rule**: If the price of one factor rises while other factor prices remain fixed, the firm will profit from substituting more of the other inputs for the more expensive factor. A rise in labor's price, P_L , will reduce MP_L/P_L . Firms will respond by reducing employment and increasing land use until equality of marginal products per dollar of input is restored—thus lowering the amount of needed L and increasing the demand for land acres. A rise in land's price alone will, by the same logic, cause labor to be substituted for more expensive land. Like the least-cost rule, the substitution rule and the derived demand for factors apply to both perfect and imperfect competition in product markets.

SUPPLY OF FACTORS OF PRODUCTION

A complete analysis of the determination of factor prices and of incomes must combine both the demand for inputs just described and the supplies of different factors. The general principles of supply vary from input to input, and this topic will be explored in depth in the following chapters. At this point we provide a few introductory comments.

In a market economy, most factors of production are privately owned. People “own” their labor in the sense that they control its use; but this crucial “human capital” can today only be rented, not sold. Capital and land are generally privately owned by households and by businesses.

Decisions about *labor* supply are determined by many economic and noneconomic factors. The important determinants of labor supply are the price of labor (i.e., the wage rate) and demographic factors, such as age, gender, education, and family structure. The quantity of *land* and other natural resources is determined by geology and cannot be significantly changed, although the quality of land is affected by conservation, settlement patterns, and improvements. The supply of *capital* depends upon past investments made by businesses, households, and governments. In the short run, the stock of capital is fixed like land, but in the long run the supply of capital reacts to economic factors such as risks, taxes, and rates of return.

Can we say anything about the elasticity of supply of inputs? Actually, the supply curve may slope positively or be vertical and might even have a negative slope. For most factors, we would expect that the supply responds positively to the factor's price in the long run; in this case, the supply curve would slope upward and to the right. The *total* supply of land is usually thought to be unaffected by price, and in this case the *total* supply of land will be perfectly inelastic, with a vertical supply curve. In some special cases, when the return to the factor increases, owners may supply less of the factor to the market. For example, if people feel they can afford to work fewer hours when wages rise, the supply curve for labor might bend backward at high wage rates, rather than slope upward.

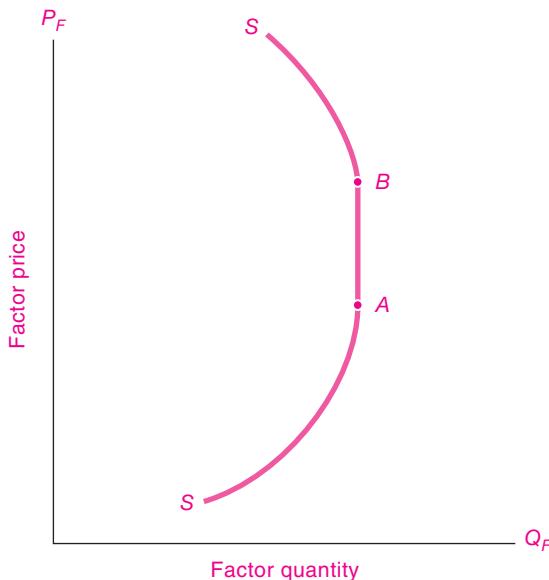


FIGURE 12-4. Supply Curve for Factors of Production

Supplies of factors of production depend upon characteristics of the factors and the preferences of their owners. Generally, supplies will respond positively to price, as in the region below A. For factors that are fixed in supply, like land, the supply curve will be perfectly inelastic, as from A to B. In special cases where a higher price of the factor increases the income of its owner greatly, as with labor or oil, the supply curve may bend backward, as in the region above B.

The different possible elasticities for the supply of factors are illustrated by the SS supply curve shown in Figure 12-4.

DETERMINATION OF FACTOR PRICES BY SUPPLY AND DEMAND

A full analysis of the distribution of income must combine the supply of and demand for factors of production. Earlier parts of this section provided the underpinnings for analysis of demand and gave a brief description of supply. We showed that, for given factor prices, profit-maximizing firms would choose input combinations according to their marginal revenue products. As the price of land falls, each farmer would substitute land for other inputs such as labor, machinery, and fertilizer. Each farmer therefore would show a demand for cornland inputs like that in Figure 12-2(b).

How do we obtain the *market demand* for inputs (whether cornland, unskilled labor, or computers)? We add together the individual demands of each of the firms. Thus at a given price of land, we add together all the demands for land of all the firms at that price; and we do the same at every price of land. In other words, *we add horizontally the demand curves for land of all the individual firms to obtain the market demand curve for land*. We follow the same procedure for any input, summing up all the derived demands of all the businesses to get the market demand for each input. And in each case, the derived demand for the input is based on the marginal revenue product of the input under consideration.³ Figure 12-5 shows a general demand curve for a factor of production as the DD curve.

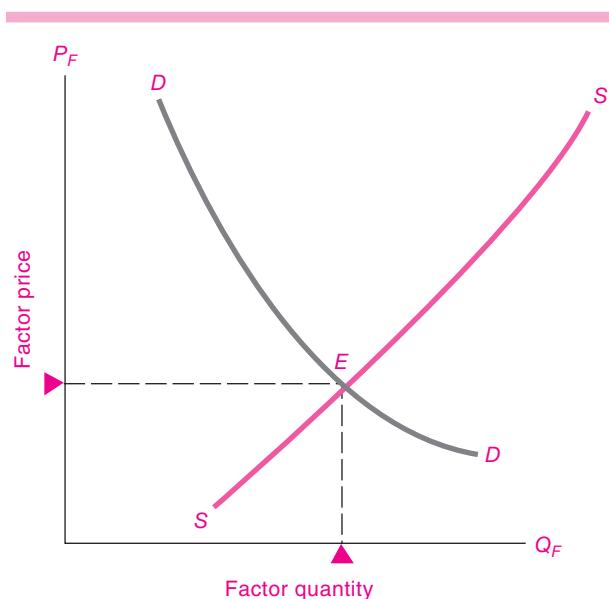


FIGURE 12-5. Factor Supply and Derived Demand
Interact to Determine Factor Prices and Income Distribution

Factor prices and quantities are determined by the interaction of factor supply and demand.

³ Note that this process of adding factor demand curves horizontally is exactly the same procedure that we followed in obtaining market demand curves for goods in Chapter 5.

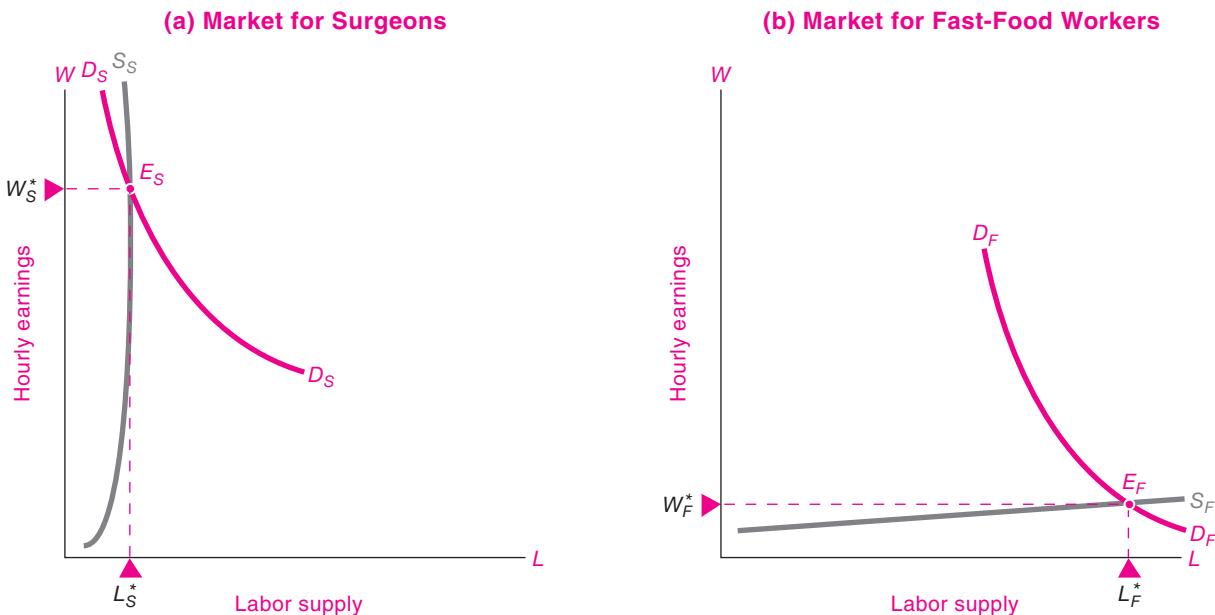


FIGURE 12-6. The Markets for Surgeons and Fast-Food Workers

In (a), we see the impact of a limited supply of surgeons: small output and high earnings per surgeon. What would be the effect on total earnings of surgeons and on the price of an operation if an aging population increased the demand for surgeons?

In (b), open entry and low skill requirements imply a highly elastic supply of fast-food workers. Wages are beaten down and employment is high. What would be the effect on wages and employment if more teenagers looked for jobs?

How do we find the overall market equilibrium? The equilibrium price of the input in a competitive market comes at that level where the quantities supplied and demanded are equal. This is illustrated in Figure 12-5, where the derived demand curve for a factor intersects its supply curve at point E . Only at this price will the amount that owners of the factor willingly supply just balance the amount that the buyers willingly purchase.



The Wages of Slicers and Flippers

We can apply these concepts to two factor markets to see why disparities in incomes are so high. Figure 12-6 shows the markets for two kinds of labor—surgeons and fast-food workers. The supply of surgeons is severely limited by the need for medical licensing and the length and cost of education and training. Demand for surgery is growing rapidly, along

with other health-care services. The result is that surgeons earn \$300,000 a year on average. Moreover, an increase in demand will result in a sharp increase in earnings, with little increase in the number of surgeons.

At the other end of the earnings scale are fast-food workers. These jobs have no skill or educational requirements and are open to virtually everyone. The supply of food workers is highly elastic. As the demand for fast foods increased in recent years, employment grew sharply. Because of the ease of entry into this market, the average full-time fast-food employee was near the bottom of the earnings pyramid at \$19,000 a year. What is the reason for the vast difference in earning power of surgeons and hamburger flippers? It is mainly the quality of labor, not the quantity of hours.

The Rich and the Rest

If you are one of the richest Americans, you might have \$50 million of interest, dividends, and other property income, while the median household earns less than

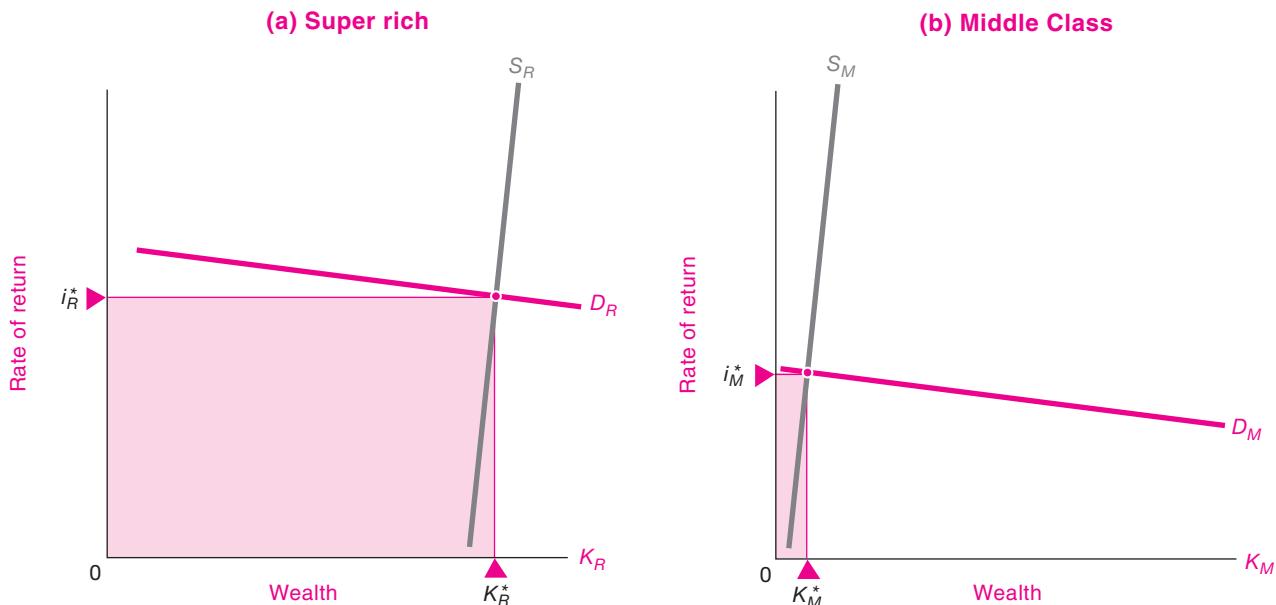


FIGURE 12-7. Differences in Total Returns to Wealth

This figure shows the demand and supply for wealth held by the super rich and the middle class. The horizontal axis shows the total wealth, while the vertical axis shows the rate of return on wealth. The shaded region is $r \times W$, or total income earned on wealth. Why is the shaded rectangle of the rich so much larger than that of the middle class? The reason is primarily that the wealth of the rich (K_R) is so much larger than that of the middle class (K_M).

\$1000 a year on its financial wealth. Figure 12-7 explains this difference. The rate of return on stocks or bonds is not that much higher for the richest than for the middle class.

Rather, the rich have a much bigger wealth base on which to earn. The shaded rectangles in Figure 12-7 show the capital earnings of the two groups. Make sure you understand that it is the amount of wealth rather than the rate of return that makes the rectangle of the top wealth holders so large.

These two examples show how factor prices and individual incomes are determined by underlying market forces. Supply and demand operate to create high returns to factors that have either limited supply or high demand as reflected in high marginal revenue product. If a factor such as surgeons becomes scarcer—say, because training requirements are tightened—the price of this factor will rise and surgeons will enjoy higher incomes. However, if demand decreases in some field like psychiatry—perhaps because insurance companies decide to cut back on psychiatric coverage, or because close substitutes like social workers and

psychologists lure away patients, or because people rely more heavily on medications than on therapy—the lower demand will produce a fall in psychiatrists' incomes. Competition giveth, but competition also taketh away.

THE DISTRIBUTION OF NATIONAL INCOME

With our new understanding of marginal-productivity theory, we can now come back to the question raised at the beginning of the chapter. In a world of intense competition, how do markets allocate national income among the many factors of production?

This section develops the neoclassical theory of factor-income distribution. It can be applied to competitive markets for any number of final products and factor inputs. But it is most easily grasped if we consider a simplified world with only one product in

which all accounts are kept in “real” units, that is, in terms of goods. The goods could be corn or a basket of different goods and services, but we will call it Q . Moreover, by setting the price equal to 1, we can conduct the entire discussion in real terms, with the value of output being Q and with the wage rate being the real wage in terms of goods or Q . In this situation, a production function tells how much Q is produced for each quantity of labor-hours, L , and for each quantity of acres of homogeneous land, A . Note that because $P = 1$, under perfect competition $MRP = MP \times P = MP \times 1 = MP$. The wage is therefore equal to MP_L .

The analysis in the neoclassical model is as follows: A first worker has a large marginal product because there is so much land to work with. Worker 2 has a slightly smaller marginal product. But the two workers are alike, so they must get exactly the same wage. The puzzle is, which wage? The MP of worker 1, or that of worker 2, or the average of the two?

Under perfect competition, the answer is clear: Landlords will not hire a worker if the market wage exceeds that worker’s marginal product. So competition will ensure that *all* the workers receive a wage rate equal to the marginal product of the last worker.

But now there is a surplus of total output over the wage bill because earlier workers had higher MP s than the last worker. What happens to the excess MP s produced by all the earlier workers? The excess stays with the landlords as their residual earnings, which we will later call *rent*. Why, you might ask, do the landlords, who may be sitting on their yachts thousands of miles away, earn anything on the land? The reason is that each landowner is a participant in the competitive market for land and rents the land for its best price. Just as workers compete with each other for jobs, landowners compete with each other for workers. We see in this competitive world no labor unions keeping wages up, no landowners’ conspiracy exploiting workers, and indeed no particular fairness in the wages and rents earned—we see just the operation of supply and demand.

We have therefore determined the total wages paid to labor. Figure 12-8 shows that the marginal product curve of labor gives the demand curve of all employers in terms of real wages. Labor-supply factors determine the supply of labor (shown as SS). The equilibrium wage comes at E . The total wages paid to labor are given by $W \times L$ (for example, if

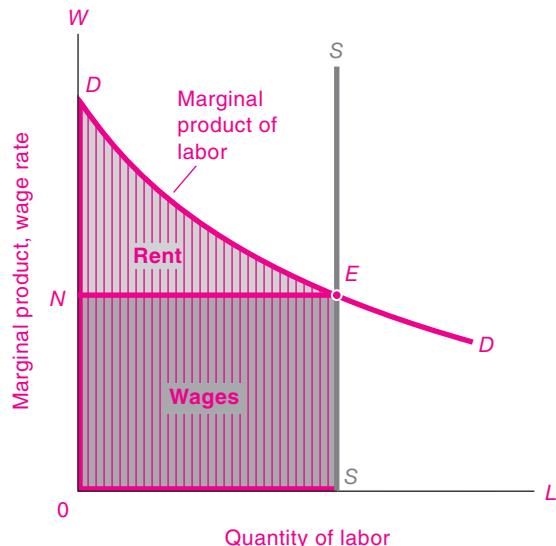


FIGURE 12-8. Marginal Product Principles Determine Factor Distribution of Income

Each vertical slice represents the marginal product of that unit of labor. Total national output $ODES$ is found by adding all the vertical slices of MP up to the total supply of labor at S .

The distribution of output is determined by marginal product principles. Total wages are the lower rectangle (equal to the wage rate ON times the quantity of labor OS). Land rents get the residual upper triangle NDE .

$W = 5$ and $L = 1$ million, total wages = 5 million); this is shown by the dark rectangle, $OSEN$.

Surprisingly, we can also calculate the rent income of land. The light green rent triangle NDE in Figure 12-8 measures all the surplus output which was produced but was not paid out in wages. The size of the rent triangle is determined by how much the MP of labor declines as additional labor is added—that is, by the extent of diminishing returns. If there are only a few high-quality acres, additional units of labor will show sharp diminishing returns and rent’s share will be large. If, by contrast, there is a great deal of homogeneous frontier land just waiting to be cleared, there will be little diminishing returns and land’s rent triangle will be very small.

We have drawn Figure 12-8 so that labor’s wages are about 3 times larger than property’s rents. This 3-to-1 relationship reflects the fact that labor earnings constitute about three-quarters of national income.

The marginal-productivity theory described here is widely used in economics. An important application is to the impact of immigration on wages and profits, which is examined in question 8 at the end of this chapter.

Marginal-Productivity Theory with Many Inputs

The marginal-productivity theory is a great step forward in understanding the pricing of different inputs. Note additionally that the positions of land and labor could be reversed to get a complete theory of distribution. To switch the roles of labor and land, add successive units of variable land to fixed labor. Calculate each successive acre's marginal product.

Then draw a demand curve showing how many acres labor owners will demand of land at each rent rate. In the new version of Figure 12-8 that you draw, find a new E' point of equilibrium. Identify land's rectangle of rent as determined by rent times quantity of land. Identify labor's residual wage triangle. Finally, note the complete symmetry of the factors. This new graph shows that we should think of the distributive shares of each and every factor of production as being simultaneously determined by their interdependent marginal products.

That is not all. Instead of labor and land, suppose the only two factors were labor and some versatile capital goods. Suppose a smooth production function relates Q to labor and capital with the same general properties as in Figure 12-8. In this case, you can redraw Figure 12-8 and get an identical picture of income distribution between labor and capital. Indeed, we can perform the same operation for three, four, or any number of factors.

In competitive markets, the demand for inputs is determined by the marginal products of factors. In the simplified case where factors are paid in terms of the single output, we get

$$\text{Wage} = \text{marginal product of labor}$$

$$\text{Rent} = \text{marginal product of land}$$

and so forth for any factor. This distributes 100 percent of output, no more and no less, among all the factors of production.

We see, then, that the aggregate theory of the distribution of income is compatible with the competitive pricing of any number of goods produced by any

number of factors. This simple but powerful theory shows how the distribution of income is related to productivity in a competitive market economy.

AN INVISIBLE HAND FOR INCOMES?

We have now sketched how a perfectly competitive economy distributes national product among the different inputs in a simplified world.

People naturally ask, Are incomes under market capitalism fair and just? In one sense, this is like asking whether animals get their fair shares of food in the jungle. Just as the battles of the jungle distribute food without regard to right or wrong, so does a competitive market distribute wages and profits according to productivity rather than ethics.

Is there an invisible hand in the marketplace that ensures that the most deserving people will obtain their just rewards? Or that those who toil long hours or nights and weekends or in tedious or dangerous work will receive a decent standard of living? Or that those who work in developing countries will get a comfortable living standard?

In reality, competitive markets do not guarantee that income and consumption will necessarily go to the neediest or most deserving. Laissez-faire competition might lead to great inequality, to malnourished children who grow up to raise more malnourished children, and to the perpetuation of inequality of incomes and wealth for generations. There is no economic law that ensures that the poor countries of Africa will catch up to the rich countries of North America. The rich may get healthier and richer as the poor get sicker and poorer. In a market economy, the distribution of income and consumption reflects not only hard work, ingenuity, and cunning but also factors such as race, gender, location, health, and luck.

While the market can work wonders in producing a growing array of goods and services in an efficient manner, there is no invisible hand which ensures that a laissez-faire economy will produce a fair and equitable distribution of income and property.

Now that we are armed with the general principles underlying the pricing of factors of production and the determination of the distribution of income, we can turn to a detailed discussion of the special features in the three major factor markets—land, labor, and capital.



SUMMARY

A. Income and Wealth

1. Distribution theory is concerned with the basic question of *for whom* economic goods are to be produced. In examining how the different factors of production—land, labor, and capital—get priced in the market, distribution theory considers how supplies and demands for these factors are linked and how they determine all kinds of wages, rents, interest rates, and profits.
2. Income refers to the total receipts or cash earned by a person or household during a given time period (usually a year). Income consists of labor earnings, property income, and government transfer payments.
3. National income consists of the labor earnings and property income generated by the economy in a year. Government takes a share of that national income in the form of taxes and gives back part of what it collects as transfer payments. The post-tax personal income of an individual includes the returns on all the factors of production—labor and property—that the individual owns, plus transfer payments from the government, less taxes.
4. Wealth consists of the net dollar value of assets owned at a given point in time. Wealth is a stock, while income is a flow per unit of time. A household's wealth includes its tangible items such as houses and its financial holdings such as bonds. Items that are of value are called assets, while those that are owed are called liabilities. The difference between total assets and total liabilities is called wealth or net worth.

B. Input Pricing by Marginal Productivity

5. To understand the pricing of different factors of production, we must analyze the theory of production and the derived demand for factors. The demand for inputs is a derived demand: we demand pizza ovens not for their own sake but for the pizzas that they can produce for consumers. Factor demand curves are derived from demand curves for final products. An upward shift in the final demand curve causes a similar upward shift in the derived factor demand curve; greater inelasticity in commodity demand produces greater inelasticity of derived factor demand.
6. We met in earlier chapters the concepts of the production function and marginal products. The demand for a factor is drawn from its marginal

revenue product (*MRP*), which is defined as the extra revenue earned from employing an extra unit of a factor. In any market, *MRP* of a factor equals the marginal revenue earned by the sale of an additional unit of the product times the marginal product of the factor ($MRP = MR \times MP$). For competitive firms, because price equals marginal revenue, this simplifies to $MRP = P \times MP$.

7. A firm maximizes profits (and minimizes costs) when it sets the *MRP* of each factor equal to that factor's marginal cost, which is the factor's price. This can be stated equivalently as a condition in which the *MRP* per dollar of input is equalized for each input. This must hold in equilibrium because a profit-maximizing employer will hire any factor up to the point where the factor's marginal product will return in dollars of marginal revenue just what the factor costs.
8. To obtain the market demand for a factor, we add horizontally all firms' demand curves. This, along with the particular factor's own supply curve, determines the supply-and-demand equilibrium. At the market price for the factor of production, the amounts demanded and supplied will be exactly equal—only at equilibrium will the factor price have no tendency to change.
9. The marginal-productivity theory of income distribution analyzes the way total national income gets distributed among the different factors. Competition of numerous landowners and laborers drives factor prices to equal their marginal products. That process will allocate exactly 100 percent of the product. Any factor, not just labor alone, can be the varying factor. Because each unit of the factor gets paid only the *MP* of the last unit hired, there is a residual surplus of output left over from the *MPs* of early inputs. This residual is exactly equal to the incomes of the other factors under marginal productivity pricing. Hence, the marginal-productivity theory of distribution, though simplified, is a logically complete picture of the distribution of income under perfect competition.
10. Even though a competitive economy may squeeze the maximum amount of bread out of its available resources, one major reservation about a market economy remains. We have no reason to think that incomes will be fairly distributed under laissez-faire capitalism. Market incomes might produce acceptable differences or enormous disparities in income and wealth that persist for generations.

CONCEPTS FOR REVIEW

income distribution
income (flow), wealth (stock)
national income
transfer payments
personal income
marginal product, marginal revenue
product, derived demand

marginal revenue product of input i
 $= MRP_i = MR \times MP_i = P \times MP_i$
for competitive firm
neoclassical theory of income
distribution
 MP rectangle, residual rent triangle

factor demands under competition:
 $MP_i \times P =$ factor price _{i} , which
gives least-cost rule:
$$\frac{MP_L}{P_L} = \frac{MP_A}{P_A} = \dots$$

 $= \frac{1}{\text{marginal revenue}}$
fairness of market incomes

FURTHER READING AND INTERNET WEBSITES

Further Reading

The neoclassical theory of income distribution was developed by one of the pioneers of American economics, John Bates Clark. You can get a flavor of his major ideas in *The Distribution of Wealth: A Theory of Wages, Interest and Profits* (1899) in an online publication at www.econlib.org/library/Clark/clkDW0.html.

Websites

Information on the distribution of income is gathered by the Census Bureau at www.census.gov/hhes/www/income.html. The most comprehensive data on the population

is gathered in the decennial census, available at www.census.gov.

If you want to examine data on income dynamics, an exemplary site for data is that on the Panel Study on Income Dynamics at www.isr.umich.edu/src/psid.

The most comprehensive data on the wealth of Americans is collected by the Federal Reserve Board; see www.federalreserve.gov/PUBS/oss/oss2/scfindex.html.

QUESTIONS FOR DISCUSSION

1. For each of the following factors, name the final output for which the item is a derived demand: wheatland, gasoline, barber, machine tool for basketballs, wine press, economics textbook.
2. Table 12-4 shows the basic numbers for production of pizzas, holding other factors constant.
 - a. Fill in the blanks in columns (3) and (5).
 - b. Construct a diagram like that in Figure 12-3 which shows the marginal revenue product of pizza workers and labor inputs.
 - c. If the wage of pizza workers is \$30 per worker, how many workers will be employed?
- d. Assume that the price of pizzas doubles. Draw the new MRP curve. Estimate the impact on the employment of pizza workers, assuming there are no other changes.
3. Over the last century, hours of work per lifetime have declined about 50 percent while real earnings have increased by a factor of 8. Assuming that the main change was an increase in the marginal-productivity-of-labor schedule, draw supply-and-demand diagrams for labor in 1900 and 2000 that will explain this trend. In your diagrams, put the number of hours worked per lifetime on the horizontal axis and the real wage rate

Marginal Revenue Product				
(1) Unit of labor (workers)	(2) Total product (pizzas)	(3) Marginal product of labor (pizzas per worker)	(4) Price of output (\$ per pizza)	(5) Marginal revenue product of labor (\$ per worker)
0	0			
1	30		5	
2	50		5	
3	60		5	
4	65		5	
5	68		5	
6	68		5	

TABLE 12-4.

on the vertical axis. What key factor about the supply of labor must you invoke to explain this historical trend?

4. Why is each of the following incorrect? State the correct proposition.
 - a. Marginal revenue product is calculated as total revenue earned per worker.
 - b. Distribution theory is simple. You simply figure out how much each factor produces and then give the factor its share of output.
 - c. Under competition, workers get paid the total output produced minus the costs of raw materials.
5. Figure 12-1 shows that the share of labor in national income changed little from 1970 to 2007 even though total incomes (GDP) rose by a factor of three. Draw a set of economywide curves like those in Figure 12-8 which can explain these two facts.
6. Labor leaders used to say, “Without any labor there is no product. Hence labor deserves *all* the product.” Apologists for capital would reply, “Take away all capital goods, and labor scratches a bare pittance from the earth; practically all the product belongs to capital.”

Analyze the flaws in these arguments. If you were to accept the arguments, show that they would allocate 200 or 300 percent of output to two or three factors,

whereas only 100 percent can be allocated. How does the neoclassical marginal-productivity theory resolve this dispute?

7. Draw the supply and demand curves for the oil market. Now suppose that a workable electric car shifts demand away from oil. Draw the new demand curve and the new equilibrium. Describe the outcome in terms of the price of oil, the quantity consumed, and the total income of the oil producers.
8. We can use the neoclassical theory of distribution to analyze the impact of immigration on the distribution of total national income. Assume that there are two factors, homogeneous labor and capital, with returns being wages and profits. Look at Figure 12-9, which has the same variables as Figure 12-8. We begin with initial supply curve S and at equilibrium point A .

Now assume that there is a large increase in labor supply due to immigration, shifting the supply-of-labor curve from S to S' , as shown by the arrow. Assume that all other inputs are unchanged. Answer the following:

- a. Describe and draw the new equilibrium after the immigration.
- b. Explain what will happen to the wage rate.

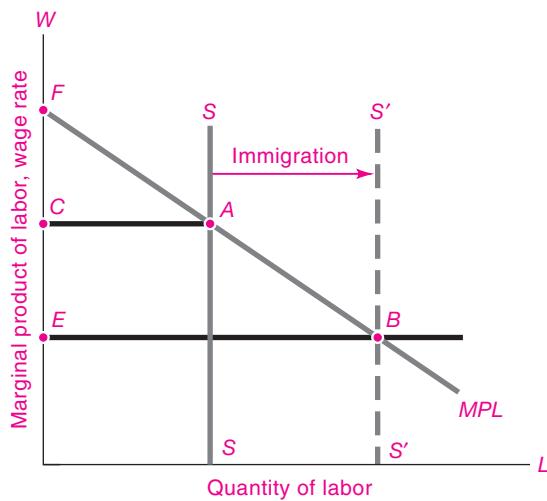


FIGURE 12-9.

- c. Explain what happens to total profits and to the rate of profit (profits per unit of capital).
 - d. Explain why you cannot tell what will happen to total wages or to the share of labor income in total national income.
 - e. Note that this question looks at the impact of immigration on total national income. This analysis appears to differ from Chapter 3's supply-and-demand analysis of the impact of immigration on different cities. Explain the reason why immigration from Mexico to the United States will affect overall wages in the United States in this example, while immigration will not affect wage differentials between Miami and Detroit in the Chapter 3 example.
9. In the marginal-productivity theory shown in Figure 12-8, let land rather than labor be the varying input. Draw a new figure and explain the theory with this new diagram. What is the residual factor?



Work is the curse of the drinking class.

Oscar Wilde

Labor is more than an abstract factor of production. Workers are people who want good jobs with high wages so that they can buy the things they need and want. This chapter explores how wages are set in a market economy. The first section reviews the supply of labor and the determination of wages under competitive conditions. This is followed by a discussion of some of the noncompetitive elements of labor markets, including labor unions and the thorny problem of labor market discrimination.

A. FUNDAMENTALS OF WAGE DETERMINATION

THE GENERAL WAGE LEVEL

In analyzing labor earnings, economists tend to look at the average **real wage**, which represents the purchasing power of an hour's work, or the money wages divided by the cost of living.¹ By that measure, American workers today are far better off than they

were 100 years ago. Figure 13-1 on page 249 shows the real average hourly wage, or the dollar wage adjusted for inflation, along with the average hours of work.

The same powerful gains for workers are found virtually everywhere. Across Western Europe, Japan, and the rapidly industrializing countries of East Asia, there has definitely been a steady, long-term improvement in the average worker's ability to buy food, clothing, and housing, as well as in the health and longevity of the population. In Europe and the United States, these gains began in earnest in the early 1800s, with the advent of the technological and social changes associated with the Industrial Revolution. Before that time real wages meandered up and down, with few long-term gains.

That is not to say that the Industrial Revolution was an unmitigated benefit to workers, especially in the laissez-faire days of the 1800s. In point of fact, a Dickens novel could hardly do justice to the dismal conditions of child labor, workplace dangers, and poor sanitation in factories of the early nineteenth century. A workweek of 84 hours was the prevailing rule, with time out for breakfast and sometimes supper. A good deal of work could be squeezed out of a 6-year-old child, and if a woman lost two fingers in a loom, she still had eight left.

Was it a mistake for people to leave the farms for the rigors of the factory? Probably not. Economic

¹ In this chapter, we will generally use the term "wages" as a shorthand expression for wages, salaries, and other forms of compensation.

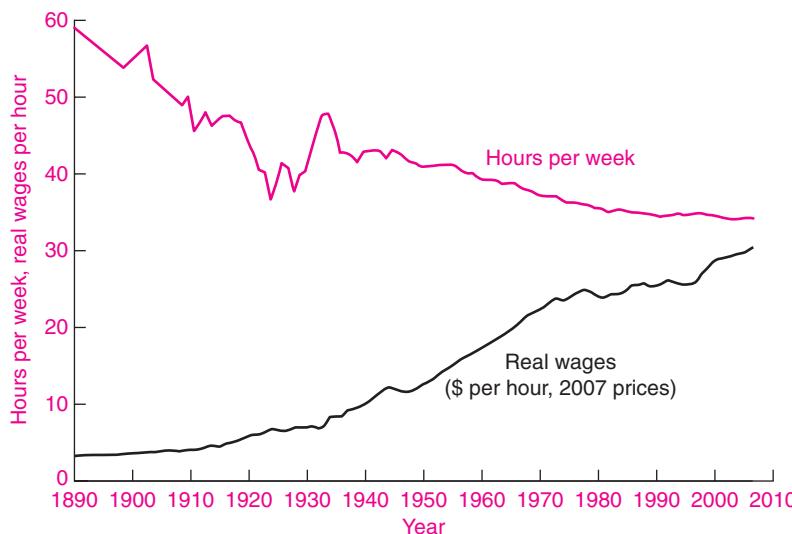


FIGURE 13-1. Wages Have Improved as Hours of Work Have Declined

With advancing technology and improved capital goods, American workers enjoy higher wages while working fewer hours. These are the fruits of long-term economic growth.

historians emphasize that even with the demanding conditions in the factories, living standards were nevertheless greatly improved over those in the earlier centuries of agrarian feudalism. The Industrial Revolution was a giant step forward for the working class, not a step back. The idyllic picture of the healthful, jolly countryside peopled by stout yeomen and happy peasantry is a historical myth unsupported by statistical research.

DEMAND FOR LABOR

Marginal Productivity Differences

We begin our examination of the general wage level by examining the factors underlying the demand for labor. The basic tools were provided in the previous chapter, where we saw that the demand for a factor of production reflects the marginal productivity of that input.

Figure 13-2 illustrates the marginal-productivity theory. Holding technology and other inputs constant, there exists a relationship between the quantity of labor inputs and the amount of output. By the law of diminishing returns, each additional unit of labor input will add a smaller and smaller slab of output. In the example shown in Figure 13-2, at 10 units of labor, the competitively determined general wage level will be \$20 per unit.

But probe deeper and ask what lies behind marginal product. To begin with, the marginal

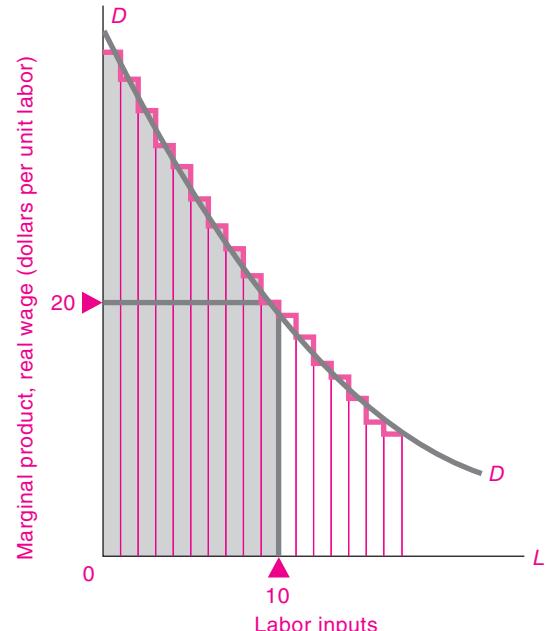


FIGURE 13-2. Demand for Labor Reflects Marginal Productivity

The demand for labor is determined by its marginal productivity in producing national output. The light blue vertical slices represent the extra output produced by the first, second, ... unit of labor. The competitively determined general wage level at 10 units of labor is \$20 per unit, equal to the marginal productivity of the tenth unit. The labor demand curve shifts up and out over time with capital accumulation, technological advance, and improvements in labor quality.

productivity of labor will rise if workers have more or better capital goods to work with. Compare the productivity of a ditchdigger using a bulldozer with that of a similar digger using a hand shovel, or the communications capabilities of medieval messengers with modern e-mail. Second, marginal productivity of better-trained or better-educated workers will generally be higher than that of workers with less “human capital.”

These reasons explain why wages and living standards rose so much during the twentieth century. Wages are high in the United States and other industrial countries because these nations have accumulated substantial capital stocks: dense networks of roads, rails, and communications; substantial amounts of plant and equipment for each worker; and adequate inventories of spare parts. Even more important are the vast improvements in technologies compared to those of an earlier era. We have seen lightbulbs replace oil lamps, airplanes replace horses, xerography replace quill and ink, computers replace abacuses, and Internet commerce invade traditional ways of doing business. Just imagine how productive the average American would be today with the technologies of 1900.

The quality of labor inputs is another factor determining the general wage level. By any measure—literacy, education, or training—the skills of the American workforce today are superior to those of 1900. Years of education are necessary to produce an engineer capable of designing precision equipment. A decade of training must precede the ability to perform successful brain surgery. As the workforce increases its education and skills, this increases the productivity of labor.

International Comparisons

The same reasoning explains why wage levels differ so dramatically across the world. Look at Table 13-1, which shows average wages plus benefits in manufacturing industries for eight countries. Note that hourly wages in the United States are lower than those in Europe but almost 20 times higher than in China.

What accounts for the enormous differences? It's not that governments in China and Mexico are suppressing wage increases, though government policies do have some impact on the minimum wage and other aspects of the labor market. Rather, real

Region	Wages and fringe benefits in manufacturing, 2006 (\$ per hour)
Germany	34.21
Italy	25.07
United States	23.82
Japan	20.20
South Korea	14.72
Mexico	2.75
China	1.37
Philippines	1.07

TABLE 13-1. General Wage Levels Vary Enormously across Countries

Western European nations, Japan, and the United States are high-wage countries, while China's hourly wages are a tiny fraction of American levels. General wage levels are determined by supply and demand for labor, but other factors such as capital, education levels, technology levels, and civil strife have a major impact on supply and demand curves.

Source: U.S. Bureau of Labor Statistics at <ftp://ftp.bls.gov/pub/special.requests/ForeignLabor/ichccpusupp02.txt> and estimates by the authors. Note these estimates use market exchange rates and not purchasing-power-parity exchange rates.

wages differ among countries primarily because of the operation of the supply and demand for labor. Look at Figure 13-3. Suppose that Figure 13-3(a) represents the state of affairs in the United States while Figure 13-3(b) describes Mexico. In Figure 13-3(a), the supply of U.S. workers is shown by the supply curve, $S_{US}S_{US}$, while the demand for workers is represented by $D_{US}D_{US}$. The equilibrium wage will settle at the level shown at E_{US} . If the wage were lower than E_{US} , shortages of labor would occur and employers would bid up wages to E_{US} , restoring the equilibrium. Similar forces determine E_M , the Mexican wage.

We see that the Mexican wage is lower than the U.S. wage principally because the Mexican demand curve for labor is far lower as a result of the low marginal productivity of labor in Mexico. The most important factor lies in the quality of the workforce. The average education level in Mexico falls far short of the American standard, with a substantial fraction of the population illiterate. Additionally, compared to the United States, a country like Mexico has much less capital to work with: many of the roads are unpaved, few computers and fax machines are

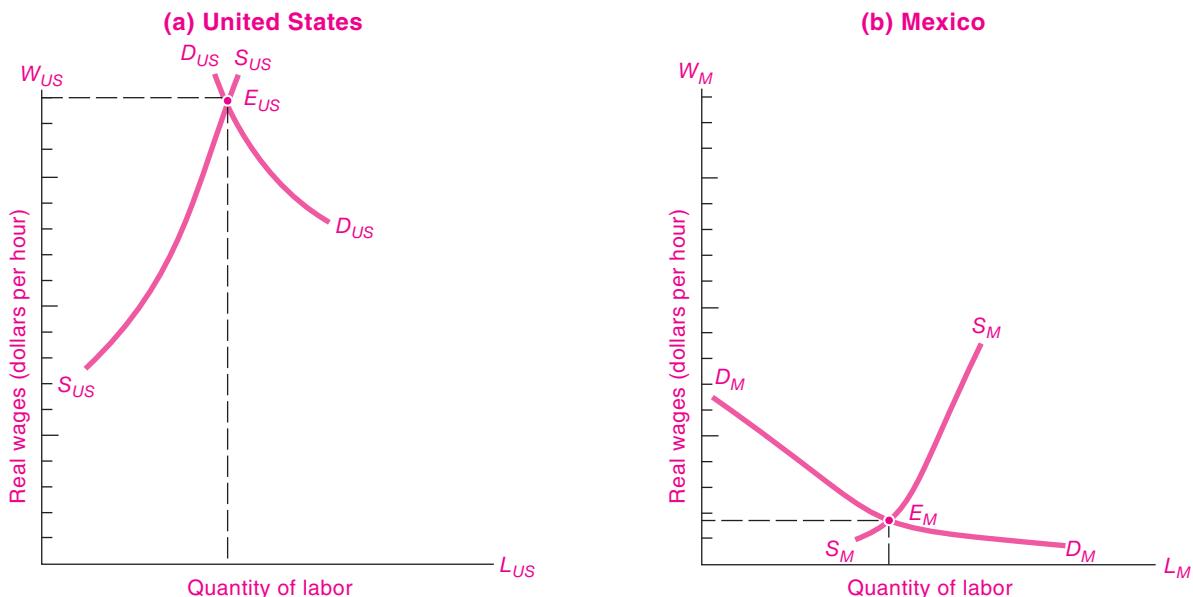


FIGURE 13-3. Favorable Resources, Skills, Management, Capital, and Technology Explain High U.S. Wages

Supply and demand determine a higher competitive wage in the United States than in Mexico. The major forces leading to high U.S. wages are a better-educated and more skilled workforce, a larger stock of capital per worker, and modern technologies.

in use, and much of the equipment is old or poorly maintained. All these factors make labor's marginal productivity low and tend to reduce wages.

This analysis can also help explain why wages have risen rapidly in East Asian regions such as Hong Kong, South Korea, and Taiwan. These economies are devoting a sizable share of their outputs to educating their populations, investing in new capital goods, and importing the latest productive technologies. The MP and DD curves for these countries have shifted greatly upward and to the right. As a result, real wages have doubled over the last 20 years in these countries, while wages have stagnated in relatively closed countries which invest less in education, public health, and tangible capital.

THE SUPPLY OF LABOR

Determinants of Supply

So far we have focused on the demand side of the labor market. Now we turn to the supply side. *Labor supply* refers to the number of hours that the

population desires to work in gainful activities. The three key elements for labor supply are hours per worker, labor-force participation, and immigration.

Hours Worked. While some people have jobs with flexible hours, most Americans work between 35 and 40 hours a week, without much leeway to increase or cut back their weekly hours. However, most people do have a great deal of control over how many hours they work over the course of their lifetimes. They may decide to go to college, to retire early, or to work part-time rather than full-time—all of these can reduce the number of total lifetime hours worked. On the other hand, the decision to take on a second job will increase the lifetime hours worked.

Suppose that wages rise. Will that increase or decrease the lifetime hours of work? Look at the supply curve of labor in Figure 13-4. Note how the supply curve rises at first; then at the critical point C , it begins to bend back. How can we explain why higher wages may first increase and then decrease the quantity of labor supplied?

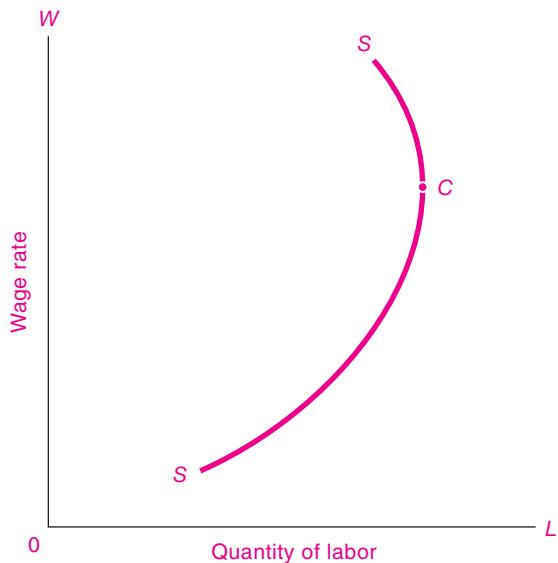


FIGURE 13-4. As Wages Rise, Workers May Work Fewer Hours

Above the critical point *C*, raising the wage rate reduces the amount of labor supplied as the income effect outweighs the substitution effect. Why? Because at higher wages workers can afford more leisure even though each extra hour of leisure costs more in wages forgone.

Put yourself in the shoes of a worker who has just been offered higher hourly rates and is free to choose the number of hours to be worked. You are tugged in two different directions. On one side is the *substitution effect*. (Chapter 5 explained that the substitution effect operates when people consume more of, or substitute in favor of, a good whose relative price falls and consume less of a good whose relative price increases.) Because each hour of work is now better paid, each hour of leisure has become more expensive; you thus have an incentive to substitute extra work for leisure.

But acting against the substitution effect is the *income effect*. With the higher wage, your income is higher. With a higher income, you will want to buy more goods and services, and, in addition, you will want more leisure time. You can afford to take longer vacations or to retire earlier than you otherwise would.

Which will be more powerful, the substitution effect or the income effect? There is no single correct

answer; it depends upon the individual. In the case shown in Figure 13-4, for all wage rates below point *C*, labor supplied increases with a higher wage: the substitution effect outweighs the income effect. But from point *C* upward, the income effect outweighs the substitution effect, and labor supplied declines as wage rates climb higher.

Labor-Force Participation. One of the most dramatic developments in recent decades has been the sharp influx of women into the workforce. The labor-force participation rate of women (i.e., the fraction of women over 15 employed or actively looking for jobs) has jumped from 34 percent in 1950 to 60 percent today. In part this can be explained by rising real wages, which have made working more attractive for women. However, a change of this magnitude cannot be explained by economic factors alone. To understand such a significant alteration in working patterns, one must look outside economics to changing social attitudes toward the role of women as mothers, homemakers, and workers.

Immigration. The role of immigration in the labor-force supply has always been important in the United States. Whereas only 5 percent of the U.S. population was foreign-born in 1970, by 2008 that number had risen to 12 percent.

The flow of legal immigrants is controlled by an intricate quota system which favors skilled workers and their families, as well as close relatives of U.S. citizens and permanent residents. In addition, there are special quotas for political refugees. Most immigrants today are undocumented ("illegal") people who enter the United States looking for better economic opportunities. In recent years, the biggest groups of legal immigrants have come from places like Mexico, the Philippines, Vietnam, and some of the Central American and Caribbean countries.

The major change in immigration in recent decades has been a change in the characteristics of immigrants. In the 1950s, Germany and Canada were the major sources, while in the 1980s and 1990s Mexico and the Philippines were the dominant sources. As a result, recent immigrants have been relatively less skilled and less educated than those of an earlier age.

From the point of view of labor supply, the overall effect of recent immigration has been an increase

in the supply of low-skilled workers in the United States relative to high-skilled workers. Studies have estimated that this change in supply has contributed to the decline in the wages of less educated groups relative to the college-educated.

Empirical Findings

Theory does not tell us whether the labor supply of a group will react positively or negatively to a wage change. Will an income-tax increase on high-income workers—which reduces their after-tax wages—cause them to reduce their work hours? Will subsidizing the wages of the working poor reduce or increase their hours worked? These vital questions must be considered by policymakers as they weigh issues of equity and efficiency. We often need to know the exact shape or elasticity of the labor supply curve.

Table 13-2 presents a summary of numerous studies of the subject. This survey shows that the labor supply curve for adult males appears to be slightly backward-bending, while the responses of other

demographic groups look more like a conventional upward-sloping supply curve. For the population as a whole, labor supply appears to respond very little to changes in real wages.

WAGE DIFFERENTIALS

While analysis of the general wage level is important for comparing different countries and times, we often want to understand *wage differentials*. In practice, wage rates differ enormously. The average wage is as hard to define as the average person. A hedge-fund manager may earn \$400 million a year, while a hedge-fund janitor may earn \$400 a week. A doctor may earn 20 times more than a lifeguard even though both are saving lives.

There are major differences in earnings among broad industry groups, as is shown in Table 13-3. Sectors with small firms such as farming, retail trade, or private households tend to pay low wages, while the larger firms in manufacturing pay twice as much. But

Group of workers	Labor-Supply Patterns		Response of labor supply to increase in real wages
	1960	2007	
Adult males (25 to 54 years)	97	91	Supply curve found to be backward-bending in most studies. Income effect dominates substitution effect. Elasticities are around -0.1 for prime-age males.
Adult females (25 to 54 years)	43	76	Females generally have shown significant positive labor-supply elasticities.
Teenagers	48	40	Teenage response is highly variable.
Seniors (65 and older)	21	16	Seniors have been responsive to relative generosity of retirement programs relative to wages.
Entire population (16 and over)	60	66	Elasticity of total labor supply is close to zero, with income effects balancing out substitution effects. Estimated labor-supply elasticity for entire population is in the range from 0.0 to 0.2.

TABLE 13-2. Empirical Estimates of Labor-Supply Responses

Economists have devoted careful study to the response of labor supply to real wages. For prime-age males (the quaint term used to designate males between 25 and 54), the supply curve is backward-bending (that is, the elasticity is negative), while teenagers and adult females generally respond positively to wages. For the economy as a whole, the labor supply curve is close to completely inelastic or vertical.

Source: U.S. Department of Labor, *Employment and Earnings*, March 2008.

Compensation by Industry	
Industry	Average earnings per full-time employee, 2006* (\$ per year)
All industries	47,000
Farms	30,400
Mining	79,200
Manufacturing	52,300
Retail trade	29,400
Finance and insurance	82,800
Securities and related	205,600
Accommodation and food services	20,800
Food services	18,900

* Total compensation per full-time equivalent worker.

TABLE 13-3. Earnings Vary by Industry

Average annual wages and salaries in broad industry groups range from a high of \$82,800 in finance to a low of \$20,800 in accommodation and food services. In narrow industry groups, earnings vary enormously between security analysts and food-service workers.

Source: U.S. Bureau of Economic Analysis at www.bea.gov, Table 6.6D in the complete NIPA tables.

within major sectors there are large variations that depend on worker skills and market conditions—fast-food workers make much less than doctors even though they all provide services.

How can we explain these wage differentials? Let's consider first a *perfectly competitive labor market*, one in which there are large numbers of workers and employers, none of which has the power to affect wage rates appreciably. Few labor markets are perfectly competitive in reality, but some (such as a large city's market for teenage workers or clerical workers) approach the competitive concept reasonably closely. If all jobs and all people are identical in a perfectly competitive labor market, competition will cause the hourly wage rates to be exactly equal. No employer would pay more for the work of one person than for that person's identical twin or for another person who possessed identical skills.

This means that to explain the pervasive wage differences across industries or individuals, we must look to either differences in jobs, differences in people, or imperfect competition in labor markets.

Differences in Jobs: Compensating Wage Differentials

Some of the tremendous wage differentials observed in everyday life arise because of differences in the quality of jobs. Jobs differ in their attractiveness; hence wages may have to be raised to coax people into the less attractive jobs.

Wage differentials that serve to compensate for the relative attractiveness, or nonmonetary differences, among jobs are called **compensating differentials**.

Window washers must be paid more than janitors because of the risks of climbing skyscrapers. Workers often receive 5 percent extra pay on the 4 P.M. to midnight "swing shift" and 10 percent extra pay for the midnight to 8 A.M. "graveyard shift." For hours beyond 40 per week or for holiday and weekend work, 1½ to 2 times the base hourly pay is customary. Jobs that involve hard physical labor, tedium, low social prestige, irregular employment, seasonal layoff, or physical risk all tend to be less attractive. No wonder, then, that companies must pay \$50,000 to \$80,000 a year to recruit people to work at dangerous and lonely jobs on offshore oil platforms or in northern Alaska. Similarly, for jobs that are especially pleasant or psychologically rewarding, such as those of park rangers and the clergy, pay levels tend to be modest.

To test whether a given difference in pay between two jobs is a compensating differential, ask people who are well qualified for both jobs: "Would you take the higher-paying job in preference to the lower?" If they are not eager to take the higher-paying job, the pay difference is probably a compensating differential that reflects the nonmonetary differences between the jobs.

Differences in People: Labor Quality

We have just seen that some wage differentials serve to compensate for the differing degrees of attractiveness of different jobs. But look around you. Garbage collectors make much less than lawyers, yet surely the legal life has higher prestige and much more pleasant working conditions. We see countless examples of high-paying jobs that are more pleasant than low-paying work. We must look to factors beyond compensating differentials to explain the reason for most wage differences.

One key to wage disparities lies in the qualitative differences among people. A biologist might classify

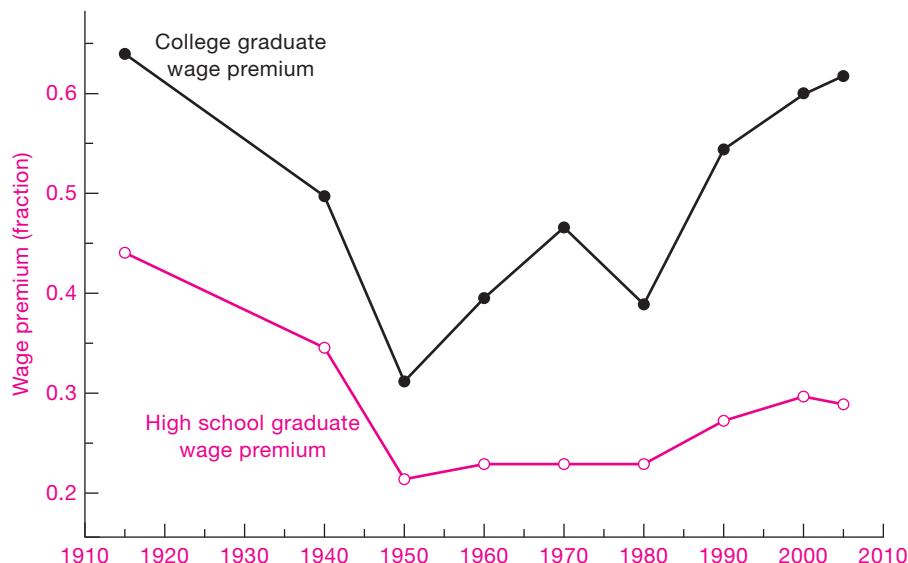


FIGURE 13-5. Relative Income Gains Have Been Dramatic for College Graduates

The education premium for college and high school has increased sharply in recent years. The college premium shows the income advantage of college graduates relative to high school graduates, while the high school premium shows the advantage relative to those who complete eighth grade. Note how sharply the college premium grew after 1980.

Source: Claudia Goldin and Lawrence F. Katz, *The Race between Education and Technology* (Harvard University Press, Cambridge, Mass., 2008).

all of us as members of the species *Homo sapiens*, but a personnel officer would insist that people differ enormously in their abilities to contribute to a firm's output.

While many of the differences in labor quality are determined by noneconomic factors, the decision to accumulate **human capital** can be evaluated economically. The term "human capital" refers to the stock of useful and valuable skills and knowledge accumulated by people in the process of their education and training. Doctors, lawyers, and engineers invest many years in their formal education and on-the-job training. They spend large sums on tuition and wages forgone and often work long hours. Part of the high salaries of these professionals should be viewed as a return on their investment in human capital—a return on the education that makes these highly trained workers a very special kind of labor.

Economic studies of incomes and education show that human capital is a good investment on average. Figure 13-5 shows the ratio of the hourly

earnings of college graduates to those of high school graduates. Relative earnings rose sharply after 1980 as the "price of skill" rose.



Should You Invest in Human Capital?

Students may be surprised to learn that every day in college is an investment in human capital. When students go to college, each year they pay thousands of dollars in tuition and earnings forgone. This cost is just as much an investment as buying a bond or a house.

Does college actually pay off? The evidence suggests that it pays off smartly for the average graduate. Look at Figure 13-5. Suppose that the total investment in college is \$200,000 and that a high school graduate earns \$40,000 per year. If the college premium is 60 percent, this says that a college graduate would earn \$64,000 per year. This represents a \$24,000 return on the investment, or around 12 percent per year. While this would not hold for

everyone, it does suggest why students are working hard to get into good colleges.

Why has the college premium risen so sharply? More and more, in today's service economy, companies are processing information rather than raw materials. In the information economy, the skills learned in college are a prerequisite for a high-paying job. A high school dropout is generally at a severe disadvantage in the job market. Even if you have to borrow for your education, put off years of gainful employment, live away from home, and pay for rent and books, your lifetime earnings in the occupations that are open only to college graduates will probably more than compensate you for the costs.

Often, people point to the role of luck in determining economic circumstances. But, as Louis Pasteur remarked, "Chance favors the prepared mind." In a world of rapidly changing technologies, education prepares people to understand and profit from new circumstances.

Differences in People: The "Rents" of Unique Individuals

For the lucky few, fame has lifted incomes to astronomical levels. Software guru Bill Gates, investment wizard Warren Buffett, basketball star Shaquille O'Neal, and even economists who consult for business can earn fabulous sums for their services.

These extremely talented people have a particular skill that is highly valued in today's economy. Outside their special field, they might earn but a small fraction of their high incomes. Moreover, their labor supply is unlikely to respond perceptibly to wages that are 20 or even 50 percent higher or lower. Economists refer to the excess of these wages above those of the next-best available occupation as a pure economic rent; these earnings are logically equivalent to the rents earned by fixed land.

Some economists have suggested that technological changes are making it easier for a small number of top individuals to serve a larger share of the market. The "winners" in athletics, entertainment, and finance far outdistance the runners-up in the race for compensation. Top entertainers or athletes can now give a single performance that reaches a billion people via television and recordings—something that was not possible just a few years ago. If this trend continues, and labor rents rise further, the income gap between the winners and the runners-up may widen even further in the years ahead.

Segmented Markets and Noncompeting Groups

Even in a perfectly competitive world where people could move easily from one occupation to another, substantial wage differentials would appear. These differences would be necessary to reflect differences in the costs of education and training or in the unattractiveness of certain occupations or to indicate rewards for unique talents.

But even after taking into account all these reasons for wage differentials, we still find a large disparity in wage rates. The major reason for the remaining difference is that labor markets are segmented into *noncompeting groups*.

A moment's thought will suggest that, instead of being a single factor of production, labor is many different, but closely related, factors of production. Doctors and economists, for example, are noncompeting groups because it is difficult and costly for a member of one profession to enter into the other. Just as there are many different kinds of houses, each commanding a different price, so are there many different occupations and skills that compete only in a general way. Once we recognize the existence of many different submarkets of the labor market, we can see why wages may differ greatly among groups.

Why is the labor market divided into so many noncompeting groups? The major reason is that, for the professions like law and medicine, it takes a large investment of time and money to become proficient. If coal mining declines because of environmental restrictions, the miners can hardly hope to land jobs teaching environmental economics overnight. Once people specialize in a particular occupation, they become part of a particular labor submarket. They are thereby subject to the supply and demand for that skill and will find that their own labor earnings rise and fall depending upon events in that occupation and industry. Because of this segmentation, the wages for one occupation can diverge substantially from the wages in other areas.

The job choice of new immigrants is a classic case of noncompeting groups. Rather than just answering random classified ads, new immigrants from a particular country tend to cluster in certain occupations. For example, in many cities, such as Los Angeles and New York, a large number of grocery stores tend to be owned by Koreans. The reason is that the Koreans can get advice and support from friends and relatives

Summary of Competitive Wage Determination	
Labor situation	Wage result
1. People are all alike—jobs are all alike.	No wage differentials
2. People are all alike—jobs differ in attractiveness.	Compensating wage differentials
3. People differ, but each type of labor is in unchangeable supply (noncompeting groups).	Wage differentials that reflect supply and demand for segmented markets
4. People differ, but there is some mobility among groups (partially competing groups).	General-equilibrium pattern of wage differentials as determined by general demand and supply (includes 1 through 3 as special cases)

TABLE 13-4. Market Wage Structure Shows Great Variety of Patterns under Competition

who also own grocery stores. As immigrants get more experience and education in the United States and become fluent in English, their job choice widens and they become part of the overall labor supply.

In addition, the theory of noncompeting groups helps us understand labor market discrimination. We will see in the next section of this chapter that much discrimination arises because workers are separated by gender, race, or other personal characteristics into noncompeting groups as a result of custom, law, or prejudice.

While the theory of noncompeting groups highlights an important aspect of labor markets, we must recognize that in the longer run entry and exit will reduce differentials. It is true that copper miners are unlikely to become computer programmers when computers and fiber optics displace rotary dials and copper wires. Consequently, we may see wage differentials arise between the two kinds of labor. But in the longer run, as more young people study computer science rather than go to work in copper mines, competition will tend to reduce the differentials of these noncompeting groups.

Table 13-4 summarizes the different forces at work in determining wage rates in competitive conditions.

B. LABOR MARKET ISSUES AND POLICIES

Our survey has up to now examined the case of competitive labor markets. In reality, distortions prevent the operation of perfect competition in

labor markets. One source of imperfect competition is labor unions. Unions represent a significant, although shrinking, fraction of workers. A second facet of labor markets is discrimination—also less important than in earlier decades, but still an issue to consider. Yet another factor acting on labor markets is government policies. By setting minimum wages (discussed in Chapter 4), encouraging or discouraging unions, or outlawing discrimination, governments have a powerful effect on labor markets.

THE ECONOMICS OF LABOR UNIONS

Sixteen million Americans, or 12 percent of wage and salary workers, belonged to labor unions in 2007. Unions definitely have market power and sometimes serve as monopoly suppliers of labor. Unions negotiate collective-bargaining agreements which specify who can fill different jobs, how much workers will be paid, and what the work rules are. And unions can decide to go on strike—withdraw their labor supply completely and even cause a factory to shut down—in order to win a better deal from an employer. The study of unions is an important part of understanding the dynamics of labor markets.

The wages and fringe benefits of unionized workers are determined by **collective bargaining**. This is the process of negotiation between representatives of firms and of workers for the purpose of establishing mutually agreeable conditions of employment. The centerpiece is the *economic package*. This includes the basic wage rates for different job categories, along with the rules for holidays and coffee breaks.

In addition, the agreement contains provisions for fringe benefits such as a pension plan, coverage for health care, and similar items.

A second important issue is *work rules*. These concern work assignments and tasks, job security, and workloads. Particularly in declining industries, the staffing requirements are a major issue because the demand for labor is falling. In the railroad industry, for example, there were decades of disputes about the number of people needed to run a train.

Collective bargaining is a complicated business, a matter of give-and-take. Much effort is spent negotiating purely economic issues, dividing the pie between wages and profits. Sometimes agreements get hung up on issues of management prerogatives, such as the ability to reassign workers or change work rules. In the end, both workers and management have a large stake in ensuring that workers are satisfied and productive on their jobs.

Government and Collective Bargaining

The legal framework is an important determinant of economic organization. Two hundred years ago, when labor first tried to organize in England and America, common-law doctrines against “conspiracy in restraint of trade” were used to block unions. In the early 1900s, unions and their members were convicted by courts, fined, jailed, and harassed by various injunctive procedures. The Supreme Court repeatedly struck down acts designed to improve working conditions for women and children and other reform legislation on hours and wages.

It was only after the pendulum swung toward support of unions and collective bargaining that the explosive growth of unions began. A major landmark was the Clayton Act (1914), designed to remove labor from antitrust prosecution. The Fair Labor Standards Act (1938) barred child labor, called for time-and-a-half pay for weekly hours over 40, and set a federal minimum wage for most nonfarm workers.

The most important labor legislation of all was the National Labor Relations (or Wagner) Act of 1935. This law stated: “Employees shall have the right to . . . join . . . labor organizations, to bargain collectively . . . , and to engage in concerted activities.” Spurred by pro-labor legislation, union membership rose from less than one-tenth of the labor force in the 1920s to one-quarter of the workforce by the end of World War II. The decline of American unions began

in the early 1970s. In essence, the monopoly power of unions was eroded by the deregulation of many industries, increased international competition, and a less favorable government attitude toward unions.

HOW UNIONS RAISE WAGES

How can labor unions raise the wages and improve the working conditions of their members? *Unions gain market power by obtaining a legal monopoly on the provision of labor services to a particular firm or industry.* Using this monopoly, they compel firms to provide wages, benefits, and working conditions that are above the competitive level. For example, if nonunion plumbers earn \$20 per hour in Alabama, a union might bargain with a large construction firm to set the wage at \$30 per hour for that firm’s plumbers.

Such an agreement is, however, valuable to the union only if the firm’s access to alternative labor supplies can be restricted. Hence, under a typical collective-bargaining agreement, firms agree not to hire nonunion plumbers, not to contract out plumbing services, and not to subcontract to non-union firms. Each of these provisions helps prevent erosion of the union’s monopoly on the supply of plumbers to the firm. In some industries, like steel and auto manufacturing unions will try to unionize the entire industry so that firm A’s unionized workers need not compete with firm B’s nonunion workers. All these steps are necessary to protect high union wage rates.

Figure 13-6 shows the impact of agreed-upon high standard wages. Here, the union forces employers to pay wages at the standard rate shown by the horizontal line *rr*. The equilibrium is at *E'*, where *rr* intersects the employers’ demand curve. Note that the union has not directly reduced supply when it sets high standard wage rates. Rather, at the high wage rates, employment is limited by the firms’ demand for labor. The number of workers who seek employment exceeds the demand by the segment *E'F*. These excess workers might be unemployed and waiting for vacancies in the high-paying union sector, or they might become discouraged and look for jobs in other sectors. The workers from *E'* to *F* are as effectively excluded from jobs as they would be if the union had directly limited entry.

The need to prevent nonunion competition also explains many of the political goals of the national

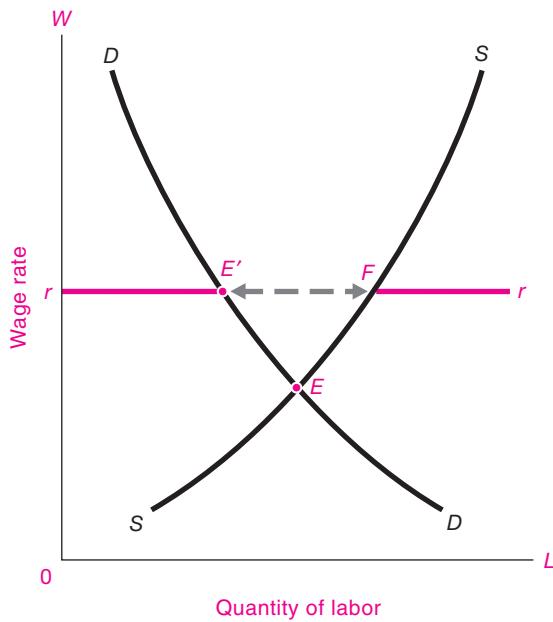


FIGURE 13-6. Unions Set High Standard Wage and Limit Employment

Raising the standard wage to rr increases wages and decreases the employment in the unionized labor market. Because of supply and demand imbalance, workers from E' to F cannot find employment in this market.

If unions push real wages too high for an entire economy, firms will demand E' while workers will supply F . Thus the blue arrow from E' to F represents the amount of classical unemployment. This source of unemployment is particularly important when a country cannot affect its price level or exchange rate, and it differs from the unemployment caused by insufficient aggregate demand.

labor movement. It explains why unions want to limit immigration; why unions support protectionist legislation to limit imports of foreign goods, which are goods made by workers who are not members of American unions; why quasi unions like medical associations fight to restrict the practice of medicine by other groups; and why unions sometimes oppose deregulation in industries such as trucking, communications, and airlines.

Theoretical Indeterminacy of Collective Bargaining

In most collective-bargaining negotiations, the workers press for higher wages while management holds

out for lower compensation costs. This is a situation known as *bilateral monopoly*—where there is but one buyer and one seller. The outcome of bilateral monopoly cannot be predicted by economic forces of costs and demands alone; it depends as well on psychology, politics, and countless other intangible factors.

EFFECTS ON WAGES AND EMPLOYMENT

The advocates of labor unions claim that unions have raised real wages and have benefited workers. Critics argue that the result of raising wages is high unemployment, inflation, and distorted resource allocation. What are the facts?

Has Unionization Raised Wages?

Let's start by reviewing the effects of unions on relative wages. If we look at all private industrial workers in 2006, union workers had average hourly earnings about 15 percent above those of nonunion workers. However, this raw number does not reflect the fact that the skill, educational, and industrial composition of union workers differs from that of nonunion workers.

Taking into account worker differences, economists have concluded that union workers receive on average a 10 to 15 percent wage differential over nonunion workers. The differential ranges from a negligible amount for hotel workers and barbers to 25 to 30 percent higher earnings for skilled construction workers or coal miners. The pattern of results suggests that where unions can effectively monopolize labor supply and control entry, they will be most effective in raising wages. There is some evidence that the impact of unions on wages has declined in recent years.

Overall Impacts. Let us assume that unions can in fact raise the wages of their members above competitive levels. Would this lead to an increase in the average wage of the entire economy? Economists who study this question conclude that the answer is no. They find that unions redistribute income from nonunion labor to union labor. Put differently, if unions succeed in raising their wages above competitive levels, their gains come at the expense of the wages of nonunion workers.

This analysis is supported by empirical evidence showing that the share of national income going to labor has changed little over the last six decades. Once cyclical influences are removed, we can see no appreciable impact of unionization on the share of wages in the United States (see Figure 12-1 on page 231). Moreover, the evidence from heavily unionized European countries suggests that when unions succeed in raising money wage rates, they sometimes trigger an inflationary wage-price spiral with little or no permanent effect upon real wages.

Unions and Classical Unemployment

If unions do not affect overall real wage levels, this suggests that their impact lies primarily upon relative wages. That is, wages in unionized industries would rise relative to those in nonunionized industries. Moreover, employment would tend to be reduced in unionized industries and expanded in nonunionized industries.

When powerful unions raise real wages to artificially high levels, the result is an excess supply of labor that is called *classical unemployment*. This case is also illustrated by Figure 13-6. Assume that unions raise wages above the market-clearing wage at E to a higher real wage at rr . Then, if the supply of and demand for labor in general are unchanged, the arrow between E' and F will represent the number of workers who want to work at wage rr but cannot find work. This is called classical unemployment because it results from real wages that are above competitive levels.

Economists often contrast classical unemployment with the unemployment that occurs in business cycles, often called Keynesian unemployment, which results from insufficient aggregate demand. The effects of too high real wages were seen after the economic unification of Germany in 1990. The economic union fixed East German wages at a level estimated to be at least twice as high as could be justified by labor's marginal revenue product. The result was a sharp decline in employment in eastern Germany after unification.

This analysis suggests that when an economy gets locked into real wages that are too high, high levels of unemployment may result. The unemployment will not respond to the traditional macroeconomic policy of increasing aggregate spending but, rather, will require remedies that lower real wages.



Declining Unionism in the United States

One of the major trends in American labor markets has been the gradual erosion of labor unions since World War II. Whereas unions had organized one-quarter of the labor force in 1955, the fraction has fallen sharply since 1980. The share of unionized workers in manufacturing has shrunk dramatically in the last two decades; only in the public sector are unions still a powerful force.

One of the reasons for the decline in unions is the waning power of the strike, which is the ultimate threat in collective bargaining. In the 1970s U.S. labor unions used that weapon regularly, averaging almost 300 strikes per year. More recently, though, strikes have become relatively uncommon; in fact, they have virtually disappeared from the American labor market. The reason for the decline is that strikes have often backfired on workers. In 1981, the striking air-traffic controllers were all fired by President Reagan. When the professional football players went on strike in 1987, they were forced back to work when the football owners put on the games with replacement players. In 1992, workers striking at Caterpillar Inc., a huge maker of heavy equipment, had to end their 6-month strike when Caterpillar threatened to fill their jobs with permanent replacements. The inability to hurt firms through strikes has led to a significant weakening in the overall power of labor unions in the previous two decades.

You might wonder if the declining power of unions will reduce labor compensation. Economists generally hold that a decline in union power will lower the relative wages of union workers rather than lower the overall share of labor. Look back at Figure 12-1 to examine the share of labor in national income. Can you determine any effect of the declining power of unions after 1980 on labor's share? Most economists believe not.

DISCRIMINATION

Racial, ethnic, and gender discrimination has been a pervasive feature of human societies since the beginning of recorded history. At one extreme, seen before the Civil War in the United States, black slaves were considered property, had virtually no rights, and were often treated harshly. In other times or places, such as in the United States during the segregation period or under apartheid in South Africa until the 1990s, blacks were segregated in housing

and transportation and faced prohibitions against interracial marriage and the most desirable forms of employment. Even today, in an era when discrimination is illegal, subtle forms of informal, premarket, criminal-justice, and statistical discrimination continue to produce disparate outcomes between men and women and particularly among different racial and ethnic groups.

Those who study or experience discrimination know that it extends far beyond the marketplace. Our discussion is limited to economic discrimination, focusing primarily on employment. We want to know why group differences persist decades after discrimination became illegal. We need to understand the sources of the differences between the wages of different groups. Why do African-American and Hispanic citizens in the United States continue to have a measurably lower level of income and wealth than other groups? Why are women excluded from many of the best jobs in business? These are troubling questions that need answers.

ECONOMIC ANALYSIS OF DISCRIMINATION

Definition of Discrimination

When economic differences arise because of irrelevant personal characteristics such as race, gender, sexual orientation, or religion, we call this **discrimination**. Discrimination typically involves either (a) disparate treatment of people on the basis of personal characteristics or (b) practices (such as tests) that have an “adverse impact” on certain groups.

Economists who first began to study discrimination, like the University of Chicago’s Gary Becker, realized that a fundamental puzzle arises: If two groups of workers have equivalent productivity, but one has lower wages, why don’t competitive profit-maximizing firms hire the low-wage workers and increase their profits? For example, suppose that a group of managers in a competitive market decides to pay blue-eyed workers more than equally productive brown-eyed workers. Nondiscriminating firms could enter the market, undercut the costs and prices of the discriminating firms by hiring mainly brown-eyed workers, and drive the discriminating firms out of business. Thus, even if some employers are biased against a group of workers, their bias should not be sufficient to reduce that group’s income. Becker’s

analysis suggests, therefore, that forces other than pure discriminating attitudes are necessary to maintain income disparities between equivalent groups.

Discrimination by Exclusion

The most pervasive form of discrimination is to exclude certain groups from employment or housing. The history of black Americans illustrates how social processes depressed their wages and social status. After slavery was abolished, the black population of the American south fell into a caste system of peonage under “Jim Crow” legislation. Even though legally free and subject to the laws of supply and demand, black workers had earnings far below those of whites. Why? Because they had inferior schooling and were excluded from the best jobs by trade unions, local laws, and customs. They were consequently shunted into menial, low-skilled occupations that were effectively noncompeting groups. Employment segregation allowed discrimination to persist for decades.

Supply and demand can illustrate how exclusion lowers the incomes of groups that are targets of discrimination. Under discrimination, certain jobs are reserved for the privileged group, as is depicted in Figure 13-7(a). In this labor market, the supply of privileged workers is shown by $S_p S_p$, while the demand for such labor is depicted as $D_p D_p$. Equilibrium wages occur at the high level shown at E_p .

Meanwhile, Figure 13-7(b) shows what is happening for minority workers, who, because they live in areas with poor schools and cannot afford private education, do not receive training for the high-paying jobs. With low levels of skills, they take low-skill jobs and have low marginal revenue products, so their wages are depressed to the low-wage equilibrium at E_m .

Note the difference between the two markets. Because minorities are excluded from good jobs, market forces have decreed that they earn much lower wages than the privileged workers. Someone might even argue that minorities “deserve” lower wages because their competitive marginal revenue products are lower. But this rationalization overlooks the root of the wage differential, which is that wage differences arose because certain groups were excluded from the good jobs by their inability to obtain education and training and by the force of custom, law, or collusion.

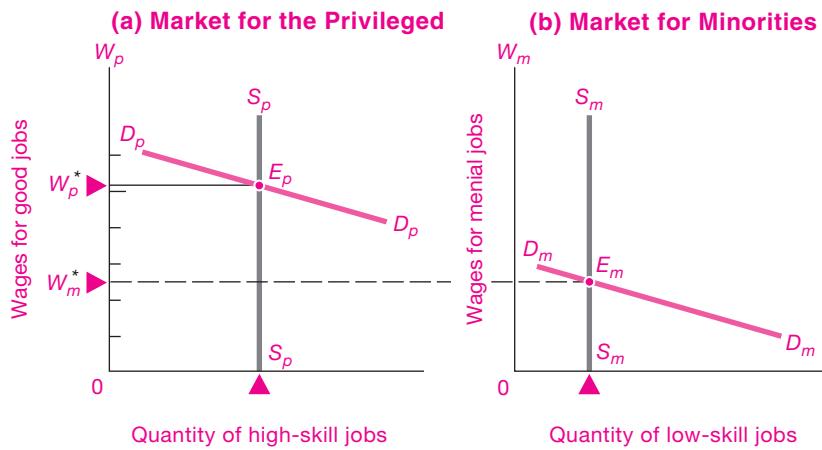


FIGURE 13-7. Discrimination by Exclusion Lowers the Wage Rates of Excluded Minorities

Discrimination is often enforced by excluding certain groups from privileged jobs. If minorities are excluded from good jobs in market (a), they must work in inferior jobs in (b). The privileged group enjoys high wage rates at E_p , while minorities earn low wage rates at E_m in market (b).

Taste for Discrimination

The exclusion example still raises the issue of why some profit-maximizing firms do not evade the laws or customs to undercut their competitors. One explanation proposed by Becker was that either firms or their customers have a “taste for discrimination.” Perhaps some managers do not like hiring black workers; maybe salespersons are prejudiced and don’t want to sell to Hispanic customers. Critics complain that this approach is tautological, in essence saying, “Things are the way they are because people like them that way.”

Statistical Discrimination

One of the most interesting variants of discrimination occurs because of the interplay between incomplete information and perverse incentives. This is known as **statistical discrimination**, in which individuals are treated on the basis of the average behavior of members of the group to which they belong rather than on the basis of their personal characteristics.

One common example arises when an employer screens employees on the basis of their college. The employer may have observed that people who graduate from better schools are *on average* more productive; in addition, grade point averages are often

difficult to compare because of differences in grading standards. Employers therefore often hire people on the basis of their college rather than of their grades. A more careful screening process would show that there are many highly qualified workers from the less well-known schools. We see here a common form of statistical discrimination based on average quality of schooling.

Statistical discrimination leads to economic inefficiencies because it reinforces stereotypes and reduces the incentives of individual members of a group to develop skills and experience. Consider someone who goes to a little-known school. She knows that she will be largely judged by the quality of her schooling credentials. The grade point average, the difficulty of the courses taken, her actual knowledge, and her on-the-job experience may be ignored. The result is that, when subject to statistical discrimination, individuals have greatly reduced incentives to invest in activities that will improve their skills and make them better workers.

Statistical discrimination is particularly pernicious when it involves race, gender, or ethnic groups. If employers treat all black youths as “unproductive” because of average experience with hiring black youths, then gifted individuals not only will

be treated as the average worker but will have little incentive to upgrade their skills.

Statistical discrimination is seen in many areas of society. Life insurance and automobile insurance generally average the risks of people who are careful with those who live dangerously; this tends to reduce the incentive to behave cautiously and leads to a decrease in the average amount of caution in the population. Women were traditionally excluded from quantitatively oriented professions like engineering; as a result, women were more likely to choose humanities and social sciences for their majors and their careers, thereby reinforcing the stereotype that women were uninterested in engineering.

Statistical discrimination not only stereotypes individuals on the basis of group characteristics; it also reduces the incentives of individuals to make investments in education and training and thereby tends to reinforce the original stereotype.

ECONOMIC DISCRIMINATION AGAINST WOMEN

The largest group to suffer from economic discrimination is women. A generation ago, women earned about 70 percent of the wages of men. Part of this was due to differences in education, job experience, and other factors. Today, the gender gap has shrunk sharply. Most of the remaining difference is the “family gap”—a wage penalty against women with children.

What lay behind the income differentials between men and women? The causes are complex, grounded in social customs and expectations, statistical discrimination, and economic factors such as education and work experience. In general, women are not paid less than men for the same job. Rather, the lower pay of women arose because women were excluded from certain high-paying professions, such as engineering, construction, and coal mining. In addition, women tended to interrupt their careers to have children and perform household duties, and this continues to persist in the family gap. Also, economic inequality of the sexes was maintained because, until recently, few women were elected to the boards of directors of large corporations, to senior partnerships in major law firms, or to tenured professorships in top universities.

EMPIRICAL EVIDENCE

Having analyzed the mechanisms by which discrimination is enforced, let us next examine empirical evidence on earnings differentials. On average, women and minorities earn less than do white men. For example, women who worked full-time had earnings equal to 60 percent of men's earnings in 1967. By 2007, that number had risen to 80 percent.

Labor economists emphasize that earnings differentials are not the same as discrimination. Wage differentials often reflect differences in skill and productivity. Many Hispanic workers, particularly immigrants, have historically received less education than have native whites; women customarily spend more time out of the labor force than do men. Since both education and continuing work experience are linked to higher pay, it is not surprising that some earnings differentials exist.

How much of the earnings differentials is due to discrimination rather than productivity differences? Here are some recent findings:

- For women, the extent of discrimination has declined markedly in recent years. Statisticians have uncovered a family gap, which refers to the fact that women who leave the labor force to care for children have an earnings penalty. Aside from the family gap, women appear to have approximately the same earnings as equally qualified men.
- The gap between African-Americans and whites was extremely large for most of American history. However, African-American workers made major progress in the first seven decades of the twentieth century. Data from the 1990s indicate that African-Americans suffer a 5 to 15 percent loss in earnings due to labor market discrimination.
- One of the major encouraging trends is the crumbling of barriers to employment of women and minorities in highly paid professions. In the period from 1950 to 2000, the fraction of women and minorities employed as physicians, engineers, lawyers, and economists has grown sharply. This is particularly striking for women in professional schools. The proportion of women in law schools increased from 4 percent in 1963 to 44 percent in 2006, while for medical schools the proportion rose from 5 percent in 1960 to almost 50 percent in 2006. We see similar trends in other occupations that were once traditionally tied to gender or race.

REDUCING LABOR MARKET DISCRIMINATION

Over the last half-century, government has taken numerous measures to end discriminatory practices. The major steps were legal landmarks, such as the Civil Rights Act of 1964 (which outlaws employment discrimination based on race, color, religion, sex, or national origin) and the Equal Pay Act of 1963 (which requires that employers pay men and women equally for the same work).

Such laws helped dismantle the most blatant discriminatory practices, but more subtle barriers remain. To counter them, more aggressive and controversial policies have been introduced, including measures such as *affirmative action*. This requires that employers show they are taking extra steps to locate and hire underrepresented groups. Studies indicate that this approach has had a positive effect on the hiring and wages of women and minorities. Affirmative action has, however, been widely criticized in recent

years as representing “reverse discrimination,” and some states have banned its use in employment and education.

Uneven Progress

Discrimination is a complex social and economic process. It was enforced by laws that denied disadvantaged groups equal access to jobs, housing, and education. Even after equality under law was established, separation of races and sexes perpetuated social and economic stratification.

The progress in narrowing the earnings gaps among different groups slowed over the last three decades. The disintegration of the traditional nuclear family, cuts in government social programs, harsh drug laws and imprisonment rates, a backlash against many antidiscrimination programs, and the declining relative wages of the unskilled have led to declining living standards for many minority groups. Progress is uneven, and substantial differences in incomes, wealth, and jobs persist.



SUMMARY

A. Fundamentals of Wage Determination

1. The demand for labor, as for any factor of production, is determined by labor's marginal product. Therefore, a country's general wage level tends to be higher when its workers are better trained and educated, when it has more and better capital to work with, and when it uses more advanced production techniques.
2. For a given population, the supply of labor depends on three key factors: population size, average number of hours worked, and labor-force participation. For the United States, immigration has been a major source of new workers in recent years, increasing the proportion of relatively unskilled workers.
3. As wages rise, there are two opposite effects on the supply of labor. The substitution effect tempts each worker to work longer because of the higher pay for each hour of work. The income effect operates in the opposite direction because higher wages mean that workers can now afford more leisure time along with other good things of life. At some critical wage, the supply curve may bend backward. The labor supply of very gifted, unique people is quite inelastic: their wages are largely pure economic rent.

4. Under perfect competition, if all people and jobs were identical, there would be no wage differentials. But once we drop unrealistic assumptions concerning the uniformity of people and jobs, we find substantial wage differentials even in a perfectly competitive labor market. Compensating wage differentials, which compensate for nonmonetary differences in the quality of jobs, explain some of the differentials. Differences in the quality of labor explain many of the other differentials. In addition, the labor market is made up of innumerable categories of noncompeting and partially competing groups.

B. Labor Market Issues and Policies

5. Labor unions occupy an important but diminishing role in the American economy, in terms of both membership and influence. Management and labor representatives meet together in collective bargaining to negotiate a contract. Such agreements typically contain provisions for wages, fringe benefits, and work rules. Unions affect wages by bargaining for standard rates. However, in order to raise real wages above prevailing market-determined levels, unions must prevent entry or competition from nonunion workers.

6. While unions may raise the wages of their members above those of non-union workers, they probably do not increase a country's real wages or labor's share of national income. They are likely to increase unemployment among union members who would prefer to wait for recall from layoff of their high-paid jobs rather than move or take low-paying jobs in other industries. And in a nation with inflexible prices, real wages that are too high may induce classical unemployment.
7. By an accident of history, a tiny minority of white males in the world has enjoyed the greatest affluence. Even more than a century after the abolition of slavery, inequality of opportunity and economic, racial, and gender discrimination continue to lead to loss of income by underprivileged groups.
8. There are many sources of discrimination. One important mechanism is the establishment and maintenance of noncompeting groups. In addition, statistical discrimination occurs when individuals are treated on the basis of the average behavior of members of the group to which they belong. This subtle form of discrimination stereotypes individuals on the basis of group characteristics, reduces the incentives of individuals to engage in self-improvement, and thereby reinforces the original stereotype.
9. Many steps have been taken to reduce labor market discrimination over the last half-century. Early approaches focused on outlawing discriminatory practices, while later steps mandated policies such as affirmative action.

CONCEPTS FOR REVIEW

Wage Determination under Perfect Competition

elements in demand for labor:

labor quality

technology

quality of other inputs

elements in supply of labor:

hours

labor-force participation

immigration

income effect vs. substitution effect
compensating differentials in wages
rent element in wages
segmented markets and non-competing groups

discrimination
earnings differentials: quality differences vs. discrimination
statistical discrimination
antidiscrimination policies

Labor Market Issues

collective bargaining
unions as monopolies
classical unemployment

FURTHER READING AND INTERNET WEBSITES

Further Reading

The elements of the theory of human capital are given in Gary S. Becker, *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, 3rd ed. (University of Chicago Press, 1993).

Labor economics is an active area. Many important topics are covered in advanced surveys, such as Ronald G. Ehrenberg and Robert S. Smith, *Modern Labor Economics: Theory and Public Policy*, 9th ed. (Addison-Wesley, Reading, Mass., 2005).

An excellent overview of the economics of discrimination is contained in the symposium on discrimination in product, credit, and labor markets in *Journal of Economic Perspectives*, Spring 1998.

An important source on the impact of immigration is George Borjas, Richard Freeman, and Lawrence Katz, "How Much Do Immigration and Trade Affect Labor Market Outcomes?" *Brookings Papers on Economic Activity*, vol. 1, 1997, pp. 1-90.

Websites

Analysis of the labor market data for the United States comes from the Bureau of Labor Statistics, at www.bls.gov. This site also has an online version of *The Monthly Labor Review*, which is an excellent source for studies about wages and employment.

An excellent review of trends in labor markets with special reference to new technologies and discrimination is in *Economic Report of the President*, 2000, chap. 4, "Work and Learning in the 21st Century," available online at [w3.access.gpo.gov/eop/](http://www.access.gpo.gov/eop/).

For an international perspective, visit the site of the International Labour Organization at www.ilo.org. If you want a detailed reading list on labor economics, visit the MIT open course website at ocw.mit.edu/OcwWeb/Economics/14-64Spring-2006/Readings/index.htm.

QUESTIONS FOR DISCUSSION

1. What steps could be taken to break down the segmented markets shown in Figure 13-7?
2. Explain, both in words and with a supply-and-demand diagram, the impact of each of the following upon the wages and employment in the affected labor market:
 - a. *Upon union bricklayers:* The bricklayers' union negotiated a lower standard work rule, from 60 bricks per hour to 50 bricks per hour.
 - b. *Upon airline pilots:* After the deregulation of the airlines, nonunion airlines increased their market share by 20 percent.
 - c. *Upon M.D.s:* Many states began to allow nurses to assume more of physicians' responsibilities.
 - d. *Upon American autoworkers:* Japan agreed to limit its exports of automobiles to the United States.
3. Explain what would happen to wage differentials as a result of each of the following:
 - a. An increase in the cost of going to college
 - b. Free migration among the nations of Europe
 - c. Introduction of free public education into a country where education had previously been private and expensive
 - d. Through technological change, a large increase in the number of people reached by popular sports and entertainment programs
4. Discrimination occurs when disadvantaged groups like women or African-Americans are segmented into low-wage markets. Explain how each of the following practices, which prevailed in some cases until recently, helped perpetuate discriminatory labor market segmentation:
 - a. Many state schools would not allow women to major in engineering.
 - b. Many top colleges would not admit women.
 - c. Nonwhites and whites received schooling in separate school systems.
 - d. Elite social clubs would not admit women, African-Americans, or Catholics.
 - e. Employers refused to hire workers who had attended inner-city schools because the average productivities of workers from those schools were low.
5. Recent immigration has increased the number of low-skilled workers with little impact upon the supply of highly trained workers. A recent study by George Borjas, Richard Freeman, and Lawrence Katz estimated that the wages of high school dropouts declined by 4 percent relative to the wages of college graduates in the 1980s as a result of immigration and trade.
 - a. To see the impact of *immigration*, turn back to Figure 12-6 in the previous chapter. Redraw the diagrams, labeling part (a) "Market for Skilled Workers" and part (b) "Market for Unskilled Workers." Then let immigration shift the supply of unskilled labor to the right while leaving the supply of skilled workers unchanged. What would happen to the relative wages of the skilled and unskilled and to the relative levels of employment as a result of immigration?
 - b. Next analyze the impact of *international trade* on wages and employment. Suppose that globalization increases the demand for domestic skilled workers in (a) while reducing the demand for domestic unskilled workers in (b). Show that this would tend to increase the inequality between skilled and unskilled workers.
6. People often worry that high tax rates would reduce the supply of labor. Consider the impact of higher taxes with a backward-bending supply curve as follows: Define the before-tax wage as W , the post-tax wage as W_p , and the tax rate as t . Explain the relationship $W_p = (1 - t)W$. Draw up a table showing the before-tax and post-tax wages when the before-tax wage is \$20 per hour for tax rates of 0, 15, 25, and 40 percent. Now turn to Figure 13-4. For the regions above and below point C, show the impact of a lower tax rate upon the supply curve. In your table, show the relationship between the tax rate and the government's tax revenues.

Land, Natural Resources, and the Environment

14



Land is a good investment: they ain't making it no more.

Will Rogers

If you look at any economic process, you will see that it is powered by a specialized combination of the three fundamental factors of production: land, labor, and capital. In Chapter 1, we learned that land and natural resources provide the footing and fuel for our economy; that durable capital goods and intangibles are produced partners in the production process; and that human labor tills the soil, operates the capital stock, and manages the production processes.

Earlier chapters surveyed both the economic theory of pricing and the marginal productivities of factors, as well as the role of labor in the economy. The present chapter continues the study of the factors of production by looking at the workings of the markets for land, natural resources, and the environment. We will start by looking at the markets for land and natural resources, which are nonproduced factors. We then turn to the vital area of environmental economics. This topic covers an important market failure and some proposed remedies and discusses the topic of global warming.

A. THE ECONOMICS OF NATURAL RESOURCES

When sentient humans first evolved hundreds of thousands of years ago, their economies were based on hunting, fishing, and gathering, with a

rich natural environment but little capital beyond a few sharp sticks and stones. Today, we generally take for granted the bounty of clean air, plentiful water, and unspoiled land. But what is the threat to humanity if we do not respect the limits of our natural environment?

At one pole is an environmentalist philosophy of confines and perils. In this view, human activities threaten to poison our soils, deplete our natural resources, disrupt the intricate web of natural ecosystems, and trigger disastrous climate change. The environmentalist point of view is well expressed in the bleak warning from the distinguished Harvard biologist E. O. Wilson:

Environmentalism . . . sees humanity as a biological species tightly dependent on the natural world. . . . Many of Earth's vital resources are about to be exhausted, its atmospheric chemistry is deteriorating, and human populations have already grown dangerously large. Natural ecosystems, the wellsprings of a healthful environment, are being irreversibly degraded. . . . I am radical enough to take seriously the question heard with increasing frequency: Is humanity suicidal?

Believers in this dismal picture argue that humans must practice “sustainable” economic growth and learn to live within the limitations of our scarce natural resources or we will suffer dire and irreparable consequences.

At the other pole are “cornucopians,” or technological optimists, who believe that we are far from exhausting either natural resources or the capabilities of technology. In this optimistic view, we can look forward to continued economic growth and rising living standards, and human ingenuity can cope with any resource limits or environmental problems. If oil runs out, there is plenty of coal. If that doesn’t pan out, then rising energy prices will induce innovation on solar, wind, and nuclear power. Cornucopians view technology, economic growth, and market forces as the saviors, not the villains. One of the most prominent of the technological optimists was Julian Simon, who wrote:

Ask an average roomful of people if our environment is becoming dirtier or cleaner, and most will say “dirtier.” The irrefutable facts are that the air in the U.S. (and in other rich countries) is safer to breathe now than in decades past. The quantities of pollutants have been declining, especially particulates which are the main pollutant. Concerning water, the proportion of monitoring sites in the U.S. with water of good drinkability has increased since the data began in 1961. Our environment is increasingly healthy, with every prospect that this trend will continue.

Generally, mainstream economists tend to lie between the environmentalist and the cornucopian extremes. They recognize that humans have been drawing upon the earth’s resources for ages. Economists tend to emphasize that *efficient management of the economy requires proper pricing of natural and environmental resources*. In this chapter we will survey the concepts involved in the pricing of scarce natural resources and the management of the environment.

RESOURCE CATEGORIES

What are the important natural resources? They include land, water, and the atmosphere. The land gives us food and wine from fertile soils, as well as oil and other minerals from the earth’s mantle. Our waters give us fish, recreation, and a remarkably efficient medium for transportation. The precious atmosphere yields breathable air, beautiful sunsets, and flying space for airplanes. Natural resources (including land) are a set of factors of production, just like labor and capital. They are factors of production because we derive output or satisfaction from their services.

Economists make two major distinctions in analyzing natural resources. The most important is whether the resource is **appropriable** or **inappropriable**. A commodity is called **appropriable** when firms or consumers can capture its full economic value. Appropriable natural resources include land (whose fertility can be captured by the farmer who sells wheat or wine produced on the land), mineral resources like oil and gas (where the owner can sell the value of the mineral deposit), and trees (where the owner can sell the land or the trees to the highest bidder). In a well-functioning competitive market, appropriable natural resources would be efficiently priced and allocated.

On the other hand, a resource is **inappropriate** when some of the costs and benefits associated with its use do not accrue to its owner. In other words, inappropriate resources are ones involving externalities. (Recall that *externalities* are those activities in which production or consumption imposes uncompensated costs or benefits on other parties.)

Examples of inappropriate resources are found in every corner of the globe. Consider, for instance, the depletion of stocks of many important fish, such as whales, tuna, herring, and sturgeon. A school of tuna can provide not only food for the dinner table but also stock for breeding future generations of tuna. Yet the breeding potential is not reflected in the market price of fish. Consequently, when a fishing boat pulls out a yellowtail tuna, it does not compensate society for the depletion of future breeding potential. This is why unregulated fisheries often tend to be overfished.

This leads to a central result in the economics of natural resources and the environment:

When markets do not capture all the costs and benefits of using natural resources, and externalities are therefore present, markets give the wrong signals and prices are distorted. Markets generally produce too much of goods that generate negative externalities and too little of goods that produce positive externalities.

Techniques used for managing resources depend on whether the resources are renewable or nonrenewable. A **nonrenewable resource** is one whose supply is essentially fixed. Important examples are the fossil fuels, which were laid down millions of years ago and are not renewable on the time scale of

human civilizations, and nonfuel mineral resources, such as copper, silver, gold, stone, and sand.

By contrast, **renewable resources** are ones whose services are replenished regularly. If properly managed, these can yield useful services indefinitely. Solar energy, agricultural land, river water, forests, and fisheries are among the most important categories of renewable resources.

The principles of efficient management of these two classes of resources present quite different challenges. Efficient use of a nonrenewable resource entails the distribution of a finite quantity of the resource over time: Should we use our low-cost natural gas in this generation or save it for the future? By contrast, prudent use of renewable resources involves ensuring that the flow of services is efficiently maintained through, for example, appropriate forest management, protection of fish breeding grounds, and regulation of pollution entering rivers and lakes.

This chapter considers the economics of natural resources. We begin this section by focusing on land. We want to understand the principles underlying the pricing of a fixed resource. In Section B, we turn to the economics of the environment, which involves the important public-policy questions relevant to protecting the quality of our air, water, and land from pollution, as well as global issues such as climate change.

FIXED LAND AND RENTS

The single most valuable natural resource is land. Under law, ownership of “land” consists of a bundle of rights and obligations such as the rights to occupy, to cultivate, to deny access, and to build. Unless you are planning to run your company from a balloon, land is an essential factor of production for any business. The unusual feature of land is that its quantity is fixed and completely unresponsive to price.¹

Rent as Return to Fixed Factors

The price of such a fixed factor is called **rent** or **pure economic rent**. Economists apply the term “rent” not

only to land but also to any other factor that is fixed in supply. If you pay Alex Rodriguez \$30 million per year to play for your baseball team, that money would be considered rent for the use of that unique factor.

Rent is calculated as dollars per unit of the fixed factor per unit of time. The rent on land in the Arizona desert might be \$0.50 per acre per year, while that in midtown New York or Tokyo might be \$1 million per acre per year. Always remember that the word “rent” is used in a special and specific way in economics to denote payments made to factors in fixed supply. Everyday usage of the word often includes other meanings, such as payment for the use of an apartment or building.

Rent (or pure economic rent) is payment for the use of factors of production that are fixed in supply.

Market Equilibrium. The supply curve for land is completely inelastic—that is, vertical—because the supply of land is fixed. In Figure 14-1, the demand

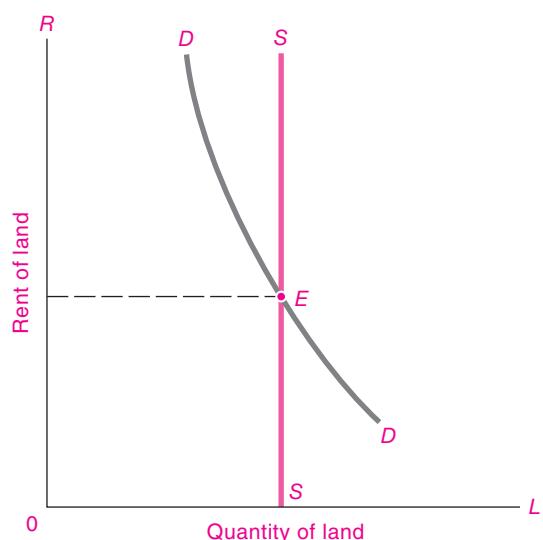


FIGURE 14-1. Fixed Land Must Work for Whatever It Can Earn

Perfectly inelastic supply characterizes the case of rent, sometimes also called pure economic rent. We run up the SS curve to the factor demand curve to determine rent. Aside from land, we can apply rent considerations to gold mines, 7-foot-tall basketball players, and anything else in fixed supply.

¹ This statement must be qualified by the possibility that swamps can be drained and in some cases land can be “produced” by filling shallow bays with landfill. The land area of Boston tripled from 1630 to 1900. Also, land can be used for different purposes, and much agricultural land has been converted to urban land around the world.

and supply curves intersect at the equilibrium point E . It is toward this factor price that the rent of land must tend. Why?

If rent were above the equilibrium, the amount of land demanded by all firms would be less than the fixed supply. Some landowners would be unable to rent their land and would have to offer their land for less and thus bid down its rent. By similar reasoning, the rent could not long remain below the equilibrium. Only at a competitive price where the total amount of land demanded exactly equals the fixed supply will the market be in equilibrium.

Suppose the land can be used only to grow corn. If the demand for corn rises, the demand curve for corn land will shift up and to the right, and the rent will rise. This leads to an important point about land: The price of corn land is high because the price of corn is high. This is a fine example of *derived demand*, which signifies that the demand for the factor is derived from the demand for the product produced by the factor.

Because the supply of land is inelastic, land will always work for whatever it can earn. Thus the value of the land derives entirely from the value of the product, and not vice versa.

Taxing Land

The fact that the supply of land is fixed has a very important consequence. Consider the land market in Figure 14-2. Suppose the government introduces a 50 percent tax on all land rents, taking care to ensure that there is no tax on buildings or improvements.

After the tax, the total demand for the land's services will not have changed. At a price (*including* tax) of \$200 in Figure 14-2, people will continue to demand the entire fixed supply of land. Hence, with land fixed in supply, the market rent on land services (including the tax) will be unchanged and must be at the original market equilibrium at point E .

What will happen to the rent received by the landowners? Demand and quantity supplied are unchanged, so the market price will be unaffected by the tax. Therefore, the tax must be completely paid out of the landowner's income.

The situation can be visualized in Figure 14-2. What the farmer pays and what the landlord receives are now two quite different things. As far as the landlords are concerned, once the government steps in to take its 50 percent share, the effect is just the same

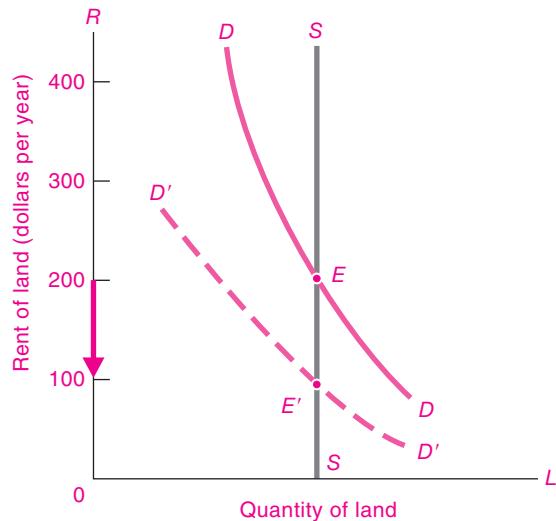


FIGURE 14-2. Tax on Fixed Land Is Shifted Back to Landowners, with Government Skimming Off Pure Economic Rent

A tax on fixed land leaves prices paid by users unchanged at E but reduces rent retained by landowners to E' . This provides the rationale for Henry George's single-tax movement, which aimed to capture for society the increased land values without distorting the allocation of resources.

as it would be if the net demand to the owners had shifted down from DD to $D'D'$. Landowners' equilibrium return after taxes is now only E' . *The entire tax has been shifted backward onto the owners of the factor in perfectly inelastic supply.*

Landowners will surely complain. But under perfect competition there is nothing they can do about it, since they cannot alter the total supply and the land must work for whatever it can get. Half a loaf is better than none.

You might at this point wonder about the effects of such a tax on economic efficiency. The striking result is that *a tax on rent will lead to no distortions or economic inefficiencies*. This surprising result comes because a tax on pure economic rent does not change anyone's economic behavior. Demanders are unaffected because their price is unchanged. The behavior of suppliers is unaffected because the supply of land is fixed and cannot react. Hence, the economy operates after the tax exactly as it did before the tax—with no distortions or inefficiencies arising as a result of the land tax.

A tax on pure economic rent will lead to no distortions or inefficiencies.



Henry George's Single-Tax Movement

The theory of pure economic rent was the basis for the single-tax movement of the late 1800s. At the time, America's population was expanding rapidly as people migrated here from all over the world. With the growth in population and the expansion of railroads into the American West, land rents soared, creating handsome profits for those who were lucky or farsighted enough to buy land early.

Why, some people asked, should landowners be permitted to receive these "unearned land increments"? Henry George (1839–1897), a journalist who thought a great deal about economics, crystallized these sentiments in his best-selling book *Progress and Poverty* (1879). He called for financing government principally through property taxes on land, while cutting or eliminating all other taxes on capital, labor, and the improvements on the land. George believed that such a "single tax" could improve the distribution of income without harming the productivity of the economy.

While the U.S. economy obviously never went very far toward the single-tax ideal, many of George's ideas were picked up by subsequent generations of economists. In the 1920s, the English economist Frank Ramsey extended George's approach by analyzing the efficiency of different kinds of taxes. This led to the development of efficient or Ramsey tax theory. This analysis shows that taxes are least distortionary if levied on sectors whose supplies or demands are highly price-inelastic.

The reasoning behind Ramsey taxes is essentially the same as that shown in Figure 14-2. If a commodity is highly inelastic in supply or demand, a tax on that sector will have very little impact on production and consumption, and the resulting distortion will be relatively small.

B. ENVIRONMENTAL ECONOMICS

In the introductory section of this chapter, we read about some of the controversies surrounding environmental problems. A stern warning from

environmentalists Paul R. Ehrlich and Ann H. Ehrlich in 2008 illustrates these concerns:

Our species has already plucked the low-hanging resource fruit and converted the richest lands to human uses. To support [population growth], metals will have to be won from ever-poorer ores, while oil, natural gas, and water will need to be obtained from ever-deeper wells and transported further. So-called "marginal" lands, often the last strongholds of the biodiversity on which we all depend for essential ecosystem services, increasingly will be converted into yet more crops to feed people, livestock, or SUVs. . . . Climate change is a major threat, even if it may not be the greatest environmental problem. Land-use change, toxification of the planet, increased probability of vast epidemics, or conflicts over scarce resources, involving, possibly, use of nuclear weapons—all population-related—may prove more menacing.

While many technological optimists believe that such concerns are exaggerated, our task is to understand the *economic forces underlying environmental degradation*. This section explores the nature of environmental externalities, describes why they produce economic inefficiencies, and analyzes potential remedies.

EXTERNALITIES

Recall that an *externality* is an activity that imposes involuntary costs or benefits on others, or an activity whose effects are not completely reflected in its market price.

Externalities come in many guises: Some are positive, while others are negative. When a firm dumps toxic wastes into a stream, doing so may kill fish and plants and reduce the stream's recreational value. This is a negative or harmful externality because the firm does not compensate people for the damages imposed on the stream. If you discover a new flu vaccine, the benefits will extend to many people who are not vaccinated because they are less likely to be exposed to the flu. This is a positive or beneficial externality.

Some externalities have pervasive effects, while others have smaller spillover components. When a carrier of bubonic plague entered a town during the Middle Ages, an entire population could be felled by the Black Death. On the other hand, when you eat an onion at a football stadium on a windy day, the external impacts are hardly noticeable.

Public vs. Private Goods

A polar case of an externality is a *public good*, which is a commodity that can be provided to everyone as easily as it can be provided to one person.

The case par excellence of a public good is national defense. Nothing is more vital to a society than its security. But national defense, as an economic good, differs completely from a *private good* like bread. Ten loaves of bread can be divided up in many ways among individuals, and what I eat cannot be eaten by others. But national defense, once provided, affects everyone equally. It matters not at all whether you are hawk or dove, old or young, ignorant or learned—you will receive the same amount of national security from the Army as does every other resident of the country.

Note therefore the stark contrast: The decision to provide a certain level of a public good like national defense will lead to a number of battalions, airplanes, and tanks to protect each of us. By contrast, the decision to consume a private good like bread is an individual act. You can eat four slices, or two, or none; the decision is purely your own and does not commit anyone else to a particular amount of bread consumption.

The example of national defense is a dramatic and extreme case of a public good. But when you think of a smallpox vaccine, the Hubble telescope, clean drinking water, or many similar government projects, you generally find elements of public goods involved. In summary:

Public goods are ones whose benefits are indivisibly spread among the entire community, whether or not individuals desire to consume the public good. **Private goods**, by contrast, are ones that can be divided up and provided separately to different individuals, with no external benefits or costs to others. Efficient provision of public goods often requires government action, while private goods can be efficiently allocated by private markets.



Global Public Goods

Perhaps the thorniest of all market failures are global public goods. These are externalities whose impacts are indivisibly spread across the entire globe. Important examples are actions to slow global warming (considered later in this chapter),

measures to prevent ozone depletion, or discoveries to prevent a global pandemic of avian flu. Global public goods pose particular problems because there are no effective market or political mechanisms available to allocate them efficiently. Markets routinely fail because individuals do not have appropriate incentives to produce these goods, while national governments cannot capture all the benefits of their investments in global public goods.

Why do global public goods differ from other goods? If a terrible storm destroys much of America's corn crop, the price system will guide farmers and consumers to equilibrate needs and availabilities. If America's public road system needs modernization, voters will lobby the government to develop an efficient transportation system. But if problems arise concerning global public goods, such as global warming or antibiotic resistance, neither market participants nor national governments have appropriate incentives to find an efficient outcome. The marginal cost of investments to any individual or nation is much less than the global marginal benefits, and underinvestment is the certain outcome.

MARKET INEFFICIENCY WITH EXTERNALITIES

Abraham Lincoln said that government should “do for the people what needs to be done, but which they cannot, by individual effort, do at all, or do so well, for themselves.” Pollution control satisfies this guideline since the market mechanism does not provide an adequate check on polluters. Firms will not voluntarily restrict emissions of noxious chemicals, nor will they always abstain from dumping toxic wastes into landfills. Pollution control is therefore generally held to be a legitimate government function.

Analysis of Inefficiency

Why do externalities like pollution lead to economic inefficiency? Take a hypothetical coal-burning electric utility. Dirty Light & Power generates an externality by spewing out tons of noxious sulfur dioxide fumes. Some of the sulfur harms the utility, requiring more frequent repainting and raising the firm's medical bills. But most of the damage is “external” to the firm, harming vegetation and buildings and causing various kinds of respiratory ailments and even premature death in people.

Dirty Light & Power must decide how much to reduce its pollution, but it also has to answer to its profit-oriented shareholders. With no pollution cleanup, its workers, plant, and profits will suffer. Cleaning up every last particle, on the other hand, will be very costly. Such a complete cleanup would cost so much that Dirty Light & Power could not hope to survive in the marketplace.

The managers therefore decide to clean up just to the point where profits are maximized. This requires that the benefits to the firm from additional abatement (“marginal private benefits”) be equal to the cost of additional cleanup (“marginal cost of abatement”). Careful economic and engineering calculations might show that the firm’s private interests are maximized when abatement is set at 50 tons. At that level, the marginal private benefits equal the marginal costs of \$10 per ton. Put differently, when Dirty Light & Power produces electricity in a least-cost manner, weighing only private costs and benefits, it will abate only 50 tons and pollute 350 tons.

Suppose, however, that a team of environmental scientists and economists is asked to examine the overall benefits of abatement to society rather than only the benefits to Dirty Light & Power. In examining the total impacts, the auditors find that the *marginal social benefits* of pollution control—including improved health and increased property values in neighboring regions—are 10 times the marginal private benefits. The impact from each extra ton on Dirty Light & Power is \$10, but the rest of society suffers an additional impact of \$90 per ton of external costs. Why doesn’t Dirty Light & Power include the \$90 of additional social benefits in its calculations? The \$90 is excluded because these benefits are external to the firm and have no effect on its profits.

We now see how pollution and other externalities lead to inefficient economic outcomes: In an unregulated environment, firms will determine their most profitable pollution levels by equating the marginal private benefit from abatement with the marginal private cost of abatement. When the pollution spillovers are significant, the private equilibrium will produce inefficiently high levels of pollution and too little cleanup activity.

Socially Efficient Pollution. Given that private decisions on pollution control are inefficient, is there

a better solution? In general, economists look to determine the socially efficient level of pollution by balancing social costs and benefits. More precisely, *efficiency requires that the marginal social benefits from abatement equal the marginal social costs of abatement*.

How might an efficient level of pollution be determined? Economists recommend an approach known as *cost-benefit analysis*, in which efficient emissions are set by balancing the marginal costs of an action against the marginal benefits of that action. In the case of Dirty Light & Power, suppose that experts study the cost data for abatement and environmental damage. They determine that marginal social costs and marginal social benefits are equalized when the amount of abatement is increased from 50 tons to 250 tons. At the efficient pollution rate, they find that the marginal costs of abatement are \$40 per ton, while the marginal social benefits from the last unit removed are also \$40 per ton.

The resulting level of pollution is *socially efficient* because such an emissions rate maximizes the net social value of production. Only at this level of pollution would the marginal social cost of abatement equal the marginal social benefit. Here again, as in many areas, we determine the most efficient outcome by equating the marginal costs and benefits of an activity.

Cost-benefit analysis will show why extreme “no-risk” or “zero-discharge” policies are generally wasteful. Reducing pollution to zero would generally impose astronomically high cleanup costs, while the marginal benefits of reducing the last few grams of pollution may be quite modest. In some cases, it may even be impossible to continue to produce with zero emissions, so a no-risk philosophy might require closing down the computer industry or banning all vehicular traffic. Generally, economic efficiency calls for a compromise, balancing the extra value of the industry’s output against the extra damage from pollution.

An unregulated market economy will generate levels of pollution (or other externalities) at which the marginal private benefit of abatement equals the marginal private cost of abatement. Efficiency requires that the marginal social benefit of abatement equals the marginal social cost of abatement. In an unregulated economy, there will be too little abatement and too much pollution.

Valuing Damages

One of the major difficulties involved in setting efficient environmental policies arises because of the need to estimate the benefits of pollution control and other policies. In cases where pollution affects only marketed goods and services, the measurement is relatively straightforward. If a warmer climate reduces wheat yields, we can measure the damage by the change in the net value of the wheat. Similarly, if a new road requires tearing down someone's house, we can calculate the market value of a replacement dwelling.

Unfortunately, many types of environmental damage are extremely difficult to value. A classic example was the proposal to ban logging across much of the Pacific Northwest in order to preserve the habitat of the spotted owl. That would cost thousands of logging jobs and raise lumber prices. How should we value the benefits in terms of the continued existence of the spotted owl? Or, to take another example, the *Exxon Valdez* oil spill in Prince William Sound, Alaska, damaged beaches and killed wildlife. How much is the life of a sea otter worth?

Economists have developed several approaches for estimating impacts, such as those on owls and otters, that do not show up directly in market prices. The most reliable techniques examine the impact of environmental damage on different activities and then put market-derived values on those activities. For example, in estimating the impact of emissions of sulfur dioxide, environmental economists first estimate the impact of higher emissions on health, and they then place a dollar value on health changes using either survey techniques or estimates that are revealed by people's actual behavior.

Some of the most difficult cases occur in situations that involve ecosystems and the survival of different species. How much should society pay to ensure that the spotted owl survives? Most people will never see a spotted owl, just as they will never see a whooping crane or actually visit Prince William Sound. They may nevertheless place a value on these natural resources. Some environmental economists use a technique called *contingent valuation*, which involves asking people how much they would be willing to pay in a hypothetical situation, say, to keep some natural resource undamaged. This technique will yield answers, but these answers have not always proved to be reliable.

Few would doubt that a healthy and clean environment has a high value, but placing reliable values on the environment, particularly on the nonmarket components, has proved a difficult business.

Graphical Analysis of Pollution

We can illustrate these points with the help of Figure 14-3. The upward-sloping market *MC* curve is the marginal cost of abatement. The downward-sloping curves are the marginal benefits of reducing pollution, with the upper, solid *MSB* line being the marginal social benefit from less pollution while the lower, dashed *MPB* line is the marginal private benefit of abatement to the polluter.



Caution on Graphing Pollution

In analyzing pollution, it is useful to think of pollution control or abatement as a "good."

In the graphs, we therefore measure mar-

ginal costs and benefits on the vertical axis and the abate-
ment or pollution removed on the horizontal axis. The
trick here is to remember that because pollution removal
is a good, it is measured positively on the horizontal axis.
You can also think of pollution as measured negatively from
the far-right point of 400. So abatement of zero is pollution
of 400, while abatement of 400 means zero pollution.

The unregulated market solution comes at point *I*, where the marginal private costs and benefits are equated. At this point, only 50 tons are removed, and the marginal private costs and benefits are \$10 per ton. But the unregulated market solution is inefficient. We can see this by performing an experiment that increases abatement by 10 tons; this is represented by the thin slice to the right of point *I*. For this additional removal, the marginal benefits are given by the total area of the slice under the *MSB* curve, while the marginal costs are given by the area of the slice under the *MC* curve. The net benefits are that part of the slice shown by the shaded area between the two curves.

The efficient level of pollution comes at point *E*, where marginal social benefits are equated to marginal costs of abatement. At that point, both *MSB* and *MC* are equal to \$40 per ton. Also, because *MSB* and *MC* are equal, the experiment of increasing

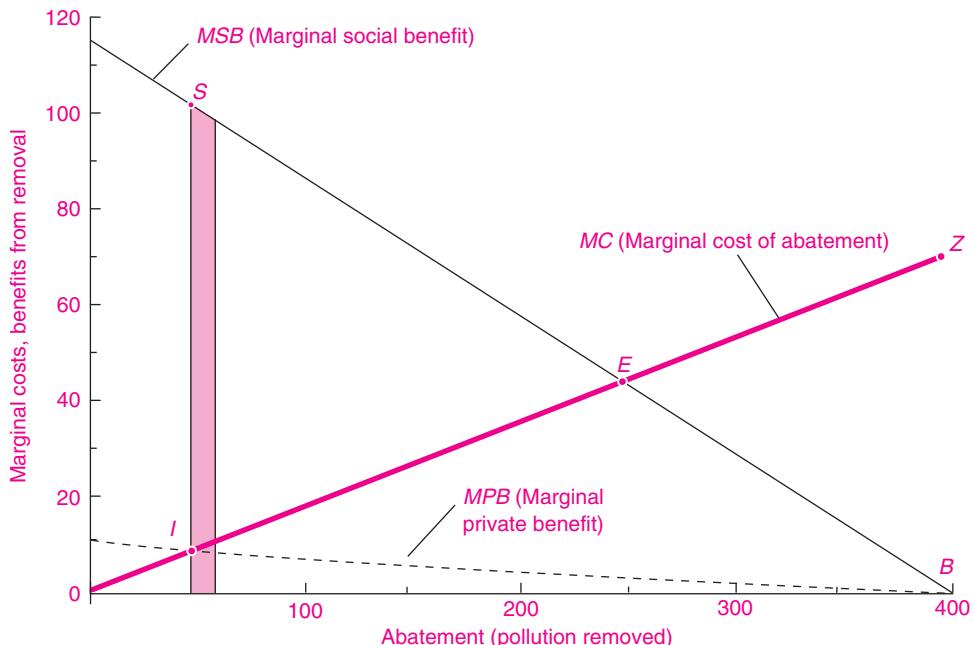


FIGURE 14-3. Inefficiency from Externalities

When marginal social benefit (MSB) diverges from marginal private benefit (MPB), markets will generate unregulated equilibrium at I , with too little abatement or pollution cleanup. Efficient cleanup comes at E , where MSB equals MC .

abatement by a tiny amount will find that there is no difference between the curves, so there is no net benefit from additional pollution control. We can also measure the net benefits of the efficient solution relative to the unregulated market by taking all the little slices of net benefit from the shaded slice to point E . This calculation shows that the area ISE represents the gains from efficient removal of pollutants.

POLICIES TO CORRECT EXTERNALITIES

What are the weapons that can be used to combat inefficiencies arising from externalities? The most visible activities are government antipollution programs that use either direct controls or financial incentives to induce firms to correct externalities. More subtle approaches use enhanced property rights to give the private sector the instruments for negotiating efficient solutions. We survey these approaches in this section.

Government Programs

Direct Controls. For almost all pollution, as well as other health and safety externalities, governments rely on direct regulatory controls; these are often called *social regulations*. For example, the 1970 Clean Air Act reduced allowable emissions of three major pollutants by 90 percent. In 1977, utilities were told to reduce sulfur emissions at new plants by 90 percent. In a series of regulations, firms were told they must phase out ozone-depleting chemicals. And so it goes with regulation.

How does the government enforce a pollution regulation? To continue our example of Dirty Light & Power, the state Department of Environmental Protection might tell Dirty Light & Power to increase its abatement to 250 tons of pollution. Under *command-and-control regulations*, the regulator would simply order the firm to comply, giving detailed instructions on what pollution-control technology to use and where to apply it. There would be little scope for novel approaches or tradeoffs within the

firm or across firms. *If* standards are appropriately set—a very big “if”—the outcome might approach the efficient pollution level described in the previous part of this section.

While it is possible that the regulator might choose a combination of pollution-control edicts that guarantees economic efficiency, in practice that is not very likely. Indeed, much pollution control suffers from extensive inefficiencies. For example, pollution regulations are often set without comparisons of marginal costs and marginal benefits, and without such comparisons there is no way to determine the most efficient level of pollution control.

In addition, standards are inherently a very blunt tool. Efficient pollution reduction requires that the marginal cost of pollution be equalized across all sources of pollution. Command-and-control regulations generally do not allow differentiation across firms, regions, or industries. Hence, regulations are usually the same for large firms and small firms, for cities and rural areas, and for high-polluting and low-polluting industries. Even though firm A might be able to reduce a ton of pollution at a tiny fraction of the cost to firm B, both firms will be required to meet the same standard; nor will there be any incentives for the low-cost firm to reduce pollution beyond the standard even though it would be economical to do just that. Study after study has confirmed that our environmental goals have proved unnecessarily costly when we use command-and-control regulation.

Market Solution: Emissions Fees. In order to avoid some of the pitfalls of direct controls, many economists have suggested that environmental policy rely instead on market-type regulations. One approach is the use of *emissions fees*, which would require that firms pay a tax on their pollution equal to the amount of external damage it causes. If Dirty Light & Power were imposing external marginal costs of \$35 per ton on the surrounding community, the appropriate emissions charge would be \$35 per ton. This is in effect *internalizing* the externality by making the firm pay the social costs of its activities. In calculating its private costs, Dirty Light & Power would find that, at point *E* in Figure 14-3, an additional ton of pollution would cost \$5 of internal costs to the firm plus \$35 in emissions fees, for an overall marginal cost of \$40 per ton of pollution. By equating the

new marginal *private* benefit (private benefit plus emissions fee) with the marginal abatement cost, the firm would set its abatement at the efficient level. *If* the emissions fee were correctly calculated—another big “if”—profit-minded firms would be led as if by a mended invisible hand to the efficient point where marginal social costs and marginal social benefits of pollution are equal.

The alternative approaches are shown graphically in Figure 14-4, which is similar to Figure 14-3. With the direct-control approach, the government instructs the firm to remove 250 tons of pollutants (or to emit no more than 150 tons). This would, in effect, place the standard at the heavy vertical line. If the standard were set at the right level, the firm would undertake the socially efficient level of abatement. Hence, with efficient regulation, the firm will choose point *E*, where *MSB* equals *MC*.

We can also see how emissions fees would operate. Suppose that the government levies a fee of \$35 per ton of pollution. Including the fee, the marginal private benefit of abatement would

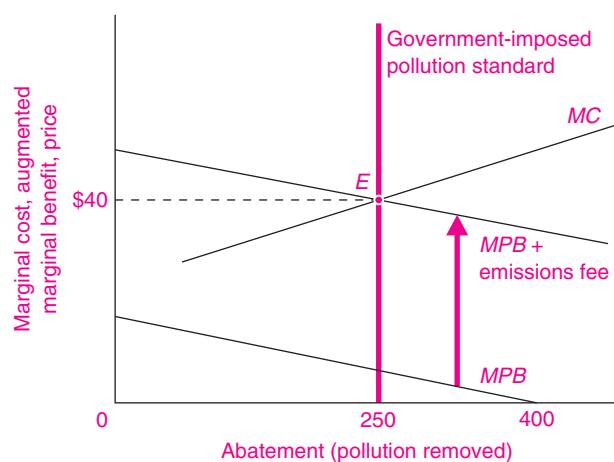


FIGURE 14-4. Pollution Standards and Emissions Fees

When government sets the pollution limitation at 150 tons, or requires removal of 250 tons, this standard will lead to efficient pollution at point *E*.

The same result can be achieved with pollution fees of \$35 per ton. The \$5 *MPB* plus the emissions fee gives a total marginal benefit of \$40 at an abatement of 250 tons. Hence the augmented marginal benefit curve (*MPB* + emissions fee) equals *MC* at the efficient level, *E*.

rise from \$5 to \$40 per ton. We show this as the augmented marginal-private-benefit schedule in Figure 14-4. Faced with the new incentives, the firm would choose efficient point *E* in Figure 14-4.

Market Solution: Tradeable Emissions Permits.

A new approach that does not require the government to legislate taxes is the use of tradeable emissions permits. With this approach, instead of telling firms that they must pay $\$x$ per unit of pollution and then allowing firms to choose the level of pollution, the government chooses the level of pollution and allocates the appropriate number of permits. The price of permits, which represents the level of the emissions fee, is then set by supply and demand in the market for permits. Assuming that firms know their costs of production and abatement, the tradeable-permits approach has the same outcome as the emissions-fee approach.



Economic Innovations: Trading Pollution Permits

Most environmental regulations use a command-and-control approach that limits the emissions from individual sources, such as power plants or automobiles. This approach cannot cap overall emissions. More important, it virtually guarantees that the overall program is inefficient because it does not satisfy the condition that emissions from all sources must have equal marginal costs of abatement.

In 1990, the United States introduced a radical new approach to environmental control in its program on control of sulfur dioxide, which is one of the most harmful environmental pollutants. Under the 1990 Clean Air Act amendments, the government allocates a limited number of pollution permits. The total number of tons permitted for the country has been gradually reduced since 1990. The innovative aspect of the plan is that the permits are freely tradeable. Electric utilities receive pollution permits and are allowed to buy and sell them with each other just like pork bellies or wheat. Those firms which can reduce their sulfur emissions most cheaply do so and sell their permits to pollute; other firms which need additional permits for new plants or have no leeway to reduce emissions find it economical to buy permits rather than install expensive antipollution equipment or shut down.

Environmental economists believe that the enhanced incentives allow the ambitious targets to be met at a much lower cost than would be paid under traditional command-and-control regulation. Studies by economist Tom Tietenberg of Colby College in Maine have determined that the traditional approaches cost 2 to 10 times as much as would cost-effective regulations like emissions trading.

The behavior of this market has produced a big surprise. Originally, the government projected that permits in the early years would sell for around \$300 per ton of sulfur dioxide. But in practice, the market price in the early years fell to below \$100 per ton. One reason for the success was that the program gave strong incentives for firms to innovate, and firms found that low-sulfur coal could be used much more easily and cheaply than had earlier been anticipated. This important experiment has given powerful support to economists who argue for market-based approaches to environmental policy.

Private Approaches

It is generally thought that some form of government intervention in the market is necessary to overcome the market failures associated with pollution and other externalities. In some cases, however, strong property rights can substitute for government regulations or taxes.

One private-sector approach relies upon *liability laws* rather than upon direct government regulations. Under this approach, the legal system makes the generator of externalities legally liable for any damages caused to other persons. In effect, by imposing an appropriate liability system, the externality is internalized.

In some areas, this doctrine is well established. For example, in most states, if you are injured by a negligent driver, you can sue for damages. Or if you are injured or become ill from a defective product, the company can be sued for product liability.

While liability rules are in principle an attractive means of internalizing the nonmarket costs of production, they are quite limited in practice. They usually involve high litigation costs, which add an additional cost to the original externality. In addition, many damages cannot be litigated because of incomplete property rights (such as those involving

clean air) or because of the large number of companies that contribute to the externality (as in the case of chemicals flowing into a stream).

A second private approach relies upon strong property rights and *negotiations among parties*. This approach was developed by the University of Chicago's Ronald Coase, who showed that voluntary negotiations among the affected parties can sometimes lead to an efficient outcome.

For example, suppose that I am a farmer using fertilizers that flow downstream and kill many of the fish in your ponds. Further, suppose that you cannot sue me for killing your fish. If your fish business is sufficiently profitable, you may try to get me to reduce my fertilizer use. In other words, if there is a net profit to be made from reorganizing our joint operations, we have a powerful incentive to get together and agree on the efficient level of fertilizer runoff. Moreover, this incentive would exist without any government antipollution program.

When property rights are well defined and transaction costs are low, particularly when there are few affected parties, strong liability laws or negotiation can sometimes operate to produce an efficient resolution in the presence of externalities.

CLIMATE CHANGE: TO SLOW OR NOT TO SLOW

Of all the environmental issues, none is so worrisome to scientists as the threat of global warming from the greenhouse effect. Climatologists and other scientists warn that the accumulation of gases like carbon dioxide (CO_2), largely produced by the combustion of fossil fuels, is likely to lead to global warming and other significant climatic changes over the next century. On the basis of climate models, scientists project that if current trends continue, the earth may warm 4° to 8° Fahrenheit over the next century. This would take the earth's climate out of the range experienced during the entire period of human civilization.

The greenhouse effect is the granddaddy of public-good problems; actions today will affect the climate for all people in all countries for centuries to come. The costs of reducing CO_2 emissions come in the near term as countries cut back their use of fossil fuels by conserving energy and using alternative

energy sources (solar energy or perhaps nuclear power), plant trees, and take other measures. In the short run, that means we will have to accept more-expensive energy, lower living standards, and lower consumption levels. The benefits of emissions reductions will come many years in the future, when lower emissions reduce future climate-induced damages—with less disruption to agriculture, seacoasts, and ecosystems.

Economists have begun to study the economic impacts of climate change in order to understand how nations might undertake sensible strategies. Economic studies indicate that the market economies in advanced countries like the United States are likely to be relatively insulated from climate change in the coming decades. The major impacts are likely to be in agriculture, forests, and fisheries, along with unmanaged ecosystems such as coral reefs.

An efficient strategy for containing climate change requires weighing the marginal costs of reducing carbon-dioxide (CO_2) emissions against the marginal benefits. Figure 14-5 shows schematically the marginal costs of reductions as MC and the marginal social benefits as MSB . The vertical axis measures costs and benefits in dollars, while the horizontal axis

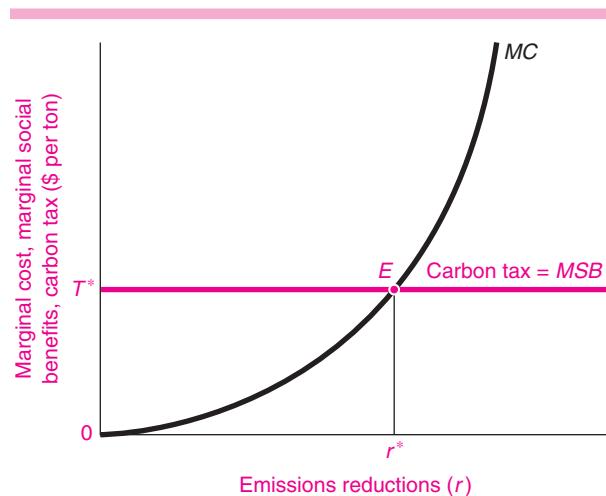


FIGURE 14-5. Carbon Taxes Can Slow Harmful Climate Change

Slowing climate change efficiently requires setting carbon taxes at T^* or limiting carbon dioxide emissions to r^* . Such measures would equate marginal costs of emissions reductions with marginal benefits of reducing damages from climate change.

measures reduction of carbon-dioxide emissions. Point *E* in the graph represents the efficient point at which marginal abatement costs equal marginal social benefits from slowing climate change. This is the point which maximizes the net economic benefits of emissions reductions. By contrast, the pure-market solution would come with emissions reductions at 0, where *MSB* is far above the zero *MC*. An extreme environmentalist solution, which attempts to reduce emissions to zero, would come at the right-hand edge of the graph, where *MC* far exceeds *MSB*.

How can point *E*, the efficient level of CO₂ reduction, be achieved? Since CO₂ emissions come from burning carbon-containing fuels, some have suggested a “carbon tax” on the carbon content of fuels. Fuels which contain more carbon, like coal, would be taxed more heavily than low-carbon fuels such as natural gas. Economists have developed models that estimate efficient paths for carbon taxes—ones that balance the economic costs of higher taxes with the benefits of reduced damages from global warming. These models can serve as a guide to policymakers in the design of policies to combat global warming. Figure 14-5 shows that if the carbon tax is set at the appropriate level, this would induce the efficient level of emissions reductions.



Global Public Goods and the Kyoto Protocol

We discussed the problem of global public goods earlier in this chapter. Nations deal

with global public goods through international agreements like treaties. These are designed to move from an inefficient noncooperative outcome to an efficient cooperative solution to the pollution game. But reaching efficient agreements often proves difficult. Measures to slow global warming provide a useful example. Although scientists have raised alarms about climate change for more than three decades, there were no major international agreements on climate change until the Framework Convention on Climate Change (FCCC) in 1992. The FCCC contained provisions in which high-income countries agreed to non-binding commitments to limit the emissions of greenhouse gases like CO₂.

When voluntary measures were ineffective, countries negotiated the 1997 Kyoto Protocol on climate change. Under the protocol, high-income countries along with

formerly socialist countries agreed to *binding commitments* to reduce by 2010 their total emissions of greenhouse gases by 5 percent (relative to 1990 levels). Each country was allocated a specific target. Based on both economic theory and the experience of the U.S. sulfur dioxide trading program (discussed above), the Kyoto Protocol included a provision for emissions trading among countries. The protocol got off to a shaky start when the Bush administration withdrew the United States from participation in 2001.

Economists have undertaken detailed analyses of the alternative approaches available to tackle the issues involved in global warming. One conclusion of such studies is that it is critical to make sure that market participants face the full costs of their actions. Currently, the climate-change externality is not “internalized” in most countries because CO₂ emissions have a zero price. Without the appropriate price signals, it is unrealistic to think that the millions of firms and billions of consumers will make decisions that reduce the use of carbon fuels. Economic studies also indicate that global participation—not just the participation of high-income countries—is critical to slowing climate change in an economical fashion. By excluding energy-intensive developing countries like China and India from reduction requirements, the cost of meeting the global emissions goal is increased greatly relative to a cost-effective global agreement.

The first round of the Kyoto Protocol went into effect in 2008, but it covers only the period through 2012. Many who are concerned about the future of the globe are looking to see whether the new Obama administration in the United States will join the effort and whether an efficient long-term solution can be designed, implemented, and enforced.

Quarrel and Pollute, or Reason and Compute?

We have seen that many environmentalists are asking gloomy questions about the future of humanity. Having surveyed the field, what should we conclude? Depending on one’s perspective, it is easy to become either optimistic or pessimistic about our ability to understand and cope with the threats to our environment. On the one hand, it is true that we are moving into uncharted waters, depleting many resources while altering others in an irreversible manner, and gambling with our world in more ways than we realize. Humans seem just as quarrelsome today as they were at the dawn of recorded history, and they have

devised weapons that are awesomely effective at avenging their grudges. At the same time, our powers of observation and analysis are also orders of magnitude more formidable.

What will prevail in this race between our tendencies to quarrel and pollute and our powers to reason and compute? Are there enough resources to allow the poor to enjoy the consumption standards of today's high-income countries, or will today's rich

pull the ladder up behind them? There are no final answers to these deep questions. But economists believe that one central answer is to employ market mechanisms to provide incentives to reduce pollution and other harmful side effects of economic growth. Wise decisions along with appropriate incentives will help to ensure that *Homo sapiens* can not only survive but also thrive for a long time to come.



SUMMARY

A. The Economics of Natural Resources

1. Natural resources are nonrenewable when they cannot regenerate quickly and are therefore essentially fixed in supply. Resources are renewable when their services are replenished regularly and they can, if properly managed, yield useful services indefinitely.
2. Natural resources are appropriable when firms or consumers can capture the full benefits of their services; examples include vineyards and oil fields. Natural resources are inappropriable when their total costs or benefits do not accrue to the owners; in other words, they involve externalities.
3. The return to fixed factors like land is called pure economic rent, or rent, for short. Since the supply curve for land is vertical and completely inelastic, the rent will be price-determined rather than price-determining.
4. A factor like land that is inelastically supplied will continue to work the same amount even though its factor reward is reduced. For this reason, Henry George pointed out that rent is in the nature of a surplus rather than a reward necessary to coax out the factor's effort. This provides the basis for his single-tax proposal to tax the unearned increment of land value, which raises tax revenues without raising prices to consumers or distorting production. Modern tax theory extends George's insight by showing that inefficiencies are minimized by taxing goods that are relatively inelastic in supply or demand because such taxes lead to relatively small distortions in behavior.

B. Environmental Economics

5. Environmental problems arise because of externalities that stem from production or consumption. An externality is an activity that imposes involuntary costs or benefits on others and whose effects are not completely reflected in market prices.

6. The polar extreme of an externality is the case of public goods, like national defense, where all consumers in a group share equally in the consumption of the good and cannot be excluded. Public health, inventions, parks, and dams also possess public-good properties. These contrast with private goods, like bread, which can be divided and provided to a single individual.
7. An unregulated market economy will produce too much pollution and too little abatement. Unregulated firms decide on abatement and other public goods by comparing the marginal *private* benefits of abatement with the marginal costs. But efficiency requires that the marginal *social* benefits of abatement equal the marginal costs.
8. Economists emphasize that the efficient management of externalities requires the proper pricing of natural and environmental resources. This involves ensuring that market participants face the full social costs of their activities.
9. There are different approaches through which governments can take steps to internalize or correct the inefficiencies arising from externalities. Alternatives include decentralized or private solutions (such as negotiations or legal liability rules) and government-imposed approaches (such as pollution-emission standards or emissions taxes). Experience indicates that no single approach will fit every circumstance, but many economists believe that greater use of market-oriented approaches would improve the efficiency of regulatory systems.
10. Global public goods present the thorniest problems because they cannot easily be solved by either markets or national governments. Nations must work together to devise new tools to forge international agreements when issues such as global warming threaten our ecosystem and our standards of living.

CONCEPTS FOR REVIEW

Land and Natural Resources

renewable vs. nonrenewable resources
appropriable vs. inappropriable resources
rent, pure economic rent
inelastic supply of land
taxation of fixed factors

Environmental Economics

externalities and public goods
private vs. public goods
inefficiency of externalities
internal vs. external costs, social vs. private benefits

remedies for externalities: standards, taxes, liability, bargaining
tradeable emissions permits
global public goods

FURTHER READING AND INTERNET WEBSITES

Further Reading

Environmental economics is a rapidly growing field. You can explore advanced topics in a textbook such as Thomas H. Tietenberg, *Environmental Economics and Policy*, 7th ed. (Addison-Wesley, New York, 2006). An excellent book of readings is Robert Stavins, ed., *Economics of the Environment: Selected Readings*, 5th ed. (Norton, New York, 2005).

The quote from Wilson is from Edward O. Wilson, “Is Humanity Suicidal?” *New York Times Magazine*, May 30, 1993, p. 27. The quotation from Julian Simon is from *Scarcity or Abundance? A Debate on the Environment* (Norton, New York, 1994), available at www.juliansimon.com/writings/Norton/NORTON01.txt. The quotation from Ehrlich and Ehrlich is from *The New York Review of Books*, February 14, 2008.

Websites

One of the best websites on resources and the environment is maintained by the nonprofit organization Resources for the Future at www.rff.org. You can consult this site for information on a wide range of issues.

Energy data are available at the Energy Information Agency’s comprehensive site at www.eia.doe.gov.

You can learn more about environmental policy at the U.S. Environmental Protection Agency’s website at www.epa.gov. International environmental policy is found at the United Nations Environmental Program’s site at www.unep.org. Information on the Kyoto Protocol and other programs to address climate change can be found at www.ipcc.ch and www.unfccc.de.

QUESTIONS FOR DISCUSSION

1. What is the difference between renewable and nonrenewable resources? Give examples of each.
2. What is meant by an inappropriable natural resource? Provide an example and explain why the market allocation of this resource is inefficient. What would be your preferred way to improve the market outcome?
3. Define “pure economic rent.”
 - a. Show that an increase in the supply of a rent-earning factor will depress its rent and lower the prices of the goods that use it.
 - b. Explain the following statement from rent theory: “It is not true that the price of corn is high because the price of corn land is high. Rather, the reverse is closer to the truth: the price of corn land is high because the price of corn is high.” Illustrate with a diagram.
4. Assume that the supply curve for top baseball players is perfectly inelastic with respect to their salaries.
 - a. Explain what completely inelastic supply means in terms of number of games played.
 - b. Next assume that because of television, the demand for the services of major-league baseball

- players increases. What would happen to their salaries? What would happen to their batting averages (other things held constant)? Does this theory fit historical trends?
5. Explain why a tax on land rent is efficient. Compare a tax on the land with a tax on the houses on the land.
6. “Local public goods” are ones that mainly benefit the residents of a specific town or state—such as beaches or schools open only to town residents. Is there any reason to think that towns might act competitively to provide the correct amount of local public goods to their residents? If so, does this suggest an economic theory of “fiscal federalism” whereby local public goods should be locally supplied?
7. Decide whether each of the following externalities is serious enough to warrant collective action. If so, which of the four remedies considered in this chapter would be most efficient?
- Steel mills emitting sulfur oxides into the Birmingham air
 - Smoking by people in restaurants
 - Smoking by students without roommates in their own rooms
 - Driving by persons under the influence of alcohol
 - Driving by persons under age 21 under the influence of alcohol
8. Get your classmates together to do a contingent-valuation analysis on the value of the following: Prohibiting drilling in all wilderness areas in the United States; preventing the extinction of spotted owls for another 10,000 years; ensuring that there are at least 1 million spotted owls in existence for another 10,000 years; reducing the chance of dying in an automobile accident from 1 in 1000 to 1 in 2000 per year. How reliable do you think this technique is for gathering information about people’s preferences?
9. Don Fullerton and Robert Stavins argue that the following are myths about how economists think about

the environment (see Chapter 1 in the Stavins book in the Further Reading section). For each, explain why it is a myth and what the correct approach is:

- Economists believe that the market solves all environmental problems.
 - Economists always recommend market solutions to environmental problems.
 - Economists always use market prices to evaluate environmental issues.
 - Economists are concerned only with efficiency and never with income distribution.
10. **Advanced problem:** Global public goods pose special problems because no single nation can capture all the benefits of its own pollution-control efforts. To see this, redraw Figure 14-5, labeling it “Emissions Reduction for the United States.” Label all the curves with “US” to indicate that they refer to costs and benefits for the United States alone. Next, draw a new *MSB* curve which is 3 times higher than the MSB_{US} at every point to indicate that the benefits to the world are 3 times higher than those to the United States alone. Consider the “nationalistic” equilibrium at *E* where the United States maximizes its own net benefits from abatement. Can you see why this is inefficient from the point of view of the entire globe? (*Hint:* The reasoning is analogous to that in Figure 14-3.)
- Consider this issue from the point of view of game theory. The Nash equilibrium would occur when each country chose the nationalistic equilibrium you have just analyzed. Describe why this is analogous to the inefficient Nash equilibrium described in Chapter 10—only here the players are nations rather than firms. Now consider the cooperative game in which nations get together to find the efficient equilibrium. Describe the efficient equilibrium in terms of global *MC* and *MSB* curves. Can you see why the efficient equilibrium would require a uniform carbon tax in each country?

Capital, Interest, and Profits

15



You can have your cake and eat it too: Lend it out at interest.

Anonymous

The United States is a “capitalist” economy. By this we mean that most of the country’s capital and other assets are privately owned. In 2008, the net stock of capital in the United States was more than \$150,000 per capita, of which 67 percent was owned by private corporations, 14 percent by individuals, and 19 percent by governments. Moreover, the ownership of the nation’s wealth was highly concentrated in the portfolios of the richest Americans. Under capitalism, individuals and private firms do most of the saving, own most of the wealth, and get most of the profits on these investments.

This chapter is devoted to the study of capital. We begin with a discussion of the basic concepts in capital theory. These include the notion of “roundaboutness” and different measures of the rate of return on investment. Then we will turn to the crucial questions of the supply and demand for capital. This overview will give us a much deeper understanding of some of the key features of a private market economy.

A. BASIC CONCEPTS OF INTEREST AND CAPITAL

What Is Capital?

We begin with a brief summary of the important concepts of capital and finance developed in this chapter. **Capital** consists of those durable produced

items that are in turn used as productive inputs for further production. Some capital might last for only a few years, while others might last for a century or more. But the essential property of capital is that it is both an input and an output.

In an earlier era, capital consisted primarily of tangible assets. Three important categories of tangible capital are structures (such as factories and homes), equipment (such as consumer durable goods like automobiles and producer durable equipment like machine tools and trucks), and inventories (such as cars in dealers’ lots).

Today, intangible capital is increasingly important. Examples include software (such as computer operating systems), patents (such as the ones on microprocessors), and brand names (such as Coca-Cola). Robert Hall of Stanford calls this “e-capital” to distinguish between traditional tangible capital and increasingly important intellectual capital.

Prices and Rentals on Investments

Capital is bought and sold in capital markets. For example, Boeing sells aircraft to airlines; the airlines then use these specialized capital goods along with software, skilled labor, land, and other inputs to produce and sell air travel.

Most capital is owned by the firms that use it. Some capital, however, is rented out by its owners.

Payments for the temporary use of capital goods are called rentals. An apartment that is owned by Ms. Landlord might be rented out for a year to a student, and the monthly payment of \$800 per month would constitute a rental. We distinguish *rent* on fixed factors like land from *rentals* on durable factors like capital.

Capital vs. Financial Assets

Individuals and businesses own a mix of different kinds of assets. One class is the productive input capital that we just discussed—items like computers, automobiles, and houses that are used to produce other goods and services. But we must distinguish these tangible assets from *financial assets*, which are essentially pieces of paper or electronic records. More precisely, financial assets are monetary claims by one party against another party. An important example is a mortgage, which is a claim against a homeowner for monthly payments of interest and principal; these payments will repay the original loan that helped finance the purchase of the house.

Often, as in the case of a mortgage, a tangible asset will lie behind (or serve as collateral for) a financial asset. In other cases, such as student loans, a financial asset may derive its value from a promise to pay based on the future earning power of an individual.

It is clear that tangible assets are an essential part of an economy because they increase the productivity of other factors. But what function do financial assets serve? These assets are crucial because of the mismatch between savers and investors. Students need money to pay for college, but they do not currently have the earnings or the savings necessary to pay the bills. Older people, who are working and saving for retirement, may have income in excess of their expenditures and can provide the savings. A vast financial system of banks, mutual funds, insurance companies, and pension funds—often supplemented by government loans and guarantees—serves to channel the funds of those who are saving to those who are investing. Without this financial system, it would not be possible for firms to make the huge investments needed to develop new products, for people to buy houses before they had saved the entire housing price, or for students to go to college without first saving the large sums necessary.

The Rate of Return on Investments

Suppose that you own some capital and rent it out or that you have some cash and lend it to a bank or to a small business. Or perhaps you want to take out a mortgage to buy a house. You will naturally want to know what you will pay to borrow or how much you will earn by lending. This amount is called the **rate of return on investments**. In the special case of the return on fixed-interest financial assets, these earnings are called the **interest rate**. From an economic point of view, interest rates or returns on investments are the price of borrowing or lending money. The returns will vary greatly depending upon the maturity, risk, tax status, and other attributes of the investment.

We will devote considerable space in this chapter to understanding these concepts. The following summary highlights the major ideas:

1. Capital consists of durable produced items that are in turn used as productive inputs for the production of other goods. Capital consists of both tangible and intangible assets.
2. Capital is bought and sold in capital markets. Payments for the temporary use of capital goods are called rentals.
3. We must distinguish financial assets, which are essentially pieces of paper deriving their value from ownership of other tangible or intangible assets.
4. The rate of return on investments, and the special case of the interest rate, is the price for borrowing and lending funds. We usually calculate rates of return on the funds using units of percent per year.

RATES OF RETURN AND INTEREST RATES

We now examine in greater detail the major concepts in capital and financial theory. We begin with the definition of a rate of return on investments, which is the most general concept. We then apply these definitions to financial assets.

Rate of Return on Capital

One of the most important tasks of any economy is to allocate its capital across different possible investments. Should a country devote its investment resources to heavy manufacturing like steel or to

information technologies like the Internet? Should Intel build a \$4 billion factory to produce the next generation of microprocessors? These questions involve costly investments—laying out money today to obtain a return in the future.

In deciding upon the best investment, we need a measure for the yield or return. One important measure is the **rate of return on investment**, which denotes the net dollar return per year for every dollar of invested capital.

Let's consider the example of a rental car company. Ugly Duckling Rental Company buys a used car for \$20,000 and rents it out. After subtracting all expenses (revenues less expenses such as wages, office supplies, and energy costs) and assuming no change in the car's price, Ugly Duckling earns a net rental of \$2400 each year. The rate of return is 12 percent per year ($12\% = \$2400/\$20,000$). Note that the rate of return is a pure or unitless number per unit of time. That is, the rate of return has the dimensions of (dollars per period)/(dollars), and it is usually calculated with units of percent per year.

These concepts are useful for comparing investments. Suppose you are considering investments in rental cars, oil wells, apartments, education, and so forth. How can you decide which investment to make?

One useful approach is to compare the rates of return on the different investments. For each possibility, calculate the dollar cost of the capital good. Then estimate the net annual dollar receipts or rentals yielded by the asset. The ratio of the annual net rental to the dollar cost is the rate of return on investment, which tells you how much money you get back for every dollar invested, measured as dollars per year per dollar of investment or percent per year.

The rate of return on investment is the annual net return (rentals less expenses) per dollar of invested capital. It is a pure or unitless number—percent per year.

Of Wine, Trees, and Drills. Here are some examples of rates of return on investments:

- I buy a plot of land for \$100,000 and sell it a year later for \$110,000. If there are no other expenses, the rate of return on this investment is \$10,000 per year/\$100,000, or 10 percent per year.

- I plant a pine tree with a labor cost of \$100. At the end of 25 years, the grown tree sells for \$430. The rate of return on this capital project is then 330 percent per quarter-century, which, as a calculator will show you, is equivalent to a return of 6 percent per year. That is, $\$100 \times (1.06)^{25} = \430 .
- I buy a \$20,000 piece of oil-drilling equipment. For 10 years it earns annual rentals of \$30,000, but I also incur annual expenses of \$26,000 for fuel, insurance, and maintenance. The \$4000 net return covers interest and repays the principal of \$20,000 over 10 years. What is the rate of return here? Statistical tables show that the rate of return is 15 percent per year.

Financial Assets and Interest Rates

For the case of financial assets, we use a different set of terms when measuring the rate of return. When you buy a bond or put money in your savings account, the financial yield on this investment is called the *interest rate*. For example, if you bought a 1-year bond in 2008, you would have earned a yield of around 3 percent per year. This means that if you bought a \$1000 bond on January 1, 2008, you would have \$1030 on January 1, 2009.

You will usually see interest rates quoted in percent per year. This is the interest that would be paid if the sum were borrowed (or loaned) for an entire year; for shorter or longer periods, the interest payment is adjusted accordingly.

THE PRESENT VALUE OF ASSETS

Most assets will produce a stream of rentals or receipts over time. If you own an apartment building, for example, you will collect rental payments over the life of the building, much as the owner of a fruit orchard will pick fruit from the trees each year.

Suppose you become weary of tending the building and decide to sell it. To set a fair price for the building, you would need to determine the value today of the entire stream of future income. The value of that stream is called the present value of the capital asset.

The present value is the dollar value today of a stream of future income. It is measured by calculating how much money invested today would be needed, at the going interest rate, to generate the asset's future stream of receipts.

Let's start with a very simple example. Say that someone offers to sell you a bottle of wine that matures in exactly 1 year and can then be sold for exactly \$11. Assuming the market interest rate is 10 percent per year, what is the present value of the wine—that is, how much should you pay for the wine today? Pay exactly \$10, because \$10 invested today at the market interest rate of 10 percent will be worth \$11 in 1 year. So the present value of next year's \$11 wine is today \$10.

Present Value for Perpetuities

We discuss the first way of calculating present value by examining the case of a *perpetuity*, which is an asset like land that lasts forever and pays $\$N$ each year from now to eternity. We are seeking the present value (V) if the interest rate is i percent per year, where the present value is the amount of money invested today that would yield exactly $\$N$ each year. This is simply

$$V = \frac{\$N}{i}$$

where V = present value of the land (\$)
 $\$N$ = perpetual annual receipts (\$ per year)
 i = interest rate in decimal terms (e.g., 0.05, or $\frac{5}{100}$ per year)

This says that if the interest rate is always 5 percent per year, an asset yielding a constant stream of income will sell for exactly 20 ($= 1 \div \frac{5}{100}$) times its annual income. In this case, what would be the present value of a perpetuity yielding \$100 every year? At a 5 percent interest rate its present value would be \$2000 ($= \$100 \div 0.05$).

The formula for perpetuities can also be used to value stocks. Suppose that a share of Spring Water Co. is expected to pay a dividend of \$1 every year into the indefinite future and that the discount rate on stocks is 5 percent per year. Then the stock price should be $P = \$1/0.05 = \20 per share. (These numbers are corrected for inflation, so the numerator is “real dividends” and the denominator is a “real interest rate” or a “real discount rate,” defined below).

General Formula for Present Value

Having seen the simple case of the perpetuity, we move to the general case of the present value of an asset with an income stream that varies over time.

The main thing to remember about present value is that future payments are worth less than current payments and they are therefore *discounted* relative to the present. Future payments are worth less than current payments just as distant objects look smaller than nearby ones. The interest rate produces a similar shrinking of time perspective.

Let's take a fantastic example.¹ Say that someone proposes to pay \$100 million to your heirs in 100 years. How much should you pay for this today? According to the general rule for present value, to figure out the value today of $\$P$ payable t years from now, ask yourself how much must be invested today to grow into $\$P$ at the end of t years. Say the interest rate is 6 percent per annum. Applying this each year to the growing amount, a principal amount of $\$V$ grows in t years to $\$V \times (1 + 0.06)^t$. Hence, we need only invert this expression to find present value: the present value of $\$P$ payable t years from now is today $\$P/(1 + 0.06)^t$. Using this formula, we determine that the present value of \$100 million paid in 100 years is \$294,723.

In most cases, there are several terms in an asset's stream of income. In present-value calculations, each dollar must stand on its own feet. First, evaluate the present value of each part of the stream of future receipts, giving due allowance for the discounting required by its payment date. Then simply add together all these separate present values. This summation will give you the asset's present value.

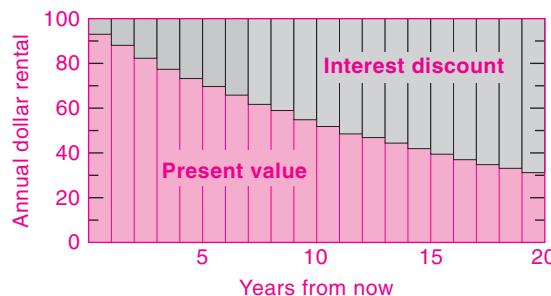
The exact formula for present value (V) is the following:

$$V = \frac{N_1}{1 + i} + \frac{N_2}{(1 + i)^2} + \cdots + \frac{N_t}{(1 + i)^t} + \cdots$$

In this equation, i is the one-period market interest rate (assumed constant). Further, N_1 is the net receipts (positive or negative) in period 1, N_2 the net receipts in period 2, N_t the net receipts in period t , and so forth. Then the stream of payments ($N_1, N_2, \dots, N_t, \dots$) will have the present value, V , given by the formula.

For example, assume that the interest rate is 10 percent per year and that I am to receive \$1100

¹ Question 9 at the end of this chapter asks about the real life example of the present value of the real estate of Manhattan when it was purchased by the Dutch.

**FIGURE 15-1. Present Value of an Asset**

The lower, green area shows the present value of a machine giving net annual rentals of \$100 for 20 years with an interest rate of 6 percent per year. The upper, blue area has been discounted away. Explain why raising the interest rate increases the blue area and therefore depresses the market price of an asset.

next year and \$2662 in 3 years. The present value of this stream is

$$V = \frac{1100}{(1.10)^1} + \frac{2662}{(1.10)^3} = 3000$$

Figure 15-1 shows graphically the calculation of present value for a machine that earns steady net annual rentals of \$100 over a 20-year period and has no scrap value at the end. Its present value is not \$2000 but only \$1157. Note how much the later dollar earnings are scaled down or discounted because of our time perspective. The total area remaining after discounting (the blue shaded area) represents the machine's total present value—the value today of the stream of all future incomes.

Acting to Maximize Present Value

The present-value formula tells us how to calculate the value of any asset once we know the future earnings. But note that an asset's future receipts usually depend on business decisions: Shall we use a truck 8 or 9 years? Overhaul it once a month or once a year? Replace it with a cheap, nondurable truck or an expensive, durable one?

There is one rule that gives correct answers to all investment decisions: Calculate the present value resulting from each possible decision. Then always act so as to maximize present value. In this way you will have more wealth to spend whenever and however you like.

Interest Rates and Asset Prices

When interest rates rise, many asset prices fall. For example, if the Federal Reserve unexpectedly tightens monetary policy and raises interest rates, you will generally read that bond and stock prices fall. We can understand the reason for this pattern using the concept of present value.

Our previous discussion showed that the present value of an asset will depend on both the stream of future returns and the interest rate. As interest rates change, so will the present value and therefore the market value of an asset. Here are some examples:

- Begin with a 1-year bond and an initial interest rate of 5 percent per year. If the bond returns \$1000 one year from now, then its current present value is $\$1000/1.05 = \952.38 . Now suppose that the interest rate rises to 10 percent per year. Then the present value of the bond would be only $\$1000/1.1 = \909.09 . The price of the asset declined as the interest rate increased.
- Take the case of a perpetuity that yields \$100 per year. At an interest rate of 5 percent per year, the perpetuity has a present value of $\$100/0.05 = \2000 . Now if the interest rate rises to 10 percent per year, the value falls by half to only \$1000.

We can now see that asset prices tend to move inversely with interest rates because their present value decreases as the interest rate increases. Note as well that the prices of longer-term assets tend to change more than do the prices of shorter-term assets. This occurs because more of the return is in the future, and the prices of long-term assets are therefore affected more by the changing interest rate.

The dependence of asset prices on interest rates is a general property of financial assets. The prices of stocks, bonds, real estate, and many other long-lived assets will decline as interest rates rise.

THE MYSTERIOUS WORLD OF INTEREST RATES

Textbooks often speak of “*the* interest rate” as if there were only one, but in fact today’s complex financial system has a vast array of interest rates. If you look at *The Wall Street Journal*, you will see page after page of financial interest rates. Interest rates depend mainly on the characteristics of the loan or of the borrower. Let us review the major differences.

Loans differ in their *term* or *maturity*—the length of time until they must be paid off. The shortest loans are overnight. Short-term securities are for periods up to a year. Companies often issue bonds that have maturities of 10 to 30 years, and mortgages are up to 30 years in maturity. Longer-term securities generally command a higher interest rate than do short-term ones because lenders are willing to sacrifice quick access to their funds only if they can increase their yield.

Loans also vary in terms of *risk*. Some loans are virtually riskless, while others are highly speculative. Investors require that a premium be paid when they invest in risky ventures. The safest assets in the world are the securities of the U.S. government. These bonds are backed by the full faith, credit, and taxing powers of the government. Intermediate in risk are borrowings of creditworthy corporations, states, and localities. Risky investments, which bear a significant chance of default or nonpayment, include those of companies close to bankruptcy, cities with shrinking tax bases, or countries like Argentina with large overseas debts and unstable political systems.

The U.S. government pays what is called the “riskless” interest rate; over the last two decades this has ranged from 0 to 15 percent per year for short-term bonds. Riskier securities might pay 1, 2, or even 10 percent per year more than the riskless rate; this premium reflects the amount necessary to compensate the lender for losses in case of default.

Assets vary in their liquidity. An asset is said to be *liquid* if it can be converted into cash quickly and with little loss in value. Most marketable securities, including common stocks and corporate and government bonds, can be turned into cash quickly for close to their current value. Illiquid assets include unique assets for which no well-established market exists. For example, if you own the only Victorian mansion in a small town, you might find it difficult to sell the asset quickly or at a price near its realistic market value—your house is an illiquid asset. Because of the higher risk and the difficulty of realizing the asset values quickly, illiquid assets or loans require higher interest rates than do liquid, riskless ones.

When these three factors (along with other considerations such as tax status and administrative costs) are considered, it is not surprising that we see so many different financial assets and so many different interest rates. Figure 15-2 and Table 15-1 show the behavior of a few important interest rates over the last five decades. In the discussion that follows,

when we speak of “the interest rate,” we are generally referring to the interest rate on short-term government securities, such as the 90-day Treasury-bill rate. As Figure 15-2 shows, most other interest rates rise and fall in step with short-term interest rates.

Real vs. Nominal Interest Rates

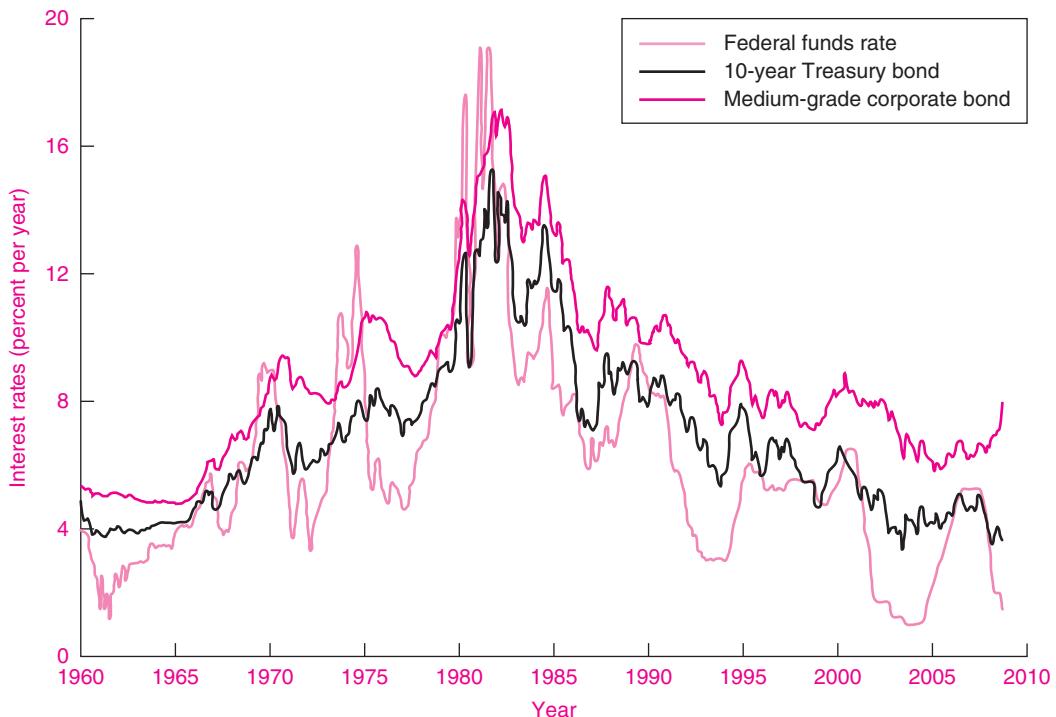
Interest is paid in dollar terms, not in terms of houses or cars or goods in general. The *nominal interest rate* measures the yield in dollars per year per dollar invested. But dollars can become distorted yardsticks. The prices of houses, cars, and goods in general change from year to year—these days prices generally rise due to inflation. Put differently, the interest rate on dollars does not measure what a lender really earns in terms of goods and services. Let us say that you lend \$100 today at 5 percent-per-year interest. You would get back \$105 at the end of a year. But because prices changed over the year, you would not be able to obtain the same quantity of goods that you could have bought at the beginning of the year if you had \$105.

Clearly, we need another concept that measures the return on investments in terms of real goods and services rather than the return in terms of dollars. This alternative concept is the *real interest rate*, which measures the quantity of goods we get tomorrow for goods forgone today. The real interest rate is obtained by correcting nominal or dollar interest rates for the rate of inflation.

The **nominal interest rate** (sometimes also called the *money interest rate*) is the interest rate on money in terms of money. When you read about interest rates in the newspaper, or examine the interest rates in Figure 15-2, you are looking at nominal interest rates; they give the dollar return per dollar of investment.

In contrast, the **real interest rate** is corrected for inflation and is calculated as the nominal interest rate minus the rate of inflation. As an example, suppose the nominal interest rate is 8 percent per year and the inflation rate is 3 percent per year; we can calculate the real interest rate as $8 - 3 = 5$ percent per year.

To take a simple example, suppose that you live in an economy where the only product is bread. Further suppose that the price of bread in the first period is \$1 per loaf and that bread inflation is 3 percent per year. If you lend \$100 at 8 percent-per-year interest, you will have \$108 at the end of the year. However, because of inflation, next year you will get

**FIGURE 15-2. Most Interest Rates Move Together**

This graph shows the major interest rates in the U.S. economy. The lowest rate is generally the federal funds rate, set by the Federal Reserve in its monetary policy. Longer-term and riskier interest rates are usually higher than safe and short-term rates.

Source: Federal Reserve System, available at www.federalreserve.gov/releases/.

Asset class	Period	Nominal rate of return (% per year)	Real rate of return (% per year)
Government securities:			
3 month	1960–2008	5.2	1.0
10 year	1960–2008	6.9	2.7
Corporate bonds:			
Safe (Aaa rated)	1960–2008	7.7	3.4
Risky (Baa rated)	1960–2008	8.7	4.4
Corporate equities			
	1960–2008	9.9	5.6
Consumer loans:			
Mortgages (fixed rate)	1971–2008	9.2	4.9
Credit cards	1972–2008	16.4	11.8
New-car loans	1972–2008	10.4	6.0

TABLE 15-1. Interest Rates on Major Financial Assets

Safe government securities have the lowest yields. Note that consumers pay a substantial penalty on credit-card debt (students beware!). The real interest rates are corrected for inflation. Note that Aaa bonds are the safest type of corporate security, while Baa securities have significant risks of bankruptcy.

Source: Federal Reserve Board, available at www.federalreserve.gov/releases/, and Department of Commerce.

back only 105 (and not 108) loaves of bread. The real (or bread) rate of interest is $8 - 3 = 5$ percent.²

During inflationary periods, we must use real interest rates, not nominal or money interest rates, to calculate the yield on investments in terms of goods earned per year on goods invested. The real interest rate is approximately equal to the nominal interest rate minus the rate of inflation.

² The exact algebra of real interest rates is as follows: Let π be the inflation rate, i the nominal interest rate, and r the real interest rate. If you invest \$1 today, you get $\$(1 + i)$ back in 1 year. However, prices have risen, so you need $\$(1 + \pi)$ in 1 year to buy the same amount of goods that you could buy with \$1 today. Instead of buying 1 unit of goods today, you can therefore buy $(1 + r)$ units tomorrow, where $(1 + r) = (1 + i)/(1 + \pi)$. For small values of i and π , $r = i - \pi$.

The World's Safest Investment

U.S. Treasury bonds are generally considered a riskless investment. Their one shortcoming is that they pay a fixed-dollar interest rate. This means that if inflation heats up, the real interest rate could easily turn negative.

In 1997, the U.S. government fixed this problem by introducing Treasury inflation-protected securities (TIPS). TIPS have their interest and principal tied to inflation, so they pay a constant real interest rate over their lifetime.

This is how these special bonds work: Each year the principal value is adjusted by the increase in the consumer price index (CPI). Let's take a specific example: In January 2000, the Treasury issued a $4\frac{1}{4}$ percent 10-year inflation-protected bond. Between January 2000 and June 2003, the CPI increased by 12 percent. Therefore, the same

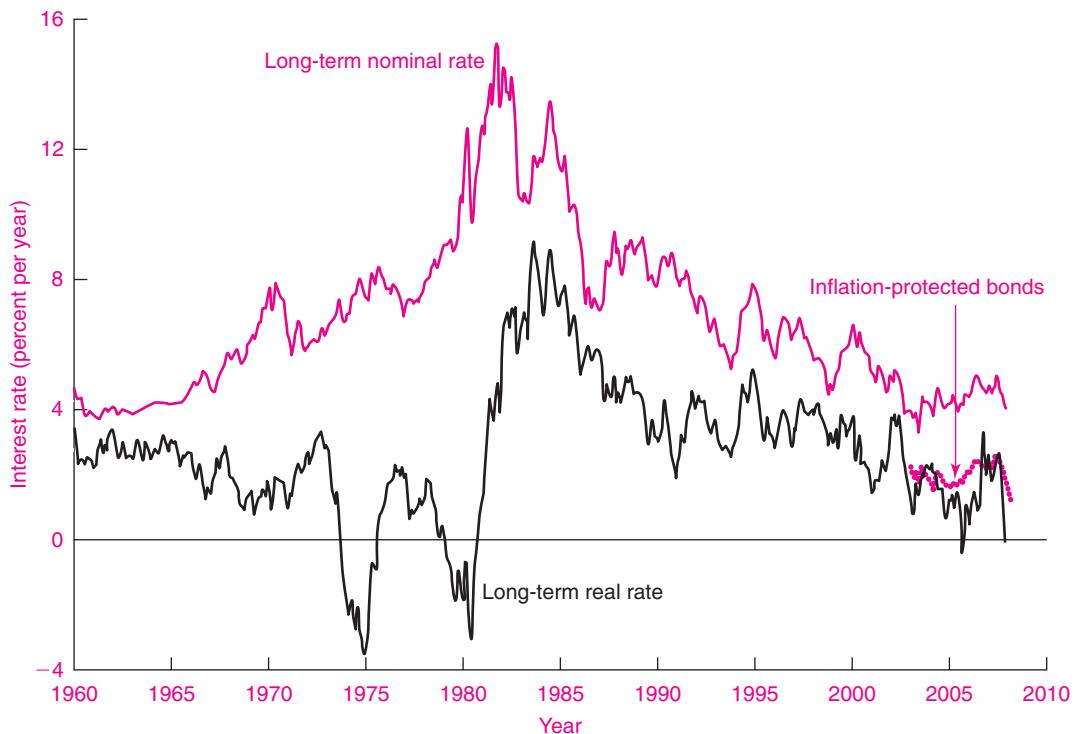


FIGURE 15-3. Nominal vs. Real Interest Rates

The long green line shows the nominal interest rate on long-term Treasury bonds. The blue line shows the “calculated” real interest rate, equal to the nominal interest rate minus the realized inflation rate over the previous year. Note that real interest rates drifted downward until 1980. After 1980, however, real interest rates moved up sharply. The short green line since 2003 shows the real interest rate on long-term inflation-indexed securities.

Source: Federal Reserve Board, Department of Labor.

\$1000 bond bought in 2000 would be valued at \$1120 in June 2003. If the Treasury made an interest payment in June 2003, it would be 4½ percent of \$1120, instead of 4½ percent of \$1000 as would be the case for a standard bond. Let's further suppose that inflation averaged 3 percent per year from 2000 to 2010. This means that the principal value of the bond upon redemption would be \$1343.92 [= $\$1000 \times (1.3)^{10}$], instead of the \$1000 for a conventional bond.

As long as people expect that there will be inflation in the coming years, the interest rate on TIPS will be less than that on standard Treasury bonds. For example, in April 2008, standard 10-year Treasury bonds had a nominal yield of 3.6 percent, while 10-year TIPS had a real yield of 1.2 percent. This indicates that the average investor expected 10-year inflation to average $3.6 - 1.2 = 2.4$ percent per year.

The difference between nominal and real interest rates on long-term bonds is illustrated in Figure 15-3. The upper line shows the nominal interest rate, while the long lower line shows the calculated real interest rate. In addition, the short green segment that begins in 2003 shows the real interest rate on TIPS. This figure shows that the rise in nominal interest rates from 1960 to 1980 was purely illusory, for nominal interest rates were just keeping up with inflation during those years. After 1980, however, real interest rates rose sharply and remained high for a decade. The data on TIPS show that the real interest rate declined sharply during the credit crisis of 2007–2008.

Economists have long been enthusiasts of indexed bonds. Such bonds can be bought by pensioners who wish to guarantee that their retirement incomes will not be eroded away by inflation. Similarly, parents who wish to save for their children's education can sock away some of their savings knowing that their investment will keep up with the general price level. Even monetary-policy makers find value in indexed bonds, for the difference between the interest on conventional bonds and that on TIPS gives an indication of what is happening to expected inflation. The main puzzle for many economists is why it took the government so long to introduce this important innovation.

B. THE THEORY OF CAPITAL, PROFITS, AND INTEREST

Now that we have surveyed the major concepts, we turn to an analysis of the *theory of capital and interest*. This theory explains how the supply and demand for

capital determines returns such as real interest rates and profits.

BASIC CAPITAL THEORY

Roundaboutness

In Chapter 2, we noted that investment in capital goods involves indirect or *roundabout* production. Instead of catching fish with our hands, we find it ultimately more worthwhile first to build boats and make nets—and then to use the boats and nets to catch many more fish than we could by hand.

Put differently, investment in capital goods involves forgoing present consumption to increase future consumption. Consuming less today frees labor for making nets to catch many more fish tomorrow. In the most general sense, capital is productive because by forgoing consumption today we get more consumption in the future.

To see this, imagine two islands that are exactly alike. Each has the same amount of labor and natural resources. Island A uses these primary factors directly to produce consumption goods like food and clothing; it uses no produced capital goods at all. By contrast, thrifty Island B sacrifices current consumption and uses its resources and labor to produce capital goods, such as plows, shovels, and looms. After this temporary sacrifice of current consumption, B ends up with a large stock of capital goods.

Figure 15-4 shows the way that Island B forges ahead of A. For each island, measure the amount of consumption that can be enjoyed while maintaining the existing capital stock. Because of its thrift, Island B, using roundabout, capital-intensive methods of production, will enjoy more future consumption than Island A. Island B gets more than 100 units of future-consumption goods for its initial sacrifice of 100 units of present consumption.

By sacrificing current consumption and building capital goods today, societies can increase their consumption in the future.

Diminishing Returns and the Demand for Capital

What happens when a nation sacrifices more and more of its consumption for capital accumulation and production becomes more and more roundabout or indirect? We would expect the law of diminishing

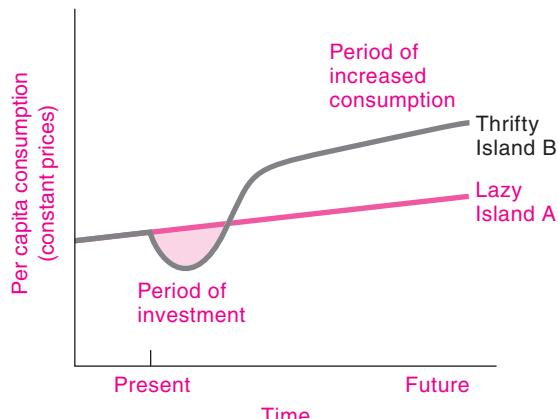


FIGURE 15-4. Investments Today Yield Consumption Tomorrow

Two islands begin with equal endowments of labor and natural resources. Lazy Island A invests nothing and shows a modest growth in per capita consumption. Thrifty Island B devotes an initial period to investment, forgoing consumption, and then enjoys the harvest of much higher consumption in the future.

returns to set in. Let's take the example of computers. The first computers were expensive and used intensively. Four decades ago, scientists would eke every last hour of time from an expensive mainframe computer that had less power than today's personal computer. By 2009, the nation's stock of computers had millions of times more computational and storage capacity. Therefore, the marginal product of computer power—the value of the last calculation or the last byte of storage—had diminished greatly as computer inputs increased relative to labor, land, and other capital. More generally, as capital accumulates, diminishing returns set in and the rate of return on the investments tends to fall.

Surprisingly, the rate of return on capital has not fallen markedly over the course of the last two centuries, even though our capital stocks have grown manyfold. Rates of return have remained high because innovation and technological change have created profitable new opportunities as rapidly as past investment has annihilated them. Even though computers are thousands of times more powerful than they were a few years ago, new applications in every corner of society from medical diagnostics to Internet commerce continue to make investments in computers profitable.



Irving Fisher: Economist as Crusader

Irving Fisher (1867–1947) was a multifaceted genius and crusader. His pioneering economic research ranged from fundamental theoretical studies on utility and capital theory to practical investigations into business cycles, index numbers, and monetary reform.

Among his fundamental contributions was the development of a complete theory of capital and interest in *The Nature of Capital and Income* (1906) and *The Theory of Interest* (1907). Fisher described the interplay between the interest rate and innumerable other elements of the economy. Yet the basic determinants of the interest rate, Fisher showed, were two fundamental pillars: impatience as reflected in “time discounting” and investment opportunity as reflected in the “marginal rate of return over cost.” It was Fisher who uncovered the deep relationship between interest and capital and the economy, as described in this summary from *The Theory of Interest*:

The truth is that the rate of interest is not a narrow phenomenon applying only to a few business contracts, but permeates all economic relations. It is the link which binds man to the future and by which he makes all his far-reaching decisions. It enters into the price of securities, land, and capital goods generally, as well as into rent, wages, and the value of all “interactions.” It affects profoundly the distribution of wealth. In short, upon its accurate adjustment depend the equitable terms of all exchange and distribution.

Fisher always aimed at research that could be empirically applied. His philosophy is embodied in the Econometric Society, which he helped found, whose constitution trumpeted a science which would lead to “the advancement of economic theory in its relation to statistics and mathematics [and] the unification of the theoretical-quantitative and the empirical-quantitative approach.”

In addition to research on pure economics, Fisher was a habitual crusader. He lobbied for a “compensated dollar” as a substitute for the gold standard. After he contracted tuberculosis, he became an impassioned advocate for improved health and developed 15 rules of personal hygiene. These included a strong advocacy of Prohibition and idiosyncrasies such as chewing 100 times before swallowing. It is said that with no alcohol and much chewing, dinner parties at the Fishers were not the liveliest gatherings in New Haven.

Fisher's most famous forecast came in 1929 when he argued that the stock market had achieved a “permanent plateau of prosperity.” He put his money behind his

forecast, and his substantial wealth was wiped out in the Great Depression.

Even though Fisher's financial acumen has been questioned, his legacy in economics has grown steadily, and he is generally regarded as the greatest American economist of all time.

Determination of Interest and the Return on Capital

We can use the classical theory of capital to understand the determination of the rate of interest. Households *supply* funds for investment by abstaining from consumption and accumulating savings over time. At the same time, businesses *demand* capital goods to combine with labor, land, and other inputs. In the end, a firm's demand for capital is driven by its desire to make profits by producing goods.

Or, as Irving Fisher put the matter a century ago:

The quantity of capital and the rate of return on capital are determined by the interaction between (1) people's *impatience* to consume now rather than accumulate more capital goods for future consumption (perhaps for old-age retirement or for that proverbial rainy day); and (2) *investment opportunities* that yield higher or lower returns to such accumulated capital.

To understand interest rates and the return on capital, consider an idealized case of a closed economy with perfect competition and without risk or inflation. In deciding whether to invest, a profit-maximizing firm will always compare its cost of borrowing funds with the rate of return on capital. If the rate of return is higher than the market interest rate at which the firm can borrow funds, it will undertake the investment. If the interest rate is higher than the rate of return on investment, the firm will not invest.

Where will this process end? Eventually, firms will undertake all investments whose rates of return are higher than the market interest rate. Equilibrium is then reached when the amount of investment that firms are willing to undertake at a given interest rate just equals the savings which that interest rate calls forth.

In a competitive economy without risk or inflation, the competitive rate of return on capital would be equal to the market interest rate. The market interest rate serves two functions: It rations out society's scarce supply of capital goods for the uses that have

the highest rates of return, and it induces people to sacrifice current consumption in order to increase the stock of capital.

Graphical Analysis of the Return on Capital

We can illustrate capital theory by concentrating on a simple case in which all physical capital goods are alike. In addition, assume that the economy is in a steady state with no population growth or technological change.

In Figure 15-5, *DD* shows the demand curve for the stock of capital; it plots the relationship between the quantity of capital demanded and the rate of return on capital. Recall from Chapter 12 that the demand for a factor like capital is a derived demand—the demand comes from the *marginal product of capital*, which is the extra output yielded by additions to the capital stock.

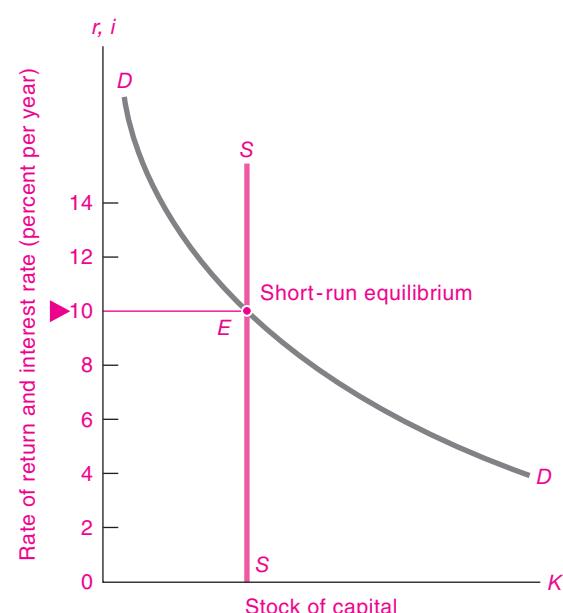


FIGURE 15-5. Short-Run Determination of Interest and Returns

In the short run, the economy has inherited a given stock of capital from the past, shown as the vertical *SS* supply-of-capital schedule. The intersection of the short-run supply curve with the demand-for-capital schedule determines the short-run return on capital, and the short-run real interest rate, at 10 percent per year.

The law of diminishing returns can be seen in the fact that the demand-for-capital curve in Figure 15-5 is downward-sloping. When capital is very scarce, the most profitable projects have a very high rate of return. Gradually, as the community exploits all the high-yield projects by accumulating capital, with total labor and land fixed, diminishing returns to capital set in. The community must then invest in lower-yield projects as it moves down the demand-for-capital curve.

Short-Run Equilibrium. We can now see how supply and demand interact. In Figure 15-5, past investments have produced a given stock of capital, shown as the vertical short-run supply curve, SS . Firms will demand capital goods in a manner shown by the downward-sloping demand curve, DD .

At the intersection of supply and demand, at point E , the amount of capital is just rationed out to the demanding firms. At this short-run equilibrium, firms are willing to pay 10 percent a year to borrow funds to buy capital goods. At that point, the lenders

of funds are satisfied to receive exactly 10 percent a year on their supplies of capital.

Thus, in our simple, riskless world, the rate of return on capital exactly equals the market interest rate. Any higher interest rate would find firms unwilling to borrow for their investments; any lower interest rate would find firms clamoring for the too scarce capital. Only at the equilibrium interest rate of 10 percent are supply and demand equilibrated. (Recall that these are *real* interest rates because there is no inflation.)

But the equilibrium at E is sustained only for the short run: At this high interest rate, people desire to accumulate more wealth, that is, to continue saving and investing. This means that the capital stock increases. However, because of the law of diminishing returns, the rate of return and the interest rate move downward. As capital increases—while other things such as labor, land, and technical knowledge remain unchanged—the rate of return on the increased stock of capital goods falls to ever-lower levels.

This process is shown graphically in Figure 15-6. Note that capital formation is taking place at point E .

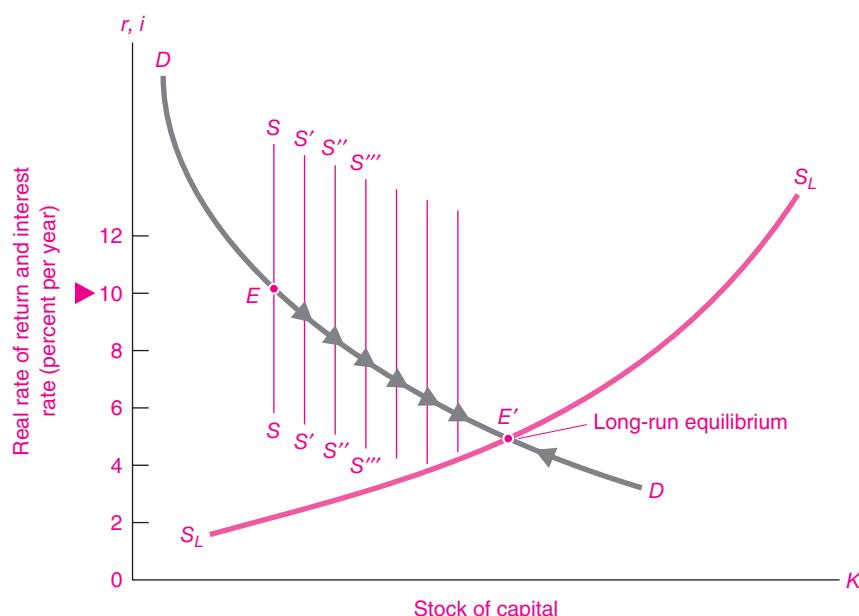


FIGURE 15-6. Long-Run Equilibration of the Supply and Demand for Capital

In the long run, society accumulates capital, so the supply curve is no longer vertical. As pictured here, the supply of capital and wealth is responsive to higher interest rates. At the original short-run equilibrium at E there is net investment, so the economy moves down the DD demand curve as shown by the blue arrows. Long-run equilibrium comes at E' , where net saving ceases.

So each year, the capital stock is a little higher as net investment occurs. As time passes, the community moves slowly down the *DD* curve as shown by the blue arrows in Figure 15-6. You can actually see a series of very thin short-run supply-of-capital curves in the figure— S , S' , S'' , S''' , These curves show how the short-run supply of capital increases with capital accumulation.

Long-Run Equilibrium. The eventual equilibrium is shown at E' in Figure 15-6; this is where the long-run supply of capital (shown as $S_L S_L'$) intersects with the demand for capital. In long-run equilibrium, the real interest rate settles at that level where the quantity of capital that firms desire to hold just matches the value of wealth that people want to own. At the long-run equilibrium, net saving stops, net capital accumulation is zero, and the capital stock is no longer growing.

Would investment gradually decline to zero as all investment opportunities are exhausted? Some economists (such as Joseph Schumpeter) have likened the investment process to a plucked violin string: In a world of unchanging technology, the string gradually comes to rest as capital accumulation drives down returns on capital. But before the economy has settled into a steady state, an outside event or invention comes along to pluck the string and set the forces of investment in motion again.

The long-run equilibrium stock of capital comes at that real interest rate where the value of assets that people want to hold exactly matches the amount of capital that firms want for production.

PROFITS AS A RETURN TO CAPITAL

Now that we have examined the determinants of the return to capital, we turn to an analysis of profits. In addition to discussing wages, interest, and rent, economists often talk about a fourth category of income called *profits*. What are profits? How do they differ from interest and the returns on capital more generally?

Reported Profit Statistics

Before we present the economic concepts, we begin with the measures used in accounting. Accountants

define profits as the difference between total revenues and total costs. To calculate profits, accountants start with total revenues and subtract all expenses (wages, salaries, rents, materials, interest, excise taxes, and the rest). The leftover residual is called profits.

It is important in analyzing profits, however, to distinguish between *accounting profits* and *economic profits*. Accounting profits (also called business income or business earnings) are the residual income measured in financial statements by accountants. Economic profits are the earnings after all costs—both money and implicit or opportunity costs—are subtracted. These concepts of profits differ because accounting profits omit some implicit returns. The opportunity costs of factors owned by firms are called *implicit returns*.

For example, most businesses own much of their capital, and there is no accounting charge for the opportunity cost or implicit return on owned capital. Accounting profits therefore include an implicit return on the capital owned by firms. In large corporations, economic profits would equal business profits minus an implicit return on the capital owned by the firm along with any other costs not fully compensated at market prices. Economic profits are generally smaller than business profits.

Determinants of Profits

What determines the rate of profit in a market economy? Profits are in fact a combination of different elements, including implicit returns on owners' capital, rewards for risk-bearing and innovative profits.

Profits as Implicit Returns. Much of reported business profits is primarily the return to the owners of the firm for the factors of production, including capital and labor provided by the owners. For example, some profits are the return on the personal work provided by the owners of the firm—such as the doctor or the lawyer who works in a small professional corporation. Another part is the rent return on the land owned by the firm. In large corporations, most profits represent the opportunity costs of invested capital.

Thus some of what is ordinarily called profit is really nothing but “implicit rentals,” “implicit rent,” and “implicit wages,” which are the earnings on factors that the firm itself owns.

Profits as Rewards for Risk-Bearing. Profits also include a reward for the riskiness of the relevant investments. Most businesses must incur a risk of default, which occurs when a loan or investment cannot be paid, perhaps because the borrower went bankrupt. In addition, there are many insurable risks, such as those for fires or hurricanes, which can be covered through the purchase of insurance. A further concern is the uninsurable or systematic risk of investments. A company may have a high degree of sensitivity to business cycles, which means that its earnings fluctuate a great deal when aggregate output goes up or down. All of these risks must either be insured against or earn a risk premium in profits.

Profits as Reward for Innovation. A third kind of profits consists of the returns to innovation and invention. A growing economy is constantly producing new goods and services—from telephones in the nineteenth century to automobiles early in the twentieth century

to computer software in the present era. These new products are the result of research, development, and marketing. We call the person who brings a new product or process to market an *innovator* or *entrepreneur*.

What do we mean by “innovators”? Innovators are people who have the vision, originality, and daring to introduce new ideas. Our economy has been revolutionized by the discoveries of great inventors like Alexander Graham Bell (telephone), Jack Kilby (integrated circuit), and Kary Mullis (polymerase chain reaction).

Every successful innovation creates a temporary pool of monopoly. We can identify innovational profits (sometimes called Schumpeterian profits) as the temporary excess return to innovators and entrepreneurs. These profit earnings are temporary and are soon competed away by rivals and imitators. But just as one source of innovational profits disappears, another is being born. An economy will generate this type of profits as long as it innovates.

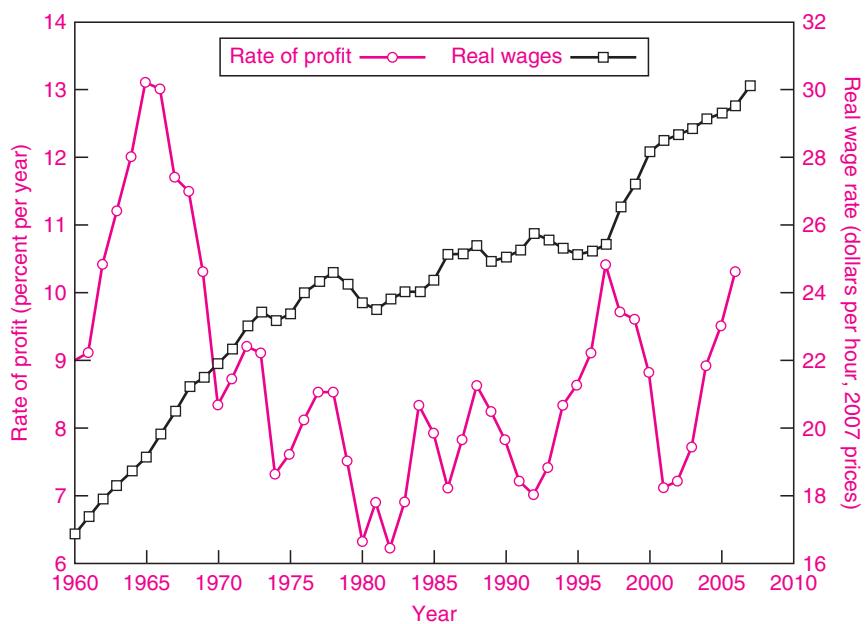


FIGURE 15-7. Trends in Wages and Rate of Profit in the United States

How have the returns to labor and capital varied in recent years? Average real wages have continued to grow. After peaking in the mid-1960s, the pretax rate of profit on American business capital fell sharply and then meandered around over the last three decades, with an average of around 8 percent per year.

Source: U.S. Departments of Commerce and Labor.

Corporate profits are the most volatile component of national income. The rights to earn corporate profits—represented by the ownership of corporate stocks or equities—must therefore provide a significant premium to attract risk-averse investors. This excess return on equities above that on risk-free investments is called the *equity premium*. Empirical studies suggest that the equity premium averaged around 5 percent per year over the twentieth century (see Table 15-1 on page 289).

Profits are a residual income item, equal to total revenues minus total costs. Profits contain elements of implicit returns (such as return on owners'

capital), return for risk-bearing, and innovational profits.

Empirical Evidence on Returns to Labor and Capital

We close with a look at the actual trends in the return to labor and capital in the United States over the last four decades, as illustrated in Figure 15-7. Real wages (which are average hourly earnings corrected for movements in the consumer price index) grew steadily. The pretax rate of profit on capital declined from its peak in the mid-1960s and has averaged around 8 percent per year for the last three decades.



SUMMARY

A. Basic Concepts of Interest and Capital

1. Recall the major concepts:
 - *Capital*: durable produced items used for further production
 - *Rentals*: net annual dollar returns on capital goods
 - *Rate of return on investment*: net annual receipts on capital divided by dollar value of capital (measured as percent per year)
 - *Interest rate*: yield on financial assets, measured as percent per year
 - *Real interest rate*: yield on funds corrected for inflation, also measured as percent per year
 - *Present value*: value today of an asset's stream of future returns
2. Interest rates are the rate of return on financial assets, measured in percent per year. People willingly pay interest because borrowed funds allow them to buy goods and services to satisfy current consumption needs or make profitable investments.
3. We observe a wide variety of interest rates. These rates vary because of many factors such as the term or maturity of loans, the risk and liquidity of investments, and the tax treatment of the interest.
4. Nominal or money interest rates generally rise during inflationary periods, reflecting the fact that the purchasing power of money declines as prices rise. To calculate the interest yield in terms of real goods and services, we use the real interest rate, which equals the nominal interest rate minus the rate of inflation.

5. Assets generate streams of income in future periods. By calculating the present value of the asset, we can convert the stream of future returns into a single value today. This is done by asking what sum today will generate the total value of all future returns when invested at the market interest rate.

6. The exact present-value formula is as follows: Each dollar payable t years from now has a present value (V) of $\$1/(1 + i)^t$. So for any net-receipt stream (N_1, N_2, \dots, N_t), where N_t is the dollar value of receipts t years in the future, we have

$$V = \frac{N_1}{1 + i} + \frac{N_2}{(1 + i)^2} + \cdots + \frac{N_t}{(1 + i)^t} + \cdots$$

B. The Theory of Capital, Profits, and Interest

7. A third factor of production is capital, a produced durable item that is used in further production. In the most general sense, investing in capital represents deferred consumption. By postponing consumption today and instead producing buildings or equipment, society increases consumption in the future. It is an economic fact that roundabout production yields a positive rate of return.
8. Interest is a device that serves two functions in the economy: As a motivating device, it provides an incentive for people to save and accumulate wealth. As a rationing device, interest allows society to select only those investment projects with the highest rates of

return. However, as more and more capital is accumulated, and as the law of diminishing returns sets in, the rate of return on capital and the interest rate will be beaten down by competition. Falling interest rates are a signal to society to adopt more capital-intensive projects with lower rates of return.

9. Saving and investing involve waiting for future consumption rather than consuming today. Such thrift interacts with the net productivity of capital to determine interest rates, the rate of return on capital, and the capital stock. The funds or financial assets needed to purchase capital are provided by households that are willing to sacrifice consumption today in return for larger consumption tomorrow. The demand for capital comes from firms that have a variety of roundabout investment projects. In long-run equilibrium, the interest rate is thus determined by the interaction

between the net productivity of capital and the willingness of households to sacrifice consumption today for consumption tomorrow.

10. Profits are revenues minus costs. Remember that economic profits differ from those measured by accountants. Economics distinguishes between three categories of profits: (a) An important source is profits as implicit returns. Firms generally own many of their own nonlabor factors of production—capital, natural resources, and patents. In these cases, the implicit return on owned inputs is part of the profits. (b) Another source of profits is uninsured or uninsurable risk, particularly that associated with the business cycle. (c) Finally, innovational profits will be earned by entrepreneurs who introduce new products or innovations.

CONCEPTS FOR REVIEW

capital, capital goods
tangible assets vs. financial assets
rentals, rate of return on capital,
interest rate, profits
present value
interest rate, real and nominal
interest-rate premiums due to
maturity, risk, illiquidity

inflation-indexed bonds
investment as abstaining from current
consumption
present value
twin elements in interest
determination:
returns to roundaboutness
impatience

elements of profits:
implicit returns
risk
innovation

FURTHER READING AND INTERNET WEBSITES

Further Reading

The foundations of capital theory were laid by Irving Fisher, *The Theory of Interest* (Macmillan, New York, 1930). You can pursue advanced topics in finance theory in an intermediate textbook such as Lawrence S. Ritter, William L. Silber, and Gregory F. Udell, *Principles of Money, Banking, and Financial Markets*, 11th ed. (Addison Wesley Longman, New York, 2003). The standard reference on U.S. monetary history is Milton Friedman and Anna Jacobson Schwartz, *Monetary History of the United States 1867–1960* (Princeton University Press, Princeton, N.J., 1963).

Modern capital and finance theories are very popular subjects and are often covered in the macroeconomics part of an introductory course or in special courses. A good book on the subject is Burton Malkiel, *A Random Walk down Wall Street* (Norton, New York, 2003). A recent book surveying financial history and theory and arguing that the stock market was extraordinarily overvalued in the bull market of 1981–2000 is Robert Shiller, *Irrational Exuberance*, 2nd ed. (Princeton University Press, Princeton, N.J., 2005). A recent summary of evidence on the efficient-market theory by Burton Malkiel and

Robert Shiller is found in *Journal of Economic Perspectives*, Winter 2003.

Websites

Data on financial markets are plentiful. See finance.yahoo.com for an entry point into stock and bond markets as well

as information on individual companies. Also see www.bloomberg.com for up-to-date financial information.

Data on financial markets are also produced by the Federal Reserve System at www.federalreserve.gov.

QUESTIONS FOR DISCUSSION

1. Calculate the present value of each of the following income streams, where I_t = the income t years in the future and i is the constant interest rate in percent per year. Round to two decimal points where the numbers are not integers.
 - a. $I_0 = 10, I_1 = 110, I_3 = 133.1; i = 10.$
 - b. $I_0 = 17, I_1 = 21, I_2 = 33.08, I_3 = 23.15; i = 5.$
 - c. $I_0 = 0, I_1 = 12, I_2 = 12, I_3 = 12, \dots; i = 5.$
2. Contrast the following four returns on durable assets: (a) rent on land, (b) rental of a capital good, (c) rate of return on a capital good, and (d) real interest rate. Give an example of each.
3. Interest-rate problems (which may require a calculator):
 - a. You invest \$2000 at an interest rate of 13.5 percent per year. What is your total balance after 6 months?
 - b. Interest is said to be “compounded” when you earn interest on whatever interest has already been paid; most interest rates quoted today are compounded. If you invest \$10,000 for 3 years at a compound annual interest rate of 10 percent, what is the total value of the investment at the end of each year?
 - c. Consider the following data: The consumer price index in 1977 was 60.6, and in 1981 it was 90.9. Interest rates on government securities in 1978 through 1981 (in percent per year) were 7.2, 10.0, 11.5, and 14.0. Calculate the average nominal and real interest rates for the 4-year period 1978–1981.
 - d. Treasury bills (T-bills) are usually sold on a discounted basis; that is, a 90-day T-bill for \$10,000 would sell today at a price such that collecting \$10,000 at maturity would produce the market interest rate. If the market interest rate is 6.6 percent per year, what would be the price on a 90-day \$10,000 T-bill?
4. Present-value questions:
 - a. Consider the 1-year bond in the discussion of present value. Calculate the present value of the bond if the interest rate is 1, 5, 10, and 20 percent.
5. What is the value of a perpetuity yielding \$16 per year at interest rates of 1, 5, 10, and 20 percent per year?
6. Compare the answers to a and b. Which asset is more sensitive to interest-rate changes? Quantify the difference.
7. Using the supply-and-demand analysis of interest, explain how each of the following would affect interest rates in capital theory:
 - a. An innovation that increased the marginal product of capital at each level of capital
 - b. A decrease in the desired wealth holdings of households
 - c. A 50 percent tax on the return on capital (in the short run and the long run)
8. Looking back to Figures 15-5 and 15-6, review how the economy moved from the short-run equilibrium interest rate at 10 percent per year to the long-run equilibrium. Now explain what would occur in both the long run and the short run if innovations shift up the demand-for-capital curve. What would happen if the government debt became very large and a large part of people's supply of capital was siphoned off to holdings of government debt? Draw new figures for both cases.
9. Explain the rule for calculating the present discounted value of a perpetual income stream. At a 5 percent interest rate, what is the worth of a perpetuity paying \$100 per year? Paying \$200 per year? Paying \$N per year? At 10 or 8 percent, what is the worth of a perpetuity paying \$100 per year? What does doubling the interest rate do to the capitalized value of a perpetuity—say, a perpetual bond?
10. Recall the algebraic formula for a convergent geometric progression:

$$1 + K + K^2 + \dots = \frac{1}{1 - K}$$

for any fraction K less than 1. If you set $K = 1/(1 + i)$, can you verify the present-value formula for a permanent income stream, $V = \$N/i$? Provide an alternative

proof using common sense. What would be the value of a lottery that paid you and your heirs \$5000 per year forever, assuming an interest rate of 6 percent per year?

9. The value of land in Manhattan was around \$150 billion in 2008. Imagine that it is 1626 and you are the economic adviser to the Dutch when they are considering whether to buy Manhattan from the Manhasset Indians. Further, assume that the relevant interest rate for calculating the present value is 4 percent per year. Would you advise the Dutch that a purchase price of \$24 is a good deal or not? How would your answer change if the interest rate were 6 percent? 8 percent? (*Hint:* For each interest rate, calculate the present value in 1626 of the land value as of 2008. Then compare that

with the purchase price in 1626. For this example, simplify by assuming that the owners collect no rents on the land. As an advanced further question, assume that the rent equals 2 percent of the value of the land each year.)

10. An increase in interest rates will generally lower the prices of assets. To see this, calculate the present value of the following two assets at interest rates of 5 percent, 10 percent, and 20 percent per year:
- A perpetuity yielding \$100 per year
 - A Christmas tree that will sell for \$50 one year from now

Explain why the price of the long-lived asset is more sensitive to interest-rate changes than the price of the short-lived asset.

PART FOUR

Applications of Economic Principles

Government Taxation and Expenditure

16



The spirit of a people, its cultural level, its social structure, the deeds its policy may prepare, all this and more is written in its fiscal history. . . He who knows how to listen to its messenger here discerns the thunder of world history more clearly than anywhere else.

Joseph Schumpeter

When we look at a market economy—providing all sorts of products from apples and boats to X-ray machines and zithers—it would be tempting to think that markets require little more than skilled workers and lots of capital. But history has shown that markets cannot work effectively alone. At a minimum, an efficient market economy needs police to ensure physical security, an independent judicial system to enforce contracts, regulatory mechanisms to prevent monopolistic abuses and lethal pollution, schools to educate the young, and a public health system to ward off communicable diseases. Exactly where to draw the line between government and private activities is a difficult and controversial question, and people today debate the appropriate role of government in education, health care, and income support.

As economists, we want to go beyond the partisan debates and analyze the functions of government—government's comparative advantage in the mixed economy. The present chapter examines the role of government in an advanced economy. What are the appropriate goals for economic policy in a market economy, and what instruments are available to carry them out? What principles underlie an efficient tax

system? Understanding the answers to these questions is key to developing sound public policies.

A. GOVERNMENT CONTROL OF THE ECONOMY

Debates about the role of government often take place on bumper stickers, with rallying cries such as “No new taxes” or “Balance the budget.” These simplistic phrases cannot capture the serious business of government economic policy. Say the populace decides that it wants to devote more resources to improving public health; or that more resources should be devoted to educating the young; or that unemployment in a deep recession should be reduced. A market economy cannot automatically solve these problems. Each of these objectives can be met if and only if the government changes its taxes, spending, or regulations. The thunder of world history is heard in fiscal policy because taxing and spending are such powerful instruments for social change.

THE TOOLS OF GOVERNMENT POLICY

In a modern economy, no sphere of economic life is untouched by the government. We can identify three major instruments or tools that government uses to influence private economic activity:

1. *Taxes* on incomes and goods and services. These reduce private income, thereby reducing private expenditures (on automobiles or restaurant food) and providing resources for public expenditures (on missiles and school lunches). The tax system also serves to discourage certain activities by taxing them more heavily (such as smoking cigarettes) while encouraging other activities by taxing them lightly or even subsidizing them (such as health care).
2. *Expenditures* on certain goods or services (such as roads, education, or police protection), along with *transfer payments* (like social security and food stamps) that provide resources to individuals.
3. *Regulations* or controls that direct people to perform or refrain from certain economic activities. Examples include rules that limit the amount firms

can pollute, or that divide up the radio spectrum, or that mandate testing the safety of new drugs.

Trends in the Size of Government

For more than a century, national income and production have been rising in all economies. At the same time, in most countries, government expenditures have been rising even faster than the overall economy. Each period of emergency—depression, war, or concern over social problems such as poverty or pollution—expanded the activity of government. After the crisis passed, government controls and spending never returned to their previous levels.

Before World War I, the combined federal, state, and local government expenditures or taxation amounted to little more than one-tenth of the entire U.S. national income. The war effort during World War II compelled government to consume about half the nation's greatly expanded total output. By 2007, expenditures of all levels of government in the United States ran around 33 percent of GDP.

Figure 16-1 shows the trend in taxes and expenditures for all levels of government in the United

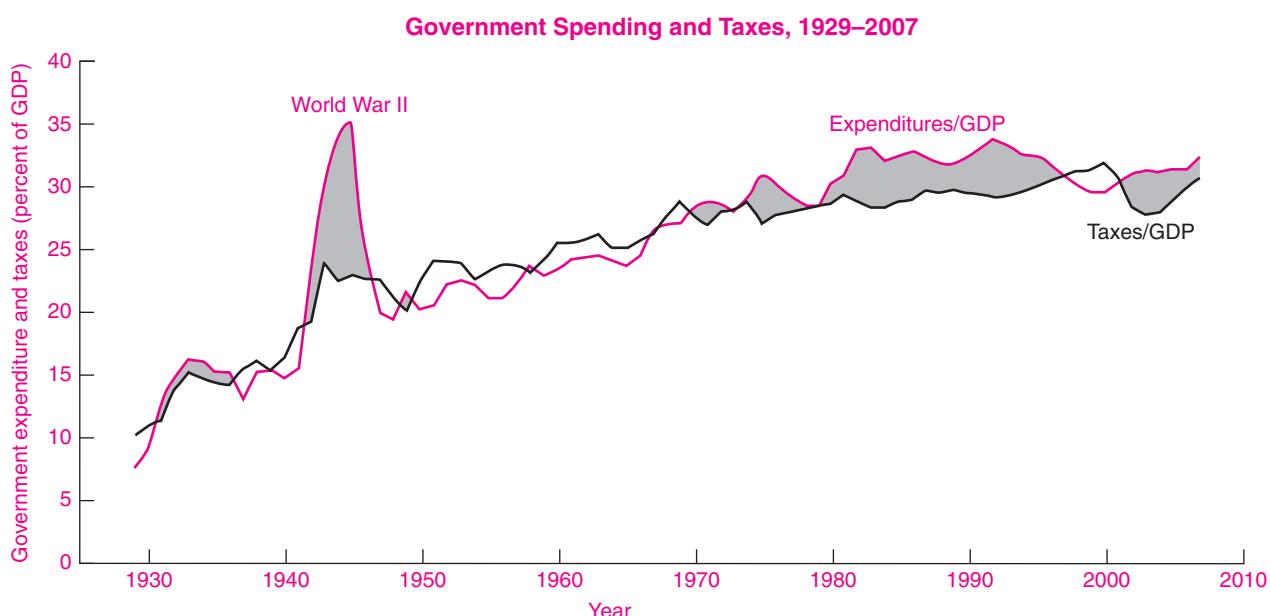


FIGURE 16-1. Government's Share of the Economy Has Grown Sharply

Government expenditures include spending on goods, services, and transfers at the federal, state, and local levels. Note how spending grew rapidly during wartime but did not return to prewar levels afterward. The difference between spending and taxes is the government deficit or surplus.

Source: U.S. Department of Commerce.

States. The rising curves indicate that the shares of government taxes and spending have grown steadily upward over recent decades.

Government's expansion has not occurred without opposition; each new spending and tax program provoked a fierce reaction. For example, when social security was first introduced in 1935, opponents denounced it as an ominous sign of socialism. But with the passage of time, political attitudes evolve. The "socialistic" social security system is today defended by politicians of all stripes as an essential part of the "social contract" between the generations. The radical doctrines of one era become accepted gospel of the next.

Figure 16-2 shows how government spending as a percentage of GDP varies among countries. High-income countries tend to tax and spend a larger fraction of GDP than do poor countries. Can we discern a pattern among wealthy countries? Within the high-income countries, no simple law

relating tax burdens and the citizenry's well-being can do justice to the true diversity of the fiscal facts of nations. For example, financing for education and health care, two of the largest components of government spending, is organized very differently across countries.

Figures 16-1 and 16-2 show the total expenditures of governments. Such expenditures include purchases of goods and services (like missiles and education) as well as transfer payments (like social security payments and interest on the government debt). Purchases of goods and services are called "exhaustive" because they make a direct claim upon the production of a country; transfer payments, by contrast, increase people's income and allow individuals to purchase goods and services but do not directly reduce the quantity of goods and services available for private consumption and investment.

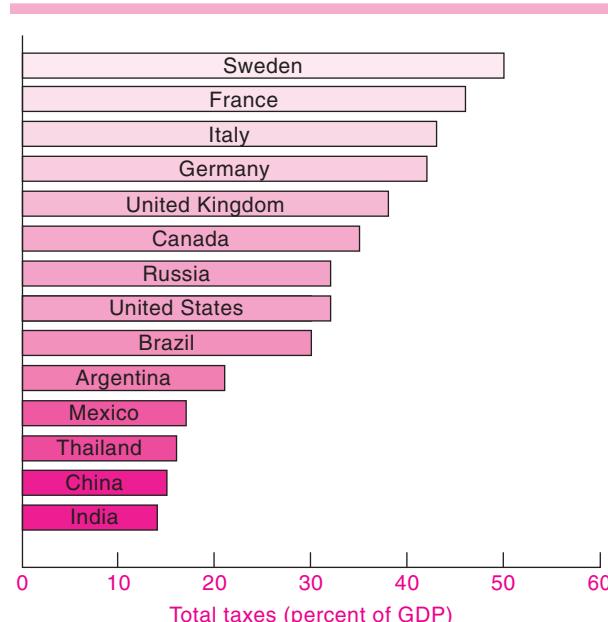


FIGURE 16-2. Government Taxation Is Highest in Rich Countries

Governments of poor countries tax and spend relatively little of national income. With affluence come greater demands for public goods and redistributive taxation to aid low-income families.

Source: United Nations for period 2000–2002, at unpan1.un.org/intradoc/groups/public/documents/un/unpan014052.pdf.

The Growth of Government Controls and Regulation

In addition to the growth in spending and taxing, there has also been a vast expansion in the laws and regulations governing economic affairs.

Nineteenth-century America came as close as any economy has come to being a pure laissez-faire society—the system that the British historian Thomas Carlyle labeled "anarchy plus the constable." This philosophy permitted people great personal freedom to pursue their economic ambitions and produced a century of rapid material progress. But critics saw many flaws in this laissez-faire idyll. Historians record periodic business crises, extremes of poverty and inequality, deep-seated racial discrimination, and poisoning of water, land, and air by pollution. Muckrakers and progressives called for a bridle on capitalism so that the people could steer this wayward beast in more humane directions.

Beginning in the 1890s, the United States gradually turned away from the belief that "government governs best which governs least." Presidents Theodore Roosevelt, Woodrow Wilson, Franklin Roosevelt, and Lyndon Johnson—in the face of strenuous opposition—pushed out the boundaries of federal control over the economy, devising new regulatory and fiscal tools to combat the economic ailments of their time.

Constitutional powers of government were interpreted broadly and used to “secure the public interest” and to “police” the economic system. In 1887, the federal Interstate Commerce Commission (ICC) was established to regulate rail traffic across state boundaries. Soon afterward, the Sherman Antitrust Act and other laws were aimed against monopolistic combinations in “restraint of trade.”

During the 1930s, a whole set of industries came under *economic regulation*, in which government sets the prices, conditions of exit and entry, and safety standards. Regulated industries since that time have included the airlines, trucking, and barge and water traffic; electric, gas, and telephone utilities; financial markets; and oil and natural gas, as well as pipelines.

In addition to regulating the prices and standards of business, the nation attempted to protect health and safety through increasingly stringent *social regulation*. Following the revelations of the muckraking era of the early 1900s, pure food and drug acts were passed. During the 1960s and 1970s, Congress passed a series of acts that regulated mine safety and then worker safety more generally; regulated air and water pollution; authorized safety standards for automobiles and consumer products; and regulated strip mining, nuclear power, and toxic wastes.

Over the last three decades, the growth in government programs slowed. Economists argued persuasively that many economic regulations were impeding competition and keeping prices up rather than down. In the area of social regulations, economists have emphasized the need to ensure that the marginal benefits of regulations exceed their marginal costs. Today, “entitlement programs” (programs available to everyone who meets certain well-defined eligibility criteria), such as pensions and health care, are now the major spending programs for most high-income countries.

Still, there is no likelihood of a return to the laissez-faire era. Government programs have changed the very nature of capitalism. Private property is less and less wholly private. Free enterprise has become progressively less free. Irreversible evolution is part of history.

THE FUNCTIONS OF GOVERNMENT

We are beginning to get a picture of how government directs and interacts with the economy. What are the appropriate economic goals for government

action in a modern mixed economy? Let’s examine the four major functions:

1. Improving economic efficiency
2. Reducing economic inequality
3. Stabilizing the economy through macroeconomic policies
4. Conducting international economic policy

Improving Economic Efficiency

A central economic purpose of government is to assist in the socially desirable allocation of resources. This is the *microeconomic* side of government policy; it concentrates on the *what* and *how* of economic life. Microeconomic policies differ among countries according to customs and political philosophies. Some countries emphasize a hands-off, laissez-faire approach, leaving most decisions to the market. Other countries lean toward heavy government regulation, or even public ownership of businesses, in which production decisions are made by government planners.

The United States is fundamentally a market economy. On any microeconomic issue, most people presume that the market will solve the economic problem at hand. But sometimes there is good reason for government to override the allocational decisions of market supply and demand.

The Limits of the Invisible Hand. Earlier chapters have explained how the invisible hand of perfect competition would lead to an efficient allocation of resources. But this invisible-hand result holds only under limited conditions. All goods must be produced efficiently by perfectly competitive firms. All goods must be private goods like loaves of bread, the total of which can be cut up into separate slices of consumption for different individuals. There can be no externalities like air pollution. Consumers and firms must be fully informed about the prices and characteristics of the goods they buy and sell.

If all these idealized conditions were met, the invisible hand could provide perfectly efficient production and distribution of national output, and there would be no need for government intervention to promote efficiency.

Yet even in this ideal case, if there were to be a division of labor among people and regions, and if a price mechanism were to work, government would

have an important role. Courts and police forces would be needed to ensure fulfillment of contracts, nonfraudulent and nonviolent behavior, freedom from theft and external aggression, and the legislated rights of property.

Inescapable Interdependencies. Laissez-faire with minimal government intervention might be a good system if the idealized conditions listed above were truly present. In reality, each and every one of the idealized conditions enumerated above is violated to some extent in all human societies. Unregulated factories do tend to pollute the air, water, and land. When contagious diseases threaten to break out, private markets have little incentive to develop effective public-health programs. Consumers are sometimes poorly informed about the characteristics of the goods they buy. The market is not ideal. There are market failures.

In other words, government often deploys its weapons to correct significant market failures, of which the most important are the following:

- *The breakdown of perfect competition.* When monopolies or oligopolies collude to fix prices or drive firms out of business, government may apply anti-trust policies or regulations.
- *Externalities and public goods.* The unregulated market may produce too much air pollution and too little investment in public health or basic science. Government can use its influence to control harmful externalities or to fund programs in science and public health. Government can levy taxes on activities which impose external public costs (such as cigarette smoking), or it can subsidize activities which are socially beneficial (such as education or prenatal health care).
- *Imperfect information.* Unregulated markets tend to provide too little information for consumers to make well-informed decisions. In an earlier era, hucksters hawked snake oil remedies that might just as easily kill you as cure you. This led to food and drug regulations requiring that pharmaceutical companies provide extensive data on the safety and efficacy of new drugs before they can be sold. The government also requires that companies provide information on energy efficiency of major household appliances like refrigerators and water heaters. In addition, government may use its spending power to collect and provide

needed information itself, as it does with automobile crash-and-safety data.

Clearly, there is much on the agenda of possible allocational problems for government to handle.

Reducing Economic Inequality

Even when the invisible hand is marvelously efficient, it may at the same time produce a very unequal distribution of income. Under laissez-faire, people end up rich or poor depending on where they were born, on their inherited wealth, on their talents and efforts, on their luck in finding oil, and on their gender or the color of their skin. To some people, the distribution of income arising from unregulated competition looks as arbitrary as the Darwinian distribution of food and plunder in the jungle.

In the poorest societies, there is little excess income to take from the better-off and provide to the unfortunate. However, as a nation becomes wealthier, it can devote more resources to provide basic necessities and social insurance for all of its residents. These activities are the role of the “welfare state”—in which governments provide a minimum living standard to all—which is surveyed in detail in the next chapter. The welfare states of North America and Western Europe now devote a significant share of their revenues to maintaining minimum standards of health, nutrition, and income.

Income redistribution is usually accomplished through taxation and spending policies. Most wealthy countries now rule that children shall not go hungry because of the economic circumstances of their parents; that the poor shall not die because of insufficient money for needed medical care; that the young shall receive free public education; and that the old shall live out their years with a minimum level of income. In the United States, these government activities are provided primarily by transfer programs, such as food stamps, Medicaid, and social security.

But attitudes about redistribution evolve as well. With rising tax burdens and government budget deficits, along with rising costs of income-support programs, taxpayers increasingly resist redistributive programs and progressive taxation.

Stabilizing the Economy through Macroeconomic Policies

Early capitalism was prone to financial panics and bouts of inflation and depression. Today government

has the responsibility of preventing calamitous business depressions by the proper use of monetary and fiscal policy, as well as regulation of the financial system. In addition, government tries to smooth out the ups and downs of the business cycle, in order to avoid either large-scale unemployment at the bottom of the cycle or high inflation at the top of the cycle. More recently, government has become concerned with finding economic policies which boost long-term economic growth. These questions are considered at length in the chapters on macroeconomics.

Conducting International Economic Policy

As we will see in Chapter 18's review of international trade, the United States has become increasingly linked to the global economy in recent years. Government now plays a critical role representing the interests of the nation on the international stage and negotiating beneficial agreements with other countries on a wide range of issues. We can group the international issues of economic policy into four main areas:

- *Reducing trade barriers.* An important part of economic policy involves harmonizing laws and reducing trade barriers so as to encourage fruitful international specialization and division of labor. In recent years, nations have negotiated a series of trade agreements to lower tariffs and other trade barriers on agricultural products, manufactured goods, and services.

Such agreements are often contentious. They sometimes harm certain groups, as when removing textile tariffs reduces employment in that industry. In addition, international agreements may require giving up national sovereignty as the price of raising incomes. Suppose that one country's laws protect intellectual property rights, such as patents and copyrights, while another country's laws allow free copying of books, videos, and software. Whose laws shall prevail?

- *Conducting assistance programs.* Rich nations have numerous programs designed to improve the lot of the poor in other countries. These involve direct foreign aid, disaster and technical assistance, the establishment of institutions like the World Bank to give low-interest-rate loans to poor countries, and concessionary terms on exports to poor nations.

- *Coordinating macroeconomic policies.* Nations have seen that fiscal and monetary policies of other nations affect inflation, unemployment, and financial conditions at home. The international monetary system cannot manage itself; establishing a smoothly functioning exchange-rate system is a prerequisite for efficient international trade. When the American credit crisis erupted in 2008, it quickly spread to Europe and threatened several European banks. Central banks needed to act in a coordinated fashion to ensure that a bank failure, or even the fear of failure, in one country did not spread like wildfire to the entire international financial system. Particularly in tightly integrated regions, like Western Europe, countries work to coordinate their fiscal, monetary, and exchange-rate policies, or even adopt a common currency, so that inflation, unemployment, or financial crises in one country do not spill over to hurt the entire area.

- *Protecting the global environment.* The most recent facet of international economic policy is to work with other nations to protect the global environment in cases where several countries contribute to or are affected by spillovers. The most active areas historically have been protecting fisheries and water quality in rivers. When the Antarctic ozone hole threatened public health, countries reached an agreement to limit the use of ozone-depleting chemicals. Other treaties are designed to reduce the threats of deforestation, global warming, and species extinction. Clearly, international environmental problems can be resolved only through the cooperation of many nations.

Even the staunchest conservatives agree that government has a major role to play in representing the national interest in the anarchy of nations.

PUBLIC-CHOICE THEORY

For the most part, our analysis has concentrated on the *normative* theory of government—on the appropriate policies that the government *should follow* to increase the welfare of the population. But economists are not starry-eyed about the government any more than they are about the market. Governments

can make bad decisions or carry out good ideas badly. Indeed, just as there are market failures such as monopoly and pollution, so are there “government failures” in which government interventions lead to waste or redistribute income in an undesirable fashion.

These issues are the domain of **public-choice theory**, which is the branch of economics and political science that studies the way that governments make decisions. Public-choice theory examines the way different voting mechanisms can function and shows that there are no ideal mechanisms to sum up individual preferences into social choices. This approach also analyzes government failures, which arise when state actions fail to improve economic efficiency or when the government redistributes income unfairly. Public-choice theory points to issues such as the short time horizons of elected representatives, the lack of a hard budget constraint, and the role of money in financing elections as sources of government failures. A careful study of government failures is crucial for understanding the limitations of government and ensuring that government programs are not excessively intrusive or wasteful.



The Economics of Politics

Economists focus most of their analysis on the workings of the marketplace. But serious economists have also pondered the government's role in society. Joseph Schumpeter pioneered public-choice theory in *Capitalism, Socialism, and Democracy* (1942), and Kenneth Arrow's Nobel Prize-winning study on social choice brought rigor to this field. The landmark study by Anthony Downs, *An Economic Theory of Democracy* (1957), sketched a powerful new theory which held that politicians choose economic policies in order to be reelected. Downs showed that this theory implies that political parties would move toward the center of the political spectrum because of electoral competition.

Among the most important applications of public-choice theory were those to economic regulation. George Stigler argued that regulatory agencies have been “captured” by the regulated and often served the industries they regulated more than consumers. Studies by James Buchanan and Gordon Tullock in *The Calculus of Consent*

(1959) defended checks and balances and advocated the use of unanimity in political decisions, arguing that unanimous decisions do not coerce anyone. Public-choice economics has been applied to such areas as farm policy and the courts, and it formed the theoretical basis for a proposed constitutional amendment to balance the budget.

B. GOVERNMENT EXPENDITURES

Nowhere can the changes in government's role be seen more clearly than in the area of government spending. Look back at Figure 16-1 on page 304. It shows the share of national output going to government spending, which includes things like purchases of goods, salaries of government workers, social security and other transfers, and interest on the government debt. You can see that government's share rose for most of the twentieth century, with temporary bulges during wartime, but it has leveled off in recent years.

FISCAL FEDERALISM

While we have been referring to government as if it were a single entity, in fact Americans face three levels of government: federal, state, and local. This reflects a division of fiscal responsibilities among the different levels of government—a system known as *fiscal federalism*. The boundaries are not always clear-cut, but in general the federal government directs activities that concern the entire nation—paying for defense, space exploration, and foreign affairs. Local governments educate children, police streets, and remove garbage. States build highways, run university systems, and administer welfare programs.

The total U.S. spending at the different levels of government is shown in Table 16-1. The dominance of the federal role is a comparatively recent phenomenon. Before the twentieth century, local government was by far the most important of the three levels. The federal government did little more than support the military, pay interest on the national

Level of government	Total expenditures, 2007 (\$, billion)	Percent of total
All levels	4,429	100.0
Federal	2,515	56.8
State	857	19.3
Local	1,058	23.9

TABLE 16-1. Federal, State, and Local Government Current Expenditures

In the early days of the Republic, most spending was at the state and local levels. Today, more than half of total government outlays are federal.

Source: U.S. Bureau of Economic Analysis.

debt, and finance a few public works. Most of its tax collection came from liquor and tobacco excises and import tariffs. But two world wars and the rise of the welfare state, with transfer programs such as social security and Medicare, increased spending gradually. The advent of the national income tax in 1913 provided a source of funds that no state or locality could match.

To understand fiscal federalism, economists emphasize that spending decisions should be allocated among the levels of government according to the spillovers from government programs. In general, localities are responsible for *local public goods*, activities whose benefits are largely confined to local residents. Since libraries are used by townspeople and streetlights illuminate city roads, decisions about these goods are appropriately made by local residents. Many federal functions involve *national public goods*, which provide benefits to all the nation's citizens. For example, an AIDS vaccine would benefit people from every state, not just those living near the laboratory where it is discovered. What about global concerns such as protecting the ozone layer or slowing global warming? These are *global public goods* because they transcend the boundaries of individual countries.

An efficient system of fiscal federalism takes into account the way the benefits of public programs spill over political boundaries. The most efficient arrangement is to locate the tax and spending decisions so that the beneficiaries of programs pay the taxes and can weigh the tradeoffs.

Federal Expenditures

Let's look now at the different levels of government. The U.S. government is the world's biggest enterprise. It buys more automobiles and steel, meets a bigger payroll, and handles more money than any other organization anywhere. The numbers involved in federal finance are astronomical—in the billions and trillions of dollars. The federal budget expenditures for 2009 are projected to be \$3107 billion, or \$3.1 trillion; this enormous number amounts to roughly \$27,000 for each American household.

Table 16-2 lists the major categories of federal expenditure for fiscal year 2009. (The federal fiscal year 2009 covers October 1, 2008, through September 30, 2009.)

The most rapidly expanding items in the last three decades have been entitlement programs, which provide benefits or payments to any persons who meet certain eligibility requirements set down by law. The major entitlements are social security (old-age, survivors, and disability insurance), health programs (including Medicare for those over 65 and Medicaid for indigent families), and income-security programs (including subsidies for food and unemployment insurance). In fact, virtually the entire growth in federal spending in recent years can be accounted for by entitlement programs, which increased from 28 percent of the budget in 1960 to 60 percent in 2009.

State and Local Expenditures

Although the battles over the federal budget command the headlines, state and local units provide many of the essential functions in today's economy. Figure 16-3 illustrates the way states and localities spend their money. By far the largest item is education because most of the nation's children are educated in schools financed primarily by local governments. By attempting to equalize the educational resources available to every child, public education helps level out the otherwise great disparities in economic opportunity.

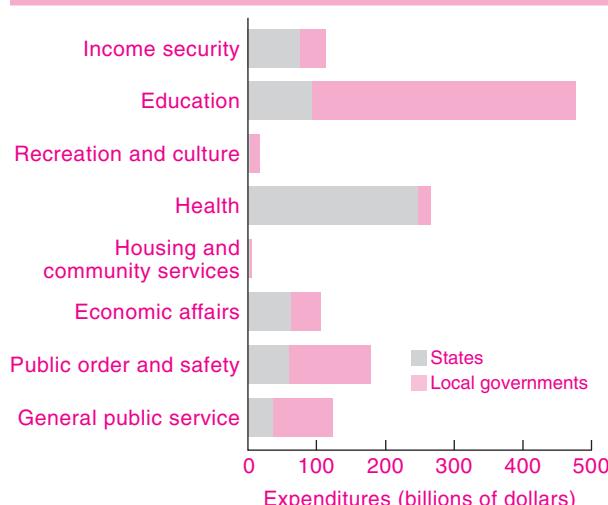
In recent years, the fastest-growing categories of spending for states and localities have been health care and prisons. In the last two decades, the number of prisoners in state prisons tripled, as the United States fought a war on crime partly by using longer prison sentences, especially for drug offenders. At the same time, state and local governments were forced to absorb their share of rising health-care costs.

Federal Expenditures, Fiscal Year 2009		
Description	Expenditures (\$, billion)	Percent of total
Total expenditures	3,107.4	100.0
National defense	675.1	21.7
Social security	649.3	20.9
Medicare	413.3	13.3
Income security	401.7	12.9
Health	299.4	9.6
Net interest	260.2	8.4
Veterans benefits and services	91.9	3.0
Education, training, employment, and social services	88.3	2.8
Transportation	83.9	2.7
Administration of justice	51.1	1.6
International affairs	38.0	1.2
Natural resources and environment	35.5	1.1
General science, space and technology	29.2	0.9
Community and regional development	23.3	0.8
General government	21.5	0.7
Agriculture	19.1	0.6
Commerce and housing credit	4.2	0.1
Energy	3.1	0.1

TABLE 16-2. Federal Spending Is Dominated by Defense and Entitlement Programs

About one-fifth of federal spending is for defense or pensions due to past wars. More than half of spending today is for rapidly growing entitlement programs—income security, social security, and health. Note how small is the traditional cost of government.

Source: Office of Management and Budget, Budget of the U.S. Government, Fiscal Year 2009, available at www.whitehouse.gov/omb/budget/fy2009/hist.html.

**FIGURE 16-3. Distribution of Spending by State and Local Governments, 2006**

State and local programs include providing education, financing hospitals, and maintaining the streets. Education and health take an increasing fraction of state and local spending.

Source: Bureau of Economic Analysis.

CULTURAL AND TECHNOLOGICAL IMPACTS

Government programs have subtle impacts on the country beyond the dollar spending. The federal government has changed the landscape through the interstate highway system. By making automotive travel much faster, this vast network lowered transportation costs, displaced the railroads, and brought goods to every corner of the country. It also helped accelerate urban sprawl and the growth of the suburban culture.

The government has put the United States on the map in many areas of science and technology. Government support gave a powerful start to the electronics industries. The development of the transistor by Bell Labs, for example, was partially funded by the U.S. military, anxious for better radar and communications. Today's computer and airplane industries were boosted in their early years by strong government support. The Internet was

developed by the Department of Defense to create a network that would continue to function in the event of nuclear war.

The government today plays an especially important role in basic science. Of all the basic research in the United States, 85 percent is funded by the government or by nonprofit institutions like universities. Often, if you follow a successful invention upstream to its source, you will find that government subsidized the inventor's education and supported basic university research. Economic studies indicate that these funds were well spent, moreover, for the social rates of return to research and development exceed the returns on investments in most other areas.

C. ECONOMIC ASPECTS OF TAXATION

Taxes are what we pay for a civilized society.

Justice Oliver Wendell Holmes

Governments must pay for their programs. The funds come mainly from taxes, and any shortfall is a deficit that is borrowed from the public.

But in economics we always need to pierce the veil of monetary flows to understand the flow of real resources. Behind the dollar flows of taxes, what the government really needs is the economy's scarce land, labor, and capital. When a nation goes to war, people argue about how to finance the military spending. But in reality, what really happens is that people are diverted from their civilian jobs, airplanes transport troops rather than tourists, and oil goes to airplanes rather than cars. When the government gives out a grant for biotechnology research, its decision really means that a piece of land that might have been used for an office building is now being used for a laboratory.

In taxing, government is in reality deciding how to draw the required resources from the nation's households and businesses for public purposes. The money raised through taxation is the vehicle by which real resources are transferred from private goods to collective goods.

PRINCIPLES OF TAXATION

Benefit vs. Ability-to-Pay Principles

Once the government has decided to collect some amount of taxes, it has many possible taxes available to it. It can tax income, tax profits, or tax sales. It can tax the rich or tax the poor, tax the old or tax the young. Are there any guidelines that can help construct a fair and efficient tax system?

Indeed there are. Economists and political philosophers have proposed two major principles for organizing a tax system:

- The **benefit principle**, which holds that individuals should be taxed in proportion to the benefit they receive from government programs. Just as people pay private goods like dollars in proportion to their consumption of private goods like bread, a person's taxes should be related to his or her use of collective goods like public roads or parks.
- The **ability-to-pay principle**, which states that the amount of taxes people pay should relate to their income or wealth. The higher the wealth or income, the higher the taxes. Usually tax systems organized on the ability-to-pay principle are also *redistributive*, meaning that they raise funds from higher-income people to increase the incomes and consumption of poorer groups.

For instance, if the construction of a new bridge is funded by tolls on the bridge, that's a reflection of the benefit principle, since you pay for the bridge only if you use it. But if the bridge were funded out of income-tax collections, that would be an example of the ability-to-pay principle.

Horizontal and Vertical Equity

Whether they are organized along benefit or ability-to-pay lines, most modern tax systems attempt to incorporate modern views about fairness or equity. One important principle is that of **horizontal equity**, which states that those who are essentially equal should be taxed equally.

The notion of equal treatment of equals has deep roots in Western political philosophy. If you and I are alike in every way except the color of our eyes, all principles of taxation would hold that we should pay equal taxes. In the case of benefit taxation, if we receive exactly the same services from the highways or parks, the principle of horizontal equity states that

we should therefore pay equal taxes. Or if a tax system follows the ability-to-pay approach, horizontal equity dictates that people who have equal incomes should pay the same taxes.

A more controversial principle is **vertical equity**, which concerns the tax treatment of people with different levels of income. Abstract philosophical principles provide little guidance in resolving the issues of fairness here. Imagine that A and B are alike in every respect except that B has 10 times the property and income of A. Does that mean that B should pay the same absolute tax dollars as A for government services such as police protection? Or that B should pay the same percentage of income in taxes? Or, since the police spend more time protecting the property of well-to-do B, is it perhaps fair for B to pay a larger fraction of income in taxes?

Be warned that general and abstract principles cannot determine the tax structure for a nation. When Ronald Reagan campaigned for lower taxes, he did so because he thought high taxes were unfair to those who had worked hard and saved for the future. A decade later, Bill Clinton said, “We now have real fairness in the tax code with over 80 percent of the new tax burden being borne by those who make over \$200,000 a year.” What looks fair to the goose seems foul to the gander.

Horizontal equity is the principle that equals should be treated equally. Vertical equity holds that people in unequal circumstances should be treated unequally and fairly, but there is no consensus on exactly how vertical equity should be applied.

Pragmatic Compromises in Taxation

How have societies resolved these thorny philosophical questions? Governments have generally adopted pragmatic solutions that are only partially based on benefit and ability-to-pay approaches. Political representatives know that taxes are highly unpopular. After all, the cry of “taxation without representation” helped launch the American Revolution. Modern tax systems are an uneasy compromise between lofty principles and political pragmatism. As the canny French finance minister Colbert wrote three centuries ago, “Raising taxes is like plucking a goose: you want to get the maximum number of feathers with the minimum amount of hiss.”

What practices have emerged? Often, public services primarily benefit recognizable groups, and those groups have no claim for special treatment by virtue of their average incomes or other characteristics. In such cases, modern governments generally rely on benefit taxes.

Thus, local roads are usually paid for by local residents. “User fees” are charged for water and sewage treatment, which are treated like private goods. Taxes collected on gasoline may be devoted (or “earmarked”) to roads.

Progressive and Regressive Taxes. Benefit taxes are a declining fraction of government revenues. Today, advanced countries rely heavily on **progressive income taxes**. With progressive taxes, a family with \$50,000 of income is taxed more than one with \$20,000 of income. Not only does the higher-income family pay a larger income tax, but it in fact pays a higher fraction of its income.

This progressive tax is in contrast to a strictly **proportional tax**, in which all taxpayers pay exactly the same proportion of income. A **regressive tax** takes a larger fraction of income in taxes from poor families than it does from rich families.

A tax is called *proportional, progressive, or regressive* depending on whether it takes from high-income people the same fraction of income, a larger fraction of income, or a smaller fraction of income than it takes from low-income people.

The different kinds of taxes are illustrated in Figure 16-4. What are some examples? A personal income tax that is graduated to take more and more out of each extra dollar of income is progressive. Economists have found, by contrast, that the cigarette tax is regressive. The reason is that the number of cigarettes purchased rises less rapidly than income. For example, some studies have determined that the income elasticity of cigarette use is around 0.6. This means that a 10 percent increase in income leads to a 6 percent increase in expenditures on cigarettes, and also to a 6 percent increase in cigarette taxes. Thus, high-income groups pay a smaller fraction of their income in cigarette taxes than do low-income groups.

Direct and Indirect Taxes. Taxes are classified as direct or indirect. **Indirect taxes** are ones that are levied on goods and services and thus only “indirectly”

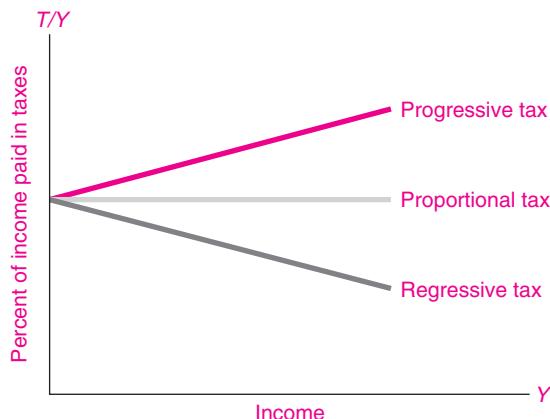


FIGURE 16-4. Progressive, Proportional, and Regressive Taxes

Taxes are progressive if they take a larger fraction of income as income rises; proportional if they are a constant fraction of income; and regressive if they place a larger relative burden on low-income families than on high-income families.

on individuals. Examples are excise and sales taxes, cigarette and gasoline taxes, tariffs on imports, and property taxes. By contrast, **direct taxes** are levied directly upon individuals or firms. Examples of direct taxes are personal income taxes, social security or other payroll taxes, and inheritance and gift taxes. Direct taxes have the advantage of being easier to tailor to fit personal circumstances, such as size of family, income, age, and more generally the ability to pay. By contrast, indirect taxes have the advantage of being easier to collect, since they can be levied at the retail or wholesale level.

FEDERAL TAXATION

Let us now try to understand the principles by which the federal system of taxation is organized. Table 16-3 provides an overview of the major taxes collected by the federal government and shows whether they are progressive, proportional, or regressive.

The Individual Income Tax

Our discussion begins with the individual income tax, which is the most complex part of the tax system. The income tax is a direct tax, and it is the tax which most clearly reflects the ability-to-pay principle.

Federal Tax Receipts, Fiscal Year 2009	
	Receipts (% of total)
Progressive:	
Individual income taxes	46.6
Estate and gift taxes	1.0
Corporate income taxes	12.6
Proportional:	
Payroll taxes	35.2
Regressive:	
Excise taxes	2.6
Customs duties	1.1
Other taxes and receipts	1.0
Total	100.0

TABLE 16-3. Income and Payroll Taxes Are the Main Federal Revenue Sources

Progressive taxes are still the leading source of federal revenues, but proportional payroll taxes are closing fast. Regressive consumption taxes have declined sharply at the federal level.

Source: See Table 16-2.

The individual income tax arrived late in our nation's history. The Constitution forbade any direct tax that was not apportioned among the states according to population. This was changed in 1913, when the Sixteenth Amendment to the Constitution provided that "Congress shall have power to lay and collect taxes on income, from whatever source derived."

How does the federal income tax work? The principle is simple, although the forms are complicated. You start by calculating your income; you next subtract certain expenses, deductions, and exemptions to obtain taxable income. You then calculate your taxes on the basis of your taxable income.

Suppose you have just graduated from college and take a job in California with a salary of \$60,000 in 2009. Table 16-4 shows a calculation of the total direct tax payments that you should expect. It will be worthwhile going line by line to understand the different items.

Line 1 begins with your salary. The first set of taxes is social insurance taxes. We will postpone our

1 Annual salary	\$60,000
2 Social security taxes:	
3 Pension	3,720
4 Medicare	870
5 Federal adjusted gross income = (1)	60,000
6 Less:	
7 Personal exemption	3,500
8 Standard deduction	5,450
9 Federal taxable income = (5) - (7) - (8)	51,050
10 Income tax:	
11 Federal	9,106
12 State (California)	2,672
13 Total taxes = (3) + (4) + (11) + (12)	16,368
14 Income after tax = (1) - (13)	43,632
15 Tax rate	
16 Average = (13)/(1)	27.3%
17 Marginal*	42.0%

*Marginal tax rate is the additional total taxes per additional dollar of income. This would be calculated by repeating all the lines for an additional \$1000 of income and then dividing the extra number of dollars of taxes by 1000.

TABLE 16-4. Calculation of Individual Income Taxes, 2009

The table shows an illustrative calculation of total taxes for a single worker living in California in 2009. The worker has a total salary of \$60,000. Social security taxes are for future social security benefits and pay health benefits for current retired workers. Income taxes are levied by the federal government and most states.

The average tax rate is 27.3 percent. Economists focus on the marginal tax rate, which is the additional tax per additional dollar of income. For our worker, the marginal tax rate is calculated to be 42 percent.

Source: Internal Revenue Service and State of California (preliminary tax tables).

discussion of these to the next section. Line 5 shows your *adjusted gross income*—that is, total wages, interest, dividends, and other income earned. If you were single, you would have a *personal exemption* of \$3500. If you do not own a house, you are likely to take the *standard deduction* of \$5450. Subtracting both of these yields your *federal taxable income* of \$51,050.

Next, you go to the tax tables. These currently show a tax of \$9106 on this income. You would also have taxes due to the state, \$2672 in this case.

Adding up all the taxes, you find you owe \$16,368. This represents 27.3 percent of your income. This is called the **effective or average tax rate**, which is equal to total taxes divided by total income.

The last row introduces an important new concept. The **marginal tax rate** is the extra tax that is paid per dollar of additional income. We have met the term “marginal” before, and it always means “extra.” If you were to earn an additional \$1000 of

income, you would pay an additional \$420 in taxes. This means that your marginal tax rate is \$420/\$1000, or 42 percent. The marginal tax rate is a critical tool for tax analysis because people and companies tend to respond to their marginal tax rates, not their average tax rates. Moreover, when marginal tax rates are extremely high, incentives to work are dulled and effort may significantly decrease.

The marginal tax rate is a central concept of tax analysis. It refers to the extra tax paid per dollar of extra income and is particularly important for understanding the incentive effects of taxation.

Figure 16-5 shows the estimated marginal tax rate for households with incomes up to \$100,000. Low-income households have a “negative income tax,” because they receive an earned-income tax credit.

The notion of marginal tax rates is extremely important in modern economics. Remember the

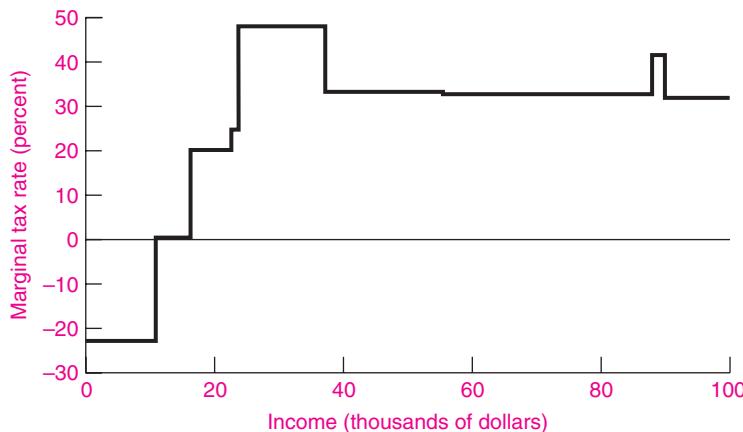


FIGURE 16-5. Marginal Tax Rate of U.S. Households by Income Category, 2005

The marginal tax rate is the extra tax that is paid per dollar of additional income. The figure shows the estimated marginal tax rates of households in 2005. These include social insurance as well as federal and average state taxes. Because of the earned-income tax credit, low-income workers get a tax rebate—this is a “negative income tax” on wages. Note that the marginal tax rates in this figure differ from those in Table 16-4 because California has relatively high taxes and because the CBO uses different assumptions about exemptions and deductions.

Source: Congressional Budget Office, *Effective Marginal Tax Rates on Labor Income*, November 2005, available at www.cbo.gov.

marginal principle. People should be concerned only with the extra costs or benefits that occur. They should “let bygones be bygones.” Under this principle, the major effect of any tax on incentives comes from the marginal tax rate.



Radical Tax Reform: The Flat Tax

The individual income tax is a powerful engine for raising revenues. But it has become enormously complex over the century since its introduction. Moreover, it is full of loopholes or “tax preferences” that provide benefits to particular forms of income or expenditure and even to individual groups of taxpayers. For example, expenditures on mortgage interest and medical care are deductible from income—they are, in effect, subsidized spending.

Economists have campaigned tirelessly for a more streamlined tax system—one that broadens the tax base, and thus raises revenues by eliminating unnecessary tax breaks, and can therefore *lower marginal tax rates*. One of the most radical and innovative proposals for fundamental tax reform is the *flat tax*, which was developed in detail by

Stanford’s Robert Hall and Alvin Rabushka.¹ Their proposal incorporates the following major features (see question 9 at the end of this chapter for an example):

- It taxes consumption rather than income. As we will discuss later in this chapter, taxing consumption serves to increase the incentive to save and can help boost the declining national savings rate.
- It integrates the corporate income tax with the individual income tax. This removes one of the major distortions in the U.S. tax code.
- It eliminates virtually all loopholes and tax preferences. Gone are subsidies for medical care, owner-occupied homes, and charitable contributions.
- It provides a basic exemption of around \$20,000 per family and then imposes a constant marginal tax rate of 19 percent above that level.

The economic effects of a flat tax would be far-reaching. Heavily taxed entities such as corporations would find their taxes lowered and would experience a major capital

¹ *The Flat Tax*, rev. ed. Hoover Institute Press, Palo Alto, Calif., 2007.

gain. High-income wage earners would find their taxes cut in half. At the same time, the amount of owner-occupied housing and medical expenditures would shrink and charitable giving would drop sharply.

Hall and Rabushka emphasize above all the importance of reducing the marginal tax rates. They argue that the flat tax would “give an enormous boost to the U.S. economy by dramatically improving incentives to work, save, invest, and take entrepreneurial risks. The flat tax would save taxpayers hundreds of billions in direct and indirect compliance costs.”

The plan’s critics point out that it would lead to a major redistribution of income to high-income people at the expense of low- and middle-income households. The losers will question whether the rich, whose share has risen dramatically over the last three decades, deserve yet another windfall. We see here yet another example of the tradeoff between fairness and efficiency that runs through many of the most controversial economic policy issues.

Social Insurance Taxes

Virtually all industries now come under the Social Security Act. Workers receive retirement benefits that depend on their earnings history and past social security taxes. The social insurance program also funds a disability program and health insurance for the poor and elderly.

To pay for these benefits, employees and employers are charged a *payroll tax*. As shown in Table 16-4, in 2008, this consisted of a total of 15.3 percent of all wage income below a ceiling of \$102,000 a year per person, along with a payroll tax of 2.9 percent of annual wage income above \$102,000. The tax is split equally between employer and employee.

Table 16-3 shows the payroll tax as a proportional tax because it taxes a fixed fraction of employment earnings. The tax incidence is more complicated, however, because the payroll tax includes only labor earnings (which makes it regressive) and finances retirement most generously for low-income people (which makes it progressive).

Corporation Taxes

The federal government collects a wide variety of other taxes, some of which are shown in Table 16-3. The *corporate income tax* is a tax on the profits of corporations.

The corporation income tax has been heavily criticized by some economists. Critics oppose the tax, arguing that corporations are but legal fictions and should not be taxed. By taxing first corporate profits and then the dividends paid by corporations and received by individuals, the government subjects corporations to double taxation.

Consumption Taxes

While the United States relies heavily on income taxes, a radically different approach is consumption taxes, which are taxes on purchases of goods and services rather than on income. The rationale is that people should be penalized for what they *use* rather than what they *produce*. Sales taxes are the most familiar example of consumption taxes. The United States has no national sales tax, although there are a number of *federal excise taxes* on specific commodities such as cigarettes, alcohol, and gasoline. Sales and excise taxes are generally regressive because they consume a larger fraction of the income of poor families than of high-income families.

Many have argued that the United States should rely more heavily on sales or consumption taxes. One tax, widely used outside the United States, is the *value-added tax*, or VAT. The VAT is like a sales tax, but it collects taxes at each stage of production. Thus, if a VAT were levied on bread, it would be collected from the farmer for wheat production, from the miller for flour production, from the baker at the dough stage, and from the grocer at the delivered-loaf stage.

The advocates of consumption taxes argue that the country is currently saving and investing less than is necessary for future needs and that by substituting consumption taxes for income taxes, the national savings rate would increase. Critics of consumption taxes respond that such a change is undesirable because sales taxes are more regressive than today's income tax. The *flat tax*, discussed earlier, is actually equivalent to a highly simplified system of personal consumption taxation (see question 9 below).

STATE AND LOCAL TAXES

Under the U.S. system of fiscal federalism, state and local governments rely on a very different set of taxes than does the federal government. Figure 16-6 illustrates the main sources of funds that finance state and local expenditures.

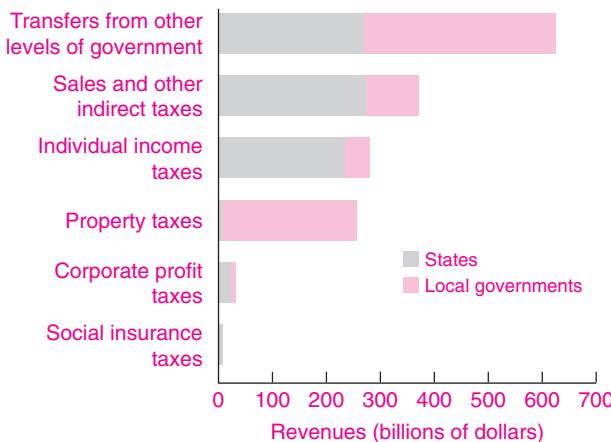


FIGURE 16-6. States and Localities Rely on Transfers and Indirect Taxes

Cities rely heavily on property taxes because houses and land cannot easily flee to the suburbs to avoid a city's tax. States get most revenues from sales and income taxes.

Source: Bureau of Economic Analysis.

Property Tax

The *property tax* is levied primarily on real estate—land and buildings. Each locality sets an annual tax rate which is levied on the assessed value of the land and structures. In many localities, the assessed value may be much smaller than the true market value. The property tax accounts for about 30 percent of the total revenues of state and local finance. Figure 16-6 shows that localities are the main recipient of property taxes.

Because about one-fourth of property values are from land, the property tax has elements of a capital tax and elements of a Henry George-type land tax. Economists believe that the land component of the property tax has little distortion, while the capital component will drive investment from high-tax central cities out to the low-tax suburbs.

Other Taxes

Most other state and local taxes are closely related to the analogous federal taxes. States get most of their revenues from *general sales taxes* on goods and services. Each purchase at the department store or restaurant incurs a percentage tax (food and other necessities are exempt in some states). States tax the net income of corporations. Forty-three states imitate the federal government, on a much smaller scale, by taxing individuals according to the size of their incomes.

There are other miscellaneous revenues. Many states levy “highway user taxes” on gasoline. A growing source of revenue is lotteries and legalized gambling, in which the states benefit from encouraging people to impoverish themselves.

EFFICIENCY AND FAIRNESS IN THE TAX SYSTEM

The Goal of Efficient Taxation

In recent years, economists have focused increasingly on the efficiency of different tax systems. The first point to recall here is that efficiency depends primarily on the marginal tax rates faced by taxpayers. Look back at Figure 16-5 to recall how the marginal tax rates differ across income groups.

Taxes on Labor Income. How do high marginal tax rates affect economic behavior? In the area of labor supply, the impacts are mixed. As we saw in Chapter 13, the impact of tax rates on hours worked is unclear because the income and substitution effects of wage changes work in opposite directions. As a result of progressive taxes, some people may choose more leisure over more work. Other people may work harder in order to make their millions. Many high-income doctors, artists, celebrities, and business executives, who enjoy their jobs and the sense of power or accomplishment that they bring, will work as hard for \$800,000 after tax as for \$1,000,000 after tax.

Figure 16-7 shows how an increase in the tax rate on labor will affect labor supply; note the paradox that hours worked may actually decline after a tax-rate cut if the labor supply curve is backward-bending.

Taxes on Capital Income. In the area of saving and investment, taxes are likely to have major effects on amounts supplied and efficiency. When taxes are high in one sector, resources will flow into more lightly taxed areas. For example, because corporate profits are double-taxed, people's savings will flow out of the corporate sector and into lightly taxed sectors. If risky investments are taxed unfavorably, investors may prefer safer investments.

Impacts of Globalization. With increased openness of economies, countries need to ensure that mobile factors of production like capital or highly skilled workers are not lured away to low-tax countries. This

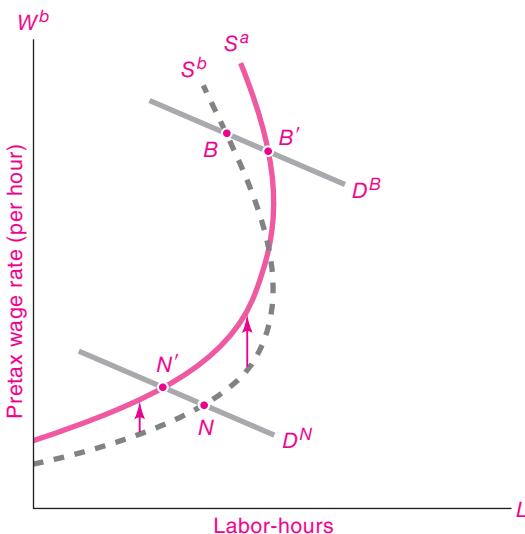


FIGURE 16-7. Response of Work to Taxes Depends on Shape of Supply Curve

Supply and demand plots labor supplied against pretax wage. Before-tax supply curve of labor (S^b) shifts vertically upward to after-tax supply (S^a) after imposition of a 25 percent income tax on labor earnings. If demand for labor intersects supply in the normal region at bottom, we see an expected decline in labor supplied from N to N' . If the labor supply is backward-bending, as at top, the labor supplied actually rises with the tax increase, going from B to B' .

concern is particularly important for company taxes, for companies can easily move their headquarters to some island tax haven.

Efficiency vs. Fairness

Economists have long been concerned with the impact of taxes on economic efficiency. Recall from Chapter 14 that Henry George argued that a tax on land will have little impact on efficiency because the supply of land is completely inelastic. The modern theory of efficient taxation puts forth the *Ramsey tax rule*, which states that the government should levy the heaviest taxes on those inputs and outputs that are most price-inelastic in supply or demand.² The rationale for the Ramsey tax rule is that if a commodity is very price-inelastic in supply or demand, a tax on

the commodity will have little impact upon consumption and production. In some circumstances, Ramsey taxes may constitute a way of raising revenues with a minimum loss of economic efficiency.

But economies and politics do not run on efficiency alone. While stiff taxation of land rents or food might be efficient, many would think them unfair. A sober reminder of the dilemma was the proposal to introduce a poll tax in Britain in 1990. A *poll tax* is a *lump-sum tax*, or a fixed tax per person. The advantage of this tax is that, like a land tax, it would induce no inefficiencies. After all, people are unlikely to decamp to Russia or commit hari-kari to avoid the tax, so the economic distortions would arguably be minimal.

Alas, the British government underestimated the extent to which the populace felt this tax to be unfair. The poll tax is highly regressive because it places a much higher proportional burden on low-income people than on high-income people. Criticism of the poll tax played a key role in bringing down the Thatcher government after 11 years in power. This illustrates clearly the difficult choice between efficiency and fairness in taxes and other areas of economic policy.



Taxing “Bads” rather than “Goods”: Green Taxes

While economists have rarely advocated poll taxes, they have favored an approach wherein the tax system would weigh more heavily on “bads” than on “goods.” The main source of inefficiency is that taxes generally tax “goods”—economic activities like working, investing in capital, saving, or taking risk—and thereby discourage these activities. An alternative approach is to tax “bads.” Traditional taxes on goods include “sin taxes”: taxes on alcohol, cigarettes, and other substances that have harmful health effects.

A new approach to taxation is to tax pollution and other undesirable externalities; such taxes are called green taxes because they are designed to help the environment as well as to raise revenues. Say that the nation decides to help slow global warming by levying a “carbon tax,” which is a tax on carbon-dioxide emissions from power plants and other sources. By standard economic reasoning we know that the tax will lead firms to lower their carbon-dioxide emissions, thereby improving the environment. In addition, this green tax will provide revenues, which the government can use either to finance its activities or to reduce

² Recall Chapter 14’s discussion of Henry George’s single tax and the extension to efficient or Ramsey taxes.

tax rates on beneficial activities like working or saving. So green taxes are doubly effective: the state gets revenue, and the environment is improved because the taxes discourage harmful externalities.

FINAL WORD

Our introductory survey of government's role in the economy is a sobering reminder of the responsibilities and shortcomings of collective action. On the one hand, governments must defend their borders, stabilize their economies, protect the public health, and

regulate pollution. On the other hand, policies often reflect primarily the attempt to redistribute income from consumers to politically powerful interest groups.

Does this mean we should abandon the visible hand of government for the invisible hand of markets? Economics cannot answer such deep political questions. But economics can examine the strengths and weaknesses of both collective and market choices, and point to mechanisms (such as green taxes or subsidies to research and development) by which a mended invisible hand may be more efficient and fair than the extremes of either pure laissez-faire or unbridled bureaucratic rulemaking.



SUMMARY

A. Government Control of the Economy

1. The economic role of government has increased sharply over the last century. The government influences and controls private economic activity by using taxes, expenditures, and direct regulation.
2. A modern welfare state performs four economic functions: (a) It remedies market failures; (b) it redistributes income and resources; (c) it establishes fiscal and monetary policies to stabilize the business cycle and promote long-term economic growth; and (d) it manages international economic affairs.
3. Public-choice theory analyzes how governments actually behave. Just as the invisible hand can break down, so there are government failures, in which government interventions lead to waste or redistribute income in an undesirable fashion.

B. Government Expenditures

4. The American system of public finance is one of fiscal federalism. The federal government concentrates its spending on issues of national concern—on national public goods like defense and space exploration. States and localities generally focus on local public goods—those whose benefits are largely confined within state or city boundaries.
5. Government spending and taxation today take approximately one-third of total national output. Of this total, about 55 percent is spent at the federal level, and the balance is divided between state and local governments. Only a small fraction of government outlays is devoted to traditional functions like police and the courts.

C. Economic Aspects of Taxation

6. Notions of “benefits” and “ability to pay” are two principal theories of taxation. A tax is progressive, proportional, or regressive as it takes a larger, equal, or smaller fraction of income from rich families than it does from poor families. Direct and progressive taxes on incomes are in contrast to indirect and regressive sales and excise taxes.
7. More than half of federal revenues come from personal and corporate income taxes. The rest comes from taxes on payrolls or consumption goods. Local governments raise most of their revenue from property taxes, while sales taxes are most important for states.
8. The individual income tax is levied on “income from whatever source derived,” less certain exemptions and deductions. The marginal tax rate, denoting the fraction paid in taxes for every dollar of additional income, is the key to determining the impact of taxes on incentives to work and save.
9. The fastest-growing federal tax is the payroll tax, used to finance social security. This is an “earmarked” levy, with funds going to provide public pensions and health and disability benefits. Because there are visible benefits at the end of the stream of payments, the payroll tax has elements of a benefit tax.
10. Economists point to the Ramsey tax rule, which emphasizes that efficiency will be promoted when taxes are levied more heavily on those activities that are relatively price-inelastic. A new approach is green taxes, which levy fees on environmental externalities, reducing harmful activities while raising revenues that would otherwise be imposed on goods or productive inputs. But in all taxes, equity and political acceptability are severe constraints.

CONCEPTS FOR REVIEW

Functions of Government

three tools of government economic control:
 taxes
 expenditures
 regulation
 market failures vs. government failures
 public-choice theory

four functions of government:
 efficiency
 distribution
 stabilization
 international representation

Government Expenditures and Taxation

fiscal federalism and local vs. national public goods

economic impact of government spending
 benefit and ability-to-pay principles
 horizontal and vertical equity
 direct and indirect taxes
 entitlement programs
 progressive, proportional, and regressive taxes
 Ramsey and green taxes

FURTHER READING AND INTERNET WEBSITES

Further Reading

An excellent review of tax issues is contained in the symposium on tax reform in *Journal of Economic Perspectives*, Summer 1987. The classic study of the flat tax referred to in the text is also online, at www.hoover.org/publications/books/3602666.html.

Websites

Data on government budget and tax trends can be found at government sites. For example, overall trends are presented by the Bureau of Economic Analysis at

www.bea.gov. Budget information for the federal government comes from the Office of Management and Budget at www.whitehouse.gov/omb.

The Internal Revenue Service (IRS) has a lively site with a plethora of tax statistics at www.irs.gov and www.irs.gov/tax-stats/index.html.

Two organizations which study taxation and have good websites are the National Tax Association at www.ntanet.org and the Brookings Institution at www.brookings.org. Policy papers by a British research institute that focuses on social security and taxation can be found at www.ifs.org.uk.

QUESTIONS FOR DISCUSSION

1. Recall Justice Oliver Wendell Holmes's statement, "Taxes are what we pay for a civilized society." Interpret this statement, remembering that in economics we always need to pierce the veil of monetary flows to understand the flow of real resources.
2. In considering whether you want a pure laissez-faire economy or government regulation, discuss whether there should be government controls over prostitution, addictive drugs, heart transplants, assault weapons, and alcohol. Discuss the relative advantages of high taxes and prohibition for such goods (recall the discussion of drug prohibition in Chapter 5).
3. Critics of the U.S. tax system argue that it harms incentives to work, save, and innovate and therefore reduces long-run economic growth. Can you see why "green

"taxes" might promote economic efficiency and economic growth? Consider, for example, taxes on sulfur or carbon-dioxide emissions or on leaky oil tankers. Construct a list of taxes that you think would increase efficiency, and compare their effects with the effects of taxes on labor or capital income.

4. Tax economists often speak of lump-sum taxes, which are levied on individuals without regard to their economic activity. Lump-sum taxes are efficient because they impose zero marginal tax rates on all inputs and outputs.

Assume that the government imposes a lump-sum tax of \$200 on each individual. Show the effect of this on the supply and demand for labor in a graph. Does the marginal revenue product of labor still equal the wage in equilibrium?

In a lifetime framework, the dynamic equivalent of a lump-sum tax is an “endowment tax,” which would tax individuals on the basis of their potential labor incomes. Would you favor such a change? Describe some of the difficulties in implementing an endowment tax.

5. Make a list of different federal taxes in order of their progressiveness. If the federal government were to trade in income taxes for consumption or sales taxes, what would be the effect in terms of overall progressiveness of the tax system?
6. Some public goods are local, spilling out to residents of small areas; others are national, benefiting an entire nation; some are global, affecting all nations. A private good is one whose spillover is negligible. Give some examples of purely private goods and of local, national, and global public goods or externalities. For each, indicate the level of government that could design relevant policies most efficiently, and suggest one or two appropriate government actions that could solve the externality.
7. Recall from our discussion of tax incidence that the incidence of a tax refers to its ultimate economic burden and to its total effect on prices, outputs, and other economic magnitudes. Below are some incidence questions that can be answered using supply and demand. Use graphs to explain your answers.
 - a. In the 1993 Budget Act, Congress raised federal gasoline taxes by 4.3 cents a gallon. Assuming the wholesale price of gasoline is determined in world markets, what is the relative impact of the tax on American producers and consumers?
 - b. Social insurance taxes are generally levied on labor earnings. What is their incidence if labor supply is perfectly inelastic? If labor supply is backward-bending?
 - c. Assume that firms must earn a given post-tax rate of return on investment, where the return is determined in world capital markets. What is the incidence of a tax on corporate income in a small open economy?
8. An interesting question involves the *Laffer curve*, named for California economist and sometime senatorial candidate Arthur Laffer. In Figure 16-8, the Laffer curve shows how revenues rise as *tax rates* are increased, reach a maximum at point *L*, and then decline to zero at a 100 percent tax rate as activity is completely discouraged. The exact shape of the Laffer curve for different taxes is highly controversial.

A common mistake in discussing taxes is the post hoc fallacy (see Chapter 1’s discussion of this). Proponents of lower taxes often invoke the Laffer curve in their arguments. They point to tax cuts of the 1960s

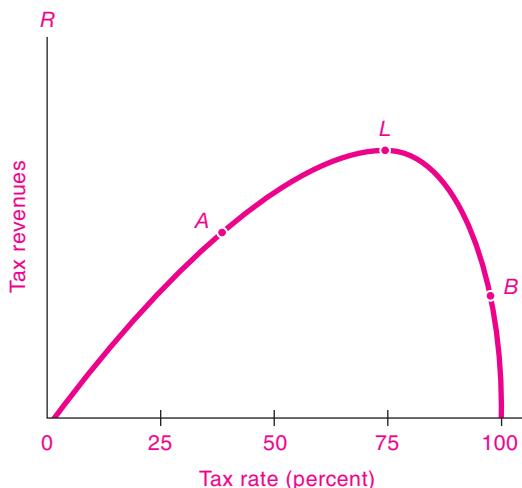


FIGURE 16-8. The Laffer Curve

to suggest that the economy is to the right of the peak of Mt. Laffer, say, at *B*. They say, in effect, “After the Kennedy-Johnson tax cuts of 1964, federal revenues actually rose from \$110 billion in 1963 to \$133 billion in 1966. Therefore, cutting taxes raises revenues.” Explain why this does not prove that the economy was to the right of *L*. Further explain why this is an example of the post hoc fallacy. Give a correct analysis.

9. Under the flat tax, all personal and corporate income is taxed only once at a low fixed rate. Table 16-5 shows how such a flat tax might work. Compare the average and marginal tax rates of the flat tax with the tax schedule shown in Table 16-4 in the text. List advantages and disadvantages of both. Which is more progressive?

(1) Adjusted gross income (\$)	(2) Deductions and exemptions (\$)	(3) Taxable income (\$)	(4) Individual income tax (\$)
5,000	20,000	0	0
10,000	20,000	0	0
20,000	20,000	0	0
50,000	20,000	30,000	6,000
100,000	20,000	80,000	16,000
1,000,000	20,000	980,000	196,000

TABLE 16-5.

Efficiency vs. Equality: The Big Tradeoff

17



[The conflict] between equality and efficiency [is] our biggest socioeconomic tradeoff, and it plagues us in dozens of dimensions of social policy. We can't have our cake of market efficiency and share it equally.

Arthur Okun (1975)

About a century ago, many Western governments began to intervene in the marketplace and introduce a social safety net as a bulwark against socialist pressures—this new conception of society was called the “welfare state.” Attitudes toward the welfare state evolved gradually into the mixed market economy found today in the democracies of Europe and North America. In these countries, the market is responsible for production and pricing of most goods and services, while governments manage the economy and provide a safety net for the poor, unemployed, and aged.

One of the most controversial aspects of government policy involves policies toward the poor. Should families have guaranteed incomes? Or perhaps just minimum levels of food, shelter, and health care? Should taxation be progressive, redistributing incomes from the rich to the poor? Or should taxation be aimed primarily at promoting economic growth and efficiency?

Surprisingly, these questions have been just as contentious as societies have become richer. You might think that as a country becomes more prosperous, it would devote a larger share of its income to programs helping the needy at home and abroad. This has not always proved to be the case. As tax burdens have risen over the last half-century, tax revolts have sparked

reductions in tax rates. People are also increasingly aware that attempts to equalize incomes can harm incentives and efficiency. Today, people ask: How much of the economic pie must be sacrificed in order to divide it more equally? How should we redesign income-support programs to retain the objective of reducing want and inequality without bankrupting the nation?

The purpose of this chapter is to examine the distribution of income along with the dilemmas of policies designed to reduce inequality. These issues are among the most controversial economic questions of today. Remember the first chapter suggestion that economics best serves the public interest in using cool heads to inform warm hearts. This chapter surveys the trends in inequality and the relative merits of different approaches and indicates how cool-headed economic analysis can help promote both fairness and continued growth of the mixed economy.

A. THE SOURCES OF INEQUALITY

To measure the inequality of control over economic resources, we need to concern ourselves with both income and wealth differences. Recall that by **personal**

income we mean the total receipts or cash earned by a person or household during a given time period (usually a year). The major components of personal income are labor earnings, property income (such as rents, interest, and dividends), and government transfer payments. **Disposable personal income** consists of personal income less any taxes paid. **Wealth** or “net worth” consists of the dollar value of financial and tangible assets minus the amount of money owed to banks and other creditors. You can refresh your memory about the major sources of income and wealth by reviewing Tables 12-1 and 12-2 (look at pages 230 and 232).

THE DISTRIBUTION OF INCOME AND WEALTH

Statistics show that in 2006 the median income of American families was \$48,200—this means that half of all families received less than this figure while half received more. This number concerns the *distribution of income*, which shows the variability or dispersion of incomes. To understand the income distribution, consider the following experiment: Suppose one person from each household writes down the yearly income of his or her household on an index card. We can then sort these cards into *income classes*. Some of the cards go into the lowest 20 percent, the group with an average income of \$11,551. Some go into the next class. A few go into the top 5 percent of households, those with an average income of \$362,514.

The actual income distribution of American households in 2006 is shown in Table 17-1. Column (1) shows the different income-class fifths, or quintiles, plus the top 5 percent of households. Column (2) shows the average income in each income class. Column (3) shows the percentage of the households in each income class, while column (4) shows the percentage of total national income that goes to the households in an income class.

Table 17-1 enables us to see at a glance the wide range of incomes in the U.S. economy. Half of the population makes less than \$50,000 per year. As you move up the distribution, the number of people gets smaller and smaller. If we made an income pyramid out of building blocks, with each layer portraying \$500 of income, the peak would be far higher than Mount Everest, but most people would be within a few feet of the ground.

How to Measure Inequality among Income Classes

How can we measure the degree of income inequality? At one pole, if incomes were absolutely equally distributed, there would be no difference between the lowest 20 percent and the highest 20 percent of the population: each quintile would receive exactly 20 percent of the nation’s income. That’s what absolute equality means.

The reality is very different. In 2006, the lowest fifth, with 20 percent of the households, earned less than 4 percent of the total income. Meanwhile the

(1) Income class of households	(2) Average	(3) Percentage of all households in this class	(4) Percentage of total income received by households in this class
Lowest fifth	\$11,551	20	3.4
Second fifth	\$29,442	20	8.7
Third fifth	\$49,968	20	14.8
Fourth fifth	\$79,111	20	23.4
Highest fifth	\$169,971	20	49.7
Top 5 percent	\$362,514	5	21.2

TABLE 17-1. Distribution of Money Incomes of American Households, 2006

How was total income distributed among households in 2006? We group households into the fifth (or quintile) with the lowest income, the fifth with the second-lowest income, and so on.

Source: U.S. Bureau of the Census, Current Population Report, *Income, Poverty, and Health Insurance Coverage in the United States: 2007*, available at www.census.gov/hhes/www/income/income.html.

situation is reversed for the top 5 percent of households, which get 21 percent of the income.

We can show the degree of inequality in a diagram known as the **Lorenz curve**, a widely used device for analyzing income and wealth inequality. Figure 17-1 is a Lorenz curve showing the amount of inequality listed in the columns of Table 17-2; that is, it contrasts the patterns of (1) absolute equality, (2) absolute inequality, and (3) actual 2006 American inequality.

Absolute equality is depicted by the numbers in column (4) of Table 17-2. When they are plotted, these become the diagonal 45° dashed green line of Figure 17-1's Lorenz diagram.

At the other extreme, we have the hypothetical case of absolute inequality, where one person has all the income. Absolute inequality is shown in column (5) of Table 17-2 and by the lowest curve on the Lorenz diagram—the dashed, right-angled blue line.

Any actual income distribution, such as that for 2006, will fall between the extremes of absolute equality and absolute inequality. The green-colored column (6) in Table 17-2 presents the data derived from the first two columns in a form suitable for plotting as an actual Lorenz curve. This actual Lorenz curve appears in Figure 17-1 as the solid green intermediate curve. The shaded area indicates the deviation from absolute equality, hence giving us a measure of the degree of inequality of income distribution.

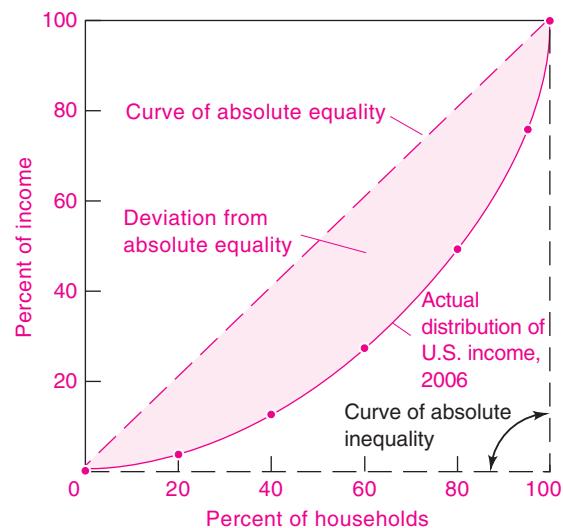


FIGURE 17-1. Lorenz Curve Shows Income Inequality

By plotting the figures from Table 17-2's column (6), we see that the solid green actual-distribution-of-income curve lies between the two extremes of absolute equality and absolute inequality. The shaded area of this Lorenz curve (as a percentage of the triangle's area) measures the relative inequality of income. (How would the curve have looked back in the roaring 1920s when inequality was greater? In an egalitarian Utopia where all have equal inheritances and opportunities?)

(1)	(2) Percentage of total income received by households in this class	(3) Percentage of households in this class and lower ones	(4) Absolute equality	(5) Absolute inequality	(6) Percentage of Income Received by This Class and Lower Ones
Income class of households					
Lowest fifth	3.4	20	20	0	3.4
Second fifth	8.7	40	40	0	12.1
Third fifth	14.8	60	60	0	26.9
Fourth fifth	23.4	80	80	0	50.3
Highest fifth	49.7	100	100	100	100.0

TABLE 17-2. Actual and Polar Cases of Inequality

By cumulating the income shares of each quintile shown in column (2), we can compare in column (6) the actual distribution with polar extremes of complete inequality and equality.

Source: Table 19-1.



The Gini Coefficient

Economists often need to calculate quantitative measures of inequality. One useful measure is the *Gini coefficient*. This is measured by calculating the shaded area in the Lorenz curve of Figure 17-1 and multiplying it by 2. The Gini coefficient is equal to 1 under complete inequality and 0 under complete equality. To see this, recall that a society with equal incomes would have the Lorenz curve run along the 45° line, so the shaded area would be zero. Conversely, when the Lorenz curve runs along the axes, the area is one-half, which, when multiplied by 2, gives a Gini coefficient of 1.

Using the Gini coefficient approach, the Census Bureau calculates that inequality was little changed from 1967 to 1980 (the Gini coefficient rose from .399 to .403) but then rose steadily from 1980 to 2006 (from .403 to .469).

Distribution of Wealth

One major source of the inequality of income is inequality of ownership of *wealth*, which is the net ownership of financial claims and tangible property. Those who are fabulously wealthy—whether because of inheritance, skill, or luck—enjoy incomes far above

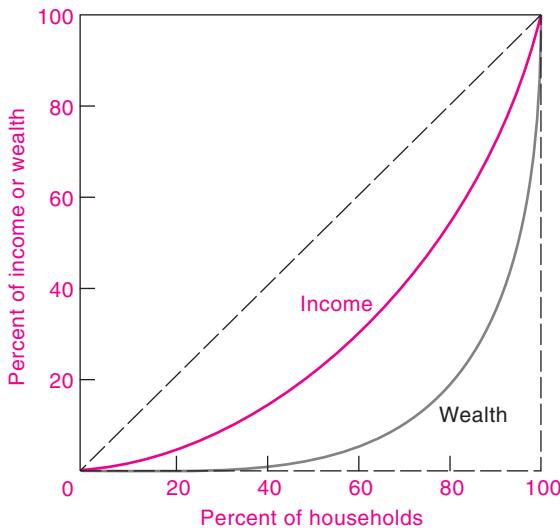


FIGURE 17-2. Inequality of Wealth Is Greater Than for Income

Holdings of wealth tend to be more concentrated than annual incomes.

Source: For income, see Table 17-1. Source for wealth is Federal Reserve Board, *Survey of Consumer Finances, 2004*, available at www.federalreserve.gov/Pubs/oss/oss2/2004/scf2004home_modify.html.

the amount earned by the average household. Those without wealth begin with an income handicap.

In market economies, wealth is much more unequally distributed than is income, as Figure 17-2 shows. In the United States, the top 10 percent of households in 2004 owned 70 percent of wealth, and the top 1 percent of the households owned around 35 percent of all wealth.

Societies are ambivalent about large wealth holdings. A century ago, President T. Roosevelt criticized “malefactors of great wealth” and introduced sharply progressive income and inheritance taxes. A century later, conservatives attempted to abolish all inheritance and gift taxes, labeling them as “death taxes.”

Inequality across Countries

Countries show quite different income distributions depending upon their economic and social structures. Table 17-3 shows the inequality of different countries as measured by the ratio of the income of

	Ratio of income of top 10% to income of bottom 10%
Japan	4.5
Czech Republic	5.2
Sweden	6.2
Germany	6.9
Korea, Republic of	7.8
France	9.0
Spain	9.0
Canada	10.0
Italy	11.7
Australia	12.7
United Kingdom	13.6
United States	15.7
South Africa	31.9
Argentina	38.9
Brazil	67.0
Namibia	129.0

TABLE 17-3. Comparative Inequality in Different Countries

This shows the ratio of the income of the top 10 percent of the population to the income of the bottom 10 percent. Inequality differs greatly across countries. Japan and Western Europe have the least inequality, while South American countries show the greatest.

Source: World Bank, *World Development Indicators, 2005*, available at devdata.worldbank.org/wdi2005/index2.htm.

the top 10 percent to that of the bottom 10 percent of the income distribution. Market-oriented countries like the United States tend to have the most unequal income distributions among the high-income countries. The welfare states of western Europe tend to have the least inequality. The sources of high inequality in the United States are discussed later in this chapter.

The experience of developing countries shows an interesting relationship. Inequality begins to rise as countries begin to industrialize, after which inequality then declines. The greatest extremes of inequality occur in middle-income countries, particularly Latin American countries like Brazil and Argentina.

POVERTY IN AMERICA

"You will always have the poor with you," according to the Scriptures. Poverty is indeed an enduring concern in the United States and in the wider world. Before we can analyze antipoverty programs, we must examine the definition of poverty.



The Elusive Concept of Poverty

The word "poverty" means different things to different people. Clearly, poverty is a condition in which people have inadequate incomes, but it is hard to draw an exact line between the poor and the nonpoor. Economists have therefore devised certain techniques which provide the official definition of poverty.

Poverty was officially defined in the 1960s in the United States as an income insufficient to buy basic food, clothing, shelter, and other necessities. This was calculated from family budgets and double-checked by examining the fraction of incomes that was spent on food. Since that time, the poverty budget has been updated by the government's consumer price index to reflect changes in the cost of living. According to the standard definition, the subsistence cost of living for a family of four was \$21,200 in 2008. This figure represents the "poverty line" or demarcation between poor and nonpoor families. The poverty line also varies by family size.

While an exact figure for measuring poverty is helpful, scholars recognize that "poverty" is a relative term. The notion of a subsistence budget includes subjective questions of taste and social convention. Housing that is today considered substandard often includes household appliances and plumbing that were unavailable to the millionaires and robber barons of an earlier age.

Because of shortcomings in the current definition, a panel of experts of the National Academy of Sciences recommended that the definition of poverty be changed to reflect *relative-income status*. The panel recommended that a family be considered poor if its consumption is less than 50 percent of the median family's consumption of food, clothing, and housing. Poverty in the relative-income sense would decline when inequality decreased; poverty would be unchanged if the economy prospered with no change in the distribution of income and consumption. In this new world, a rising tide would lift all boats but not change the fraction of the population considered poor. This new approach is being weighed carefully by the government.

Who Are the Poor?

Poverty hits some groups harder than others. Table 17-4 shows the incidence of poverty in different groups for 2006. Whites have lower poverty rates than blacks and Hispanics. The elderly no longer have above-average poverty.

Poverty in Major Groups, 2006	
Population group	Percentage of group in poverty
Total population	12.3
By racial and ethnic group:	
White (non-Hispanic)	8.2
Black	24.3
Hispanic	20.6
By age:	
Under 18 years	17.4
18 to 64 years	10.8
65 years and over	9.4
By type of family:	
Married couple	5.7
Female householder, no husband present	30.5
Male householder, no wife present	13.8

TABLE 17-4. Incidence of Poverty in Different Groups, 2006

Whites and married couples have lower-than-average poverty rates. Blacks, Hispanics, and female-headed households have above-average poverty rates.

Source: U.S. Bureau of the Census, *Poverty in the United States: 2006, CPS 2007 Annual Social and Economic Supplement*, downloaded from pubdb3.census.gov/macro/032007/pov/toc.htm.

Perhaps the most troubling trend is that single-parent families headed by women are an increasingly large share of the poor population. In 1959, about 18 percent of poor families were headed by women raising children alone. By 2006, the poverty rate of that group was 30 percent. Social scientists worry that children in single-parent families will receive inadequate nutrition and education and will find it difficult to escape from poverty when they are adults.

Why are so many female-headed and minority families poor? What is the role of discrimination? Experienced observers conclude that blatant racial or gender discrimination in which firms simply pay minorities or women less is vanishing today. Yet the relative poverty of women and blacks continues at a high rate. How can we reconcile these two apparently contradictory trends? The major factor at work is the increasing gap between earnings of highly educated and skilled workers and those of unskilled and less educated workers. Over the last 25 years, the wage differential between these two groups has grown sharply. The growing wage gap has hit minority groups particularly hard.

Who Are the Rich?

At the other extreme are the high earners. Many of the top earners get primarily *property income*, which consists of income on assets like stocks, bonds, and real estate. A generation ago, many of the richest Americans got their wealth through inheritances. Today, entrepreneurship is a much more important road to riches. Most of the richest people in America got that way by taking risks and creating profitable new businesses, such as computer software companies, television networks, and retail chains. The people who invented new products or services or organized the companies that brought them to market got rich on the “Schumpeterian profits” from these innovations. This group of wealthy individuals includes folk heroes like Bill Gates (head of software giant Microsoft), the Waltons (founders of Wal-Mart), and Warren Buffett (investment guru). In an earlier era, the rich lived on stocks, bonds, and land rents.

Another major change among top earners is that wages (including proprietorships) today account for 85 percent of the income of the top 1 percent, whereas that share was only about 50 percent at the beginning of the twentieth century. The high earners are increasingly working in finance and business.

What single profession makes the most money? In recent years, it has been investment bankers and specialists working in financial markets. The average earnings in the securities industry in 2006 was \$206,000 for all workers, and the top managers and analysts make many times that amount.

Why are there such vast differences in compensation among jobs? Some of the differences come from investments in human capital, such as the years of training needed to become a top doctor. Abilities also play a role, for example, in limiting jobs in finance to those who have a deep appreciation of the decimal point. Some jobs pay more because they are dangerous or unpleasant (recall the discussion of compensating differentials in Chapter 13). Moreover, when the supply of labor is limited in an occupation (say, because of union restrictions or professional licensing rules), the supply restrictions drive up the wages and salaries of that occupation.

Trends in Inequality

The inequality of income in the United States has gone through a complete cycle over the last century. The history of inequality in the United States is shown in Figure 17-3. This shows the ratio of the incomes received by the top fifth of families to those received by the bottom fifth. We can see three distinct periods: falling inequality until World War II, stable shares until the 1970s, and then rising inequality over the last three decades. We see that the ratio of upper- to lower-group incomes has almost doubled. Also, examine the income shares of the four top groups, shown in Figure 17-4. The most striking trend is the very top 0.1 percent of the income pyramid. The 133 thousand families in that group had an average income of \$6.3 million in 2006.

Diminishing Inequality. Inequality peaked in 1929 and then declined sharply in the Great Depression as stock prices reduced capital income of the upper groups. The long postwar boom brought prosperity to the middle-class workers, and the share of top income groups declined to its trough in the late 1960s. The share of total income going to the poorest fifth of families rose from 3.8 percent to about 5 percent between 1929 and 1975.

Why did inequality narrow over this period? Inequality declined in part because of the narrowing of wage inequality. With increasing education of

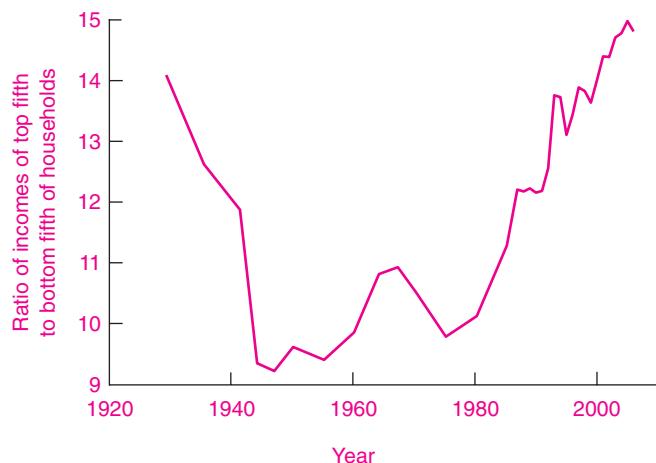


FIGURE 17-3. Trends in Inequality in the United States, 1929–2006

A useful measure of inequality is the ratio of the incomes of the top fifth of the population to those of the bottom fifth. The share of top incomes declined after 1929 with the stock market collapse of the 1930s, the low unemployment and reduced barriers to women and minorities during World War II, and the migration from the farm to the city. Since 1980, income inequality has grown sharply with higher immigration and decline of wages of the unskilled.

Source: U.S. Bureau of the Census, with historical series spliced together by authors.

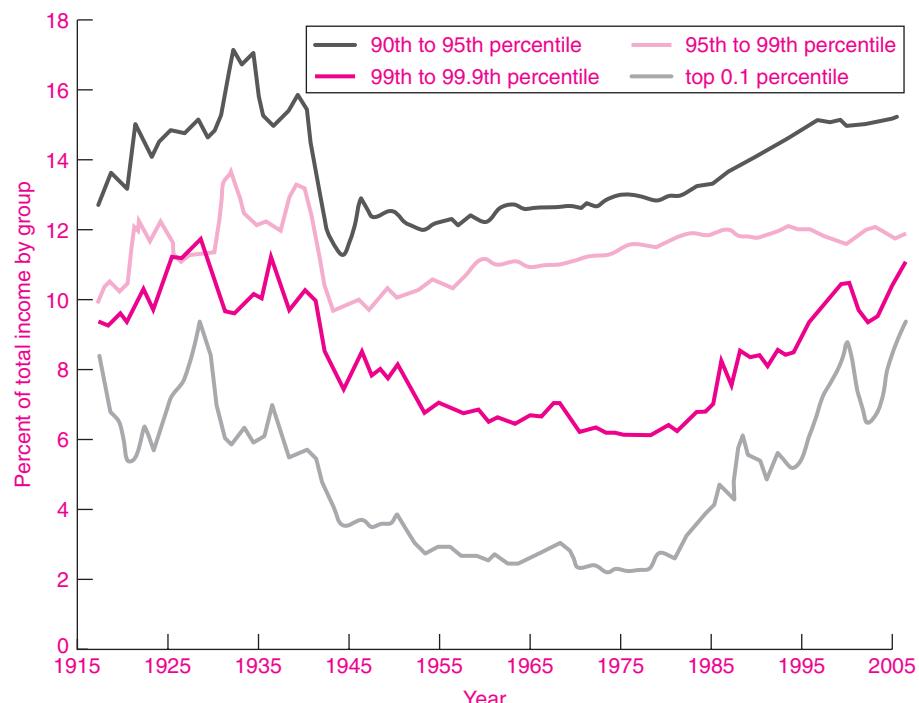


FIGURE 17-4. Income Shares of Top Income Groups, United States, 1917–2006

Inequality fell over most of the twentieth century and then began to rise around 1970. The most dramatic gains were in the very top group—the top 0.1 percent of households. Their share rose from 2 percent of income in 1975 to over 9 percent in the latest year.

Source: The methods were developed in Thomas Piketty and Emmanuel Saez, "Income Inequality in the United States, 1913–1998," *Quarterly Journal of Economics*, 2003. The data here are from their update of March 2008, downloaded from elsa.berkeley.edu/~saez/.

lower-income groups and unionization of the workforce, the wage gap declined. Government policies like social security made a big difference for the elderly population, while programs like cash assistance and food stamps for the indigent and unemployment insurance boosted the incomes of other low-income groups. Our progressive income-tax system, which taxed high incomes more heavily than low incomes, tended to reduce the degree of inequality.

Widening Gaps. In the last quarter-century, several of these trends have reversed themselves. The share of total income going to the bottom quintile declined sharply in the 1980s, sinking from 5.4 percent in 1975 to 3.4 percent in 2006. Average real incomes for families in the bottom fifth are well below their peak. Although the incomes of the poor stagnated during the last quarter-century, the share of income going to the richest Americans soared.

Why did inequality rise in recent decades? After years of intensive debate on this question, a tentative verdict has been proposed in a recent survey by Robert J. Gordon and Ian Dew-Becker. Their conclusions are:

- Virtually none of the rising inequality came from changes in the overall share of labor in national income. That share has been virtually unchanged since 1970.
- The decline of trade unions contributed slightly to increased inequality for men.
- The impact of foreign trade on relative wages appears minimal, while immigration appears to have adversely affected foreign-born workers who are close “substitutes” for immigrants.
- Technological change appears primarily to have depressed the relative wages of the middle-income groups while boosting the incomes of complementary highly skilled workers and having little effect on unskilled service-sector workers.
- The very top of the income distribution has increased its share sharply because of three phenomena. First, the pay of superstars has risen as technology has increased the audience of athletes and entertainers. Second, the incomes of top professionals, particularly in finance, have increased with the increased globalization of the U.S. economy. Third, they endorse the idea that the separation of ownership from control has allowed “the outsized gains in CEO pay.”

This concludes our description of the measurement and sources of inequality. In the next section, we turn to an analysis of government programs to combat poverty and reduce inequality. High-income democracies everywhere are rethinking these programs as they redefine the role of the state.

B. ANTIPOVERTY POLICIES

All societies take steps to provide for their poor citizens. But what is given to the poor must come from other groups, and that is undoubtedly the major point of resistance to redistributive programs. In addition, economists worry about the impact of redistribution upon the efficiency and morale of a country. In this section, we review the rise of the welfare state, consider the costs of income redistribution, and survey the current system of income maintenance.

The Rise of the Welfare State

The early classical economists believed the distribution of income was unalterable. They argued that attempts to alleviate poverty by government interventions in the economy were foolish endeavors that would simply end up reducing total national income. This view was contested by the English economist and philosopher John Stuart Mill. While cautioning against interferences with the market mechanism, he argued eloquently that government policies could reduce inequality.

A half-century later, at the end of the nineteenth century, political leaders in Western Europe took steps that marked a historic turning point in the economic role of government. Bismarck in Germany, Gladstone and Disraeli in Britain, followed by Franklin Roosevelt in the United States introduced a new concept of government responsibility for the welfare of the populace.

This marked the rise of the **welfare state**, in which government takes steps to protect individuals against specified contingencies and to guarantee people a minimum standard of living.

Important welfare-state programs include public pensions, accident and sickness insurance, unemployment insurance, health insurance, food and housing

programs, family allowances, and income supplements for certain groups of people. These policies were introduced gradually from 1880 through to the modern era. The welfare state came late to the United States, being introduced in the New Deal of the 1930s with unemployment insurance and social security. Medical care for the aged and the poor was added in the 1960s. In 1996 the federal government turned back the clock by removing the guarantee of a minimum income. The debate over redistribution never ends.

THE COSTS OF REDISTRIBUTION

One of the goals of a modern mixed economy is to provide a safety net for those who are temporarily or permanently unable to provide adequate incomes for themselves. One reason for these policies is to promote greater equality.

What are the different concepts of equality? To begin with, democratic societies affirm the principle of equality of *political rights*—generally including the right to vote, the right to trial by jury, and the right to free speech and association. In the 1960s, liberal philosophers espoused the view that people should also have equal *economic opportunity*. In other words, all people should play by the same rules on a level playing field. All should have equal access to the best schools, training, and jobs. Then discrimination on the basis of race or gender or religion would disappear. Many steps were taken to promote greater equality, but inequalities of opportunity have proved very stubborn.

A third, and the most far-reaching, ideal is equality of *economic outcomes*. In this utopia, people would have the same consumption whether they were smart or dull, eager or lazy, lucky or unfortunate. Wages would be the same for doctor and nurse, lawyer and secretary. “From each according to his abilities, to each according to his needs” was Karl Marx’s formulation of this philosophy.

Today, even the most radical socialist recognizes that some differences in economic outcome are necessary if the economy is to function efficiently. Without some differential reward for different kinds of work, how can we ensure that people will do the unpleasant as well as the enjoyable work, that they will work on dangerous offshore oil derricks as well as in beautiful parks? Insisting on equality of outcomes would severely hamper the functioning of the economy.



The Leaky Bucket

In taking steps to redistribute income from the rich to the poor, governments may harm economic efficiency and reduce the amount of national income available to distribute. On the other hand, if equality is a social good, it is one worth paying for.

The question of how much we are willing to pay in reduced efficiency for greater equity was addressed by Arthur Okun in his “leaky bucket” experiment. He noted that if we value equality, we would approve when a dollar is taken in a bucket from the very rich and given to the very poor. But, he continued, suppose the bucket of redistribution has a leak in it. Suppose only a fraction—maybe only one-half—of each dollar paid by the rich in taxes actually reaches the poor. Then redistribution in the name of equity has been at the expense of economic efficiency.¹

Okun presented a fundamental dilemma. Redistributive measures like the progressive income tax, analyzed in Chapter 16, will reduce real output by reducing incentives to work and save. As a nation considers its income-distribution policies, it will want to weigh the benefit of greater equality against the impact of these policies on total national income.

Redistribution Costs in Diagrams

We can illustrate Okun’s point by using the income-possibility curve of Figure 17-5. This graph shows the incomes available to different groups when government programs redistribute income.

We begin by dividing the population in half; the real income of the lower half is measured on the vertical axis of Figure 17-5, while the income of the upper half is measured on the horizontal axis. At point A, which is the pre-redistribution point, no taxes are levied and no transfers are given, so people simply live with their market incomes. In a competitive economy, point A will be efficient and the no-redistribution policy maximizes total national income.

However, at laissez-faire point A, the upper-income group receives substantially more income than the lower half. People might strive for greater equality by tax and transfer programs, hoping to move toward the point of equal incomes at E. If such

¹ Arthur M. Okun, *Equality and Efficiency: The Big Tradeoff* (Brookings Institution, Washington, D.C., 1975).

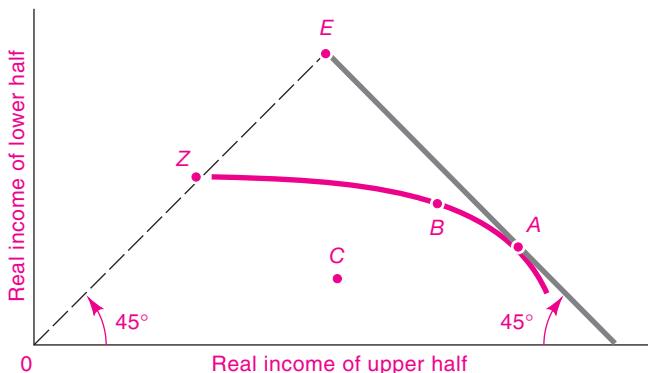


FIGURE 17-5. Redistributing Income May Harm Economic Efficiency

Point A marks the most efficient outcome, with maximal national output. If society could redistribute with no loss of efficiency, the economy would move toward point E. Because redistributive programs generally create distortions and efficiency losses, the path of redistribution might move along the green line ABZ. Society must decide how much efficiency to sacrifice to gain greater equality. Why would everyone want to avoid redistributive programs that take the economy from point B to point C?

steps could be taken without reducing national output, the economy would move along the blue line from A toward E. The slope of the AE line is -45° , reflecting the assumption about efficiency that the redistributive bucket has no leaks, so every dollar taken from the upper half increases the income of the lower half by exactly \$1. Along the -45° line, total national income is constant, indicating that redistributive programs have no impact upon the total national income.

Most redistributive programs do affect efficiency. If a country redistributes income by imposing high tax rates on the wealthiest people, their saving and work effort may be reduced or misdirected, with a resulting lower total national output. They may spend more money on tax lawyers or invest less in high-yielding but risky innovations. Also, if society puts a guaranteed floor beneath the incomes of the poor, the sting of poverty will be reduced and the poor may work less. All these reactions to redistributive programs reduce the total size of real national income.

In terms of Okun's experiment, we might find that for every \$100 of taxation on the rich, the income of the poor increased by only \$50, with the rest dissipated because of reduced effort or administrative costs. The bucket of redistribution has developed a leak. Costly redistribution is shown by the ABZ curve in Figure 17-5. Here, the hypothetical frontier of real incomes bends away from the -45° line because taxes and transfers produce inefficiencies.

The experience of socialist countries exemplifies how attempts to equalize incomes by expropriating property from the rich can end up hurting everyone. By prohibiting private ownership of businesses,

socialist governments reduced the inequalities that arise from large property incomes. But the reduced incentives for work, investment, and innovation crippled this radical experiment of "to each according to his needs" and impoverished entire countries. By 1990, comparisons of living standards in East and West had convinced many socialist countries that private ownership of business would benefit the living standards of workers as well as capitalists.

How Big Are the Leaks?

Okun characterized our redistributive system of taxes and transfers as a leaky bucket. But just how big are the leaks in the American economy? Is the country closer to Figure 17-5's point A, where the leaks are negligible? Or to B, where they are substantial? Or to Z, where the redistributive bucket is in fact a sieve? To find the answer, we must examine the major inefficiencies induced by high tax rates and by generous income-support programs: administrative costs, damage to work and saving incentives, and socioeconomic costs.

- The government must hire tax collectors to raise revenues and social security accountants to disburse them. These are clear inefficiencies or regrettable necessities, but they are small: the Internal Revenue Service spends only half a penny on administrative costs for each dollar of collected revenues.
- As the tax collector's bite grows larger and larger, might I not become discouraged and end up working less? Tax rates might conceivably be so high that total revenues are actually lower

than they would be at more modest tax rates. Empirical evidence, however, suggests that the damage of taxes on work effort is limited. For a few groups, the labor supply curve may actually be backward-bending, indicating that a tax on wages might increase rather than decrease work effort. Most studies find that taxes have only a small impact on labor effort for middle-income and high-income workers. However, there may well be substantial impacts of the tax and transfer system on the behavior of poor people.

- Perhaps the most important potential leakage from the revenue bucket is the savings component. Some believe that current government programs discourage saving and investment. Some economic studies indicate that by taxing income rather than consumption, total saving is reduced. Additionally, economists worry that the nation's saving rate has declined sharply because of generous social programs—especially social security and Medicare—that reduce the need for people to save for old age and health contingencies.
- Some claim that the leaks cannot be found in the cost statistics of the economist; instead, the costs of equality are seen in attitudes rather than in dollars. Are people so turned off by the prospect of high taxes that they turn on to drugs and idleness? Is the welfare system leading to a permanent underclass, a society of people who are trapped in a culture of dependency?
- Some people criticize the entire notion of costly redistribution, arguing as follows: Poverty is rooted in malnourishment in the early years, broken families, illiteracy at home, poor education, and lack of job training. Poverty begets poverty; the vicious cycle of malnutrition, poor education, drug dependency, low productivity, and low incomes leads to yet another generation of poor families. These analysts contend that enhanced programs to provide health care and adequate food for poor families will increase productivity and efficiency rather than decrease output. By breaking the vicious cycle of poverty today, we will be raising the skills, human capital, and productivity of the children of poverty tomorrow.

Adding Up the Leaks

When all the leaks are added up, how big are they? Okun argued that the leaks are small, particularly when funds for redistributive programs are drawn from the tap of a broad-based income tax. Others disagree strenuously, pointing to high marginal tax rates and overly generous transfer programs as confusing and destructive of economic efficiency.

What is the reality? While much research has been undertaken on the cost of redistribution, the truth has proved elusive. A cautious verdict is that there are but modest losses to economic efficiency from redistributive programs of the kind used in the United States today. For many people, the efficiency costs of redistribution are a reasonable price to pay for reducing the economic and human costs of poverty in malnutrition, poor health, lost job skills, and human misery. But countries whose welfare-state policies have gone far beyond those in the United States see major inefficiencies. Egalitarian countries like Sweden and the Netherlands, which provided cradle-to-grave protection for their citizens, found declining labor-force participation, growing unemployment, and rising budget deficits. These countries have taken steps to reduce the burden of the welfare state.

Countries need to design their policies carefully to avoid the extremes of unacceptable inequality or great inefficiency.

ANTIPOVERTY POLICIES: PROGRAMS AND CRITICISMS

All societies provide for their aged, their young, and their sick. Sometimes, the support comes from families or religious organizations. Over most of the last century, central governments have increasingly assumed the responsibility for providing income support for the poor and needy. Yet, as governments have assumed larger responsibilities for more people, the fiscal burdens of transfer programs have grown steadily. Today, most high-income countries face the prospect of rising tax burdens to finance health and retirement programs as well as income-support programs for poor families. This rising tax burden has provoked a sharp backlash against "welfare programs," particularly in the United States. Let's review the major antipoverty programs and recent reforms.

Income-Security Programs

What are the major income-security programs today? Let's look briefly at a few of the programs that have been established in the United States.

Most income-security programs are targeted at the elderly rather than the poor. The major programs are social security, which is a contributory federal retirement program, and Medicare, which is a subsidized health program for those over 65 years old. These two programs are the largest transfer programs in the United States and in most other high-income countries.

Programs specifically targeted to poor households are a patchwork quilt of federal, state, and local programs. Some of these are cash assistance. Others subsidize particular goods or services, such as the food-stamps program or Medicaid, which provides poor families with free health care. Most of the programs targeted to poor families have shrunk sharply over the last two decades.

The most controversial program was cash assistance to poor parents with small children. This program was drastically reformed in 1996, and we will discuss the reform below.

How much do all federal programs add up to in terms of budget expenditures? All federal poverty programs today amount to 20 percent of the total federal budget.

Incentive Problems of the Poor

One of the major obstacles faced by poor families is that the rules in most welfare programs severely reduce the incentives of low-income adults to seek work. If a poor person on welfare gets a job, the government will trim back food stamps, income-support payments, and rent subsidies, and the person might even lose medical benefits. We might say that poor people face high marginal "tax rates" (or, more accurately, "benefit-reduction rates") because welfare benefits are sharply reduced as earnings rise.

THE BATTLE OVER WELFARE REFORM

The traditional welfare system has few defenders. Some want to dismantle it; others, to strengthen it. Some wish to devolve responsibility for income support to states, localities, or families; others, to broaden the federal role. These disparate approaches reflect

disparate views of poverty and lead to strikingly different policy proposals.

Two Views of Poverty

Social scientists put forth a wide variety of proposals to cure or alleviate poverty. The different approaches often reflect differing views of the roots of poverty. Proponents of strong government action see poverty as the result of social and economic conditions over which the poor have little control. They stress malnutrition, poor schools, broken families, discrimination, lack of job opportunities, and a dangerous environment as central determinants of the fate of the poor. If you hold this view, you might well believe that government bears a responsibility to alleviate poverty—either by providing income to the poor or by correcting the conditions that produce poverty.

A second view holds that poverty grows out of maladaptive individual behavior—behavior that is the responsibility of individuals and is properly cured by the poor themselves. In earlier centuries, laissez-faire apologists held that the poor were shiftless, lazy, or drunk; as a charity worker wrote almost a century ago, "Want of employment . . . is, as often as not, [caused by] drink." Sometimes the government itself is blamed for breeding dependency on government programs that squelch individual initiative. Critics who hold these views advocate that the government should cut back on welfare programs so that people will develop their own resources.

The poverty debate was succinctly summarized by the eminent social scientist William Wilson:

Liberals have traditionally emphasized how the plight of disadvantaged groups can be related to the problems of the broader society, including problems of discrimination and social class subordination. . . . Conservatives, in contrast, have traditionally stressed the importance of different group values and competitive resources in accounting for the experiences of the disadvantaged.²

Much of today's debate can be better understood if these two views and their implications are factored into the political equation.

² William Julius Wilson, "Cycles of Deprivation and the Underclass Debate," *Social Service Review*, December 1985, pp. 541–559.

Income-Support Programs in the United States Today

Most high-income countries provide guaranteed income supplements for poor families with children, and that model was followed by the United States until 1996. At that time, the country took a radically different approach to increasing incomes of the poor. First, the government augmented a program to supplement wages of working families. Second, it fundamentally altered cash assistance programs, abolishing a federal entitlement for poor families.

The Earned-Income Tax Credit

The wage supplement program is called the *earned-income tax credit* or *EITC*. This credit applies to labor incomes and is in effect a wage supplement. In 2008, the EITC provided a supplement to wage income of as much as 40 percent, up to a maximum of \$4824 for a family with two children. A single father or mother would receive some credit for an income up to around \$39,000 of wages. It is known as a “refundable” credit because it is actually paid to an individual when the individual owes no taxes.

What is the difference between a traditional cash-assistance program and the earned-income tax credit? Cash assistance provides a minimum benefit for poor families and then reduces the benefit as market income increases. The earned-income tax credit, by contrast, gives nothing to those who do not work and supplements the earnings of those who do work. The philosophy of the EITC in essence is, “Those who do not work shall not get government dollars.”

The 1996 U.S. Welfare Reform

From the 1930s until 1996, poor families could also benefit from a federal cash-assistance program known as Aid to Families with Dependent Children. This was an *entitlement program*, meaning that anyone who met certain qualifications could receive the benefits as a matter of law.

President Bill Clinton had run on a platform of “reforming welfare as we know it.” In 1996, he teamed up with a Republican Congress and completely changed the rules for cash assistance. The old program was replaced by the Temporary Assistance for Needy Families (TANF) program, which removed the federal entitlement to cash benefits and turned the program over to the 50 states.

The major provisions of the new program were the following:

- The primary responsibility for the income support of poor people was turned over to state and local governments. This replaced the earlier system in which the federal government picked up most of the costs of income support.
- The entitlement for federal cash assistance under TANF was removed.
- Each family is subject to a lifetime limit of 5 years of benefits under the federally supported program. After 5 years, TANF funds can no longer be used to support the family, even if it moves to a new state or has been off the welfare rolls for a number of years.
- Adults in the program must engage in work activities after 2 years of benefits.
- Legal immigrants may be excluded from TANF benefits.
- Other major low-income-support programs were largely unchanged.

Appraisal. The 1996 welfare reform was a major change in social policy. One aspect is the effect on *labor markets*. To the extent that the loss of benefits forces people to seek work, this will increase the supply of relatively uneducated and unskilled labor. This increased supply will tend to lower wages of the lowest-paid workers and increase income inequality. (This effect operates much the same way that the sharp increase in immigration has contributed to lowering of wages of the unskilled in the last three decades.) If the equilibrium wages of some workers are driven down below the minimum wage, this may also lead to an increase in the unemployment rate of these groups.

One important feature of the new law, emphasized by social and economic conservatives, was the *transfer of responsibility* for income support for poor families to the states. The idea behind this change was that states would reverse the century-long trend of increasing generosity of welfare programs. Critics of this transfer believed that placing decision-making responsibility in the states would give strong incentives for states to trim welfare benefits to reduce the costs and the fiscal burden of the low-income population. This has been called a “race to the bottom” in which the equilibrium is for states to have the

lowest-possible benefits and drive low-income households elsewhere.

The *impacts* of the expanded EITC and 1996 welfare reform have surprised most analysts. Among the major impacts have been the following:

- The fall in welfare caseloads has been unprecedented, widespread, and continuous. From 1995 to 2008, the number of households on welfare has fallen by more than 70 percent. While a decline was expected, its size and duration were surprising.
- There was a large increase in the labor-force participation rate of single women with young children. The combination of economic incentives and a strong labor market was successful in pushing women off welfare and into jobs.

ECONOMIC POLICY FOR THE 21ST CENTURY

How should government's role in the economy be redefined? We close with three final reflections:

1. We have examined the key economic functions of the government. The government combats market failures, redistributes income, stabilizes the economy, manages international affairs, and promotes long-term economic growth. Each of these is essential. No serious person today advocates shutting down the government. No one today proposes to allow nuclear dumping, to let poor orphans starve in the streets, to privatize the central bank, or to open the borders to all flows of people and drugs. The question is not whether government should regulate the economy but how and where it should intervene.

2. While government plays a central role in a civilized society, we must constantly reassess the mission and instruments of government policy. Governments have a monopoly on political power, and this imposes a special responsibility for government to operate efficiently. Every public dollar spent on wasteful programs could be used for promoting scientific research or alleviating hunger. Every inefficient tax reduces people's consumption opportunities, whether for food or education or housing. The central premise of economics is that resources are scarce—and this applies to the government as well as to the private sector.
3. While economics can analyze the major public-policy controversies, it cannot have the final word. For underlying all public-policy debates are normative assumptions and value judgments about what is just and fair. What an economist does, therefore, is try very hard to keep positive science cleanly separated from normative judgments—to draw a line between the economic calculations of the head and the human feelings of the heart. But keeping description separate from prescription does not mean that the professional economist is a bloodless computer. Economists are as divided in their political philosophies as is the rest of the population. Conservative economists argue strenuously for reducing the scope of government and ending programs to redistribute income. Liberal economists are just as passionate in advocating reducing poverty or using macroeconomic policies to combat unemployment. Economic science cannot say which political point of view is right or wrong. But it can arm us for the great debate.



SUMMARY

A. The Sources of Inequality

1. In the previous century, the classical economists believed that inequality was a universal constant, unchangeable by public policy. This view does not stand up to scrutiny. Poverty made a glacial retreat over the early part of the twentieth century, and absolute incomes for those in the bottom part of the income

distribution rose sharply. Since around 1980, this trend has reversed, and inequality has increased.

2. The Lorenz curve is a convenient device for measuring the spreads or inequalities of income distribution. It shows what percentage of total income goes to the poorest 1 percent of the population, to the poorest 10 percent, to the poorest 95 percent, and so forth.

The Gini coefficient is a quantitative measure of inequality.

3. Poverty is essentially a relative notion. In the United States, poverty was defined in terms of the adequacy of incomes in the early 1960s. By this standard of measured income, little progress in reducing inequality has been made in the last decade.
4. Income inequality declined markedly over most of the twentieth century. Then, beginning around 1975, the gap between rich and poor began to widen. The largest income gains have gone to the very top of the income distribution, to the richest 0.1 percent of people. Analysts believe that the “rich man’s crash” of 2007–2009 will narrow income gaps at the very top. Wealth is even more unequally distributed than is income, both in the United States and in other capitalist economies.

B. Antipoverty Policies

5. Political philosophers write of three types of equality: (a) equality of political rights, such as the right to vote; (b) equality of opportunity, providing equal

access to jobs, education, and other social systems; and (c) equality of outcomes, whereby people are guaranteed equal incomes or consumptions. Whereas the first two types of equality are increasingly accepted in most advanced democracies like the United States, equality of outcomes is generally rejected as impractical and too harmful to economic efficiency.

6. Equality has costs as well as benefits; the costs show up as drains from Okun’s “leaky bucket.” That is, attempts to reduce income inequality by progressive taxation or transfer payments may harm economic incentives to work or save and may thereby reduce the size of national output.
7. Major programs to alleviate poverty are welfare payments, food stamps, Medicaid, and a group of smaller or less targeted programs. As a whole, these programs are criticized because they impose high benefit-reduction rates (or marginal “tax” rates) on low-income families when families begin to earn wages or other income.

CONCEPTS FOR REVIEW

trends of income distribution

Lorenz curve of income and wealth

Gini coefficient

poverty

welfare state

Okun’s “leaky bucket”

equality: political, of opportunity,
of outcomes

equality vs. efficiency

income-possibility curve: ideal and
realistic cases

FURTHER READING AND INTERNET WEBSITES

Further Reading

An influential book on equality versus efficiency is Arthur Okun, *Equality and Efficiency: The Big Tradeoff* (Brookings Institution, Washington, D.C., 1975).

For a nontechnical review of issues in health-care reform, see the symposium in *Journal of Economic Perspectives*, Summer 1994.

Websites

The Census Department collects poverty data. See www.census.gov/hhes/www/poverty.html. For information on

welfare and poverty, see www.welfareinfo.org. The site www.dolefa.gov describes the results of welfare reform from the perspective of individuals.

The Urban Institute (www.urban.org) and the Joint Center for Poverty Research (www.jcpr.org) are organizations devoted to analyzing trends in poverty and income distribution.

QUESTIONS FOR DISCUSSION

1. Let each member of the class anonymously write down on a card an estimate of his or her family's annual income. From these, draw up a frequency table showing the distribution of incomes. What is the median income? The mean income?
2. What effect would the following have on the Lorenz curve of after-tax incomes? (Assume that the taxes are spent by the government on a representative slice of GDP.)
 - a. A proportional income tax (i.e., one taxing all incomes at the same rate)
 - b. A progressive income tax (i.e., one taxing high incomes more heavily than low incomes)
 - c. A sharp increase in taxes on cigarettes and food

Draw four Lorenz curves to illustrate the original income distribution and the income distribution after each of the three tax categories.
3. Review Okun's leaky bucket experiment. Get a group together and have each member of the group write down on a piece of paper how large a leak should be tolerated when government transfers \$100 from the top income quintile to the bottom income quintile. Do you think it should be 99 percent? Or 50 percent? Or zero? Each person should write a short justification of the maximum number. Tabulate the results and then discuss the differences.
4. Consider two ways of supplementing the income of the poor: (a) cash assistance (say, \$500 per month) and (b) categorical benefits such as subsidized food or medical care. List the pros and cons of using each

strategy. Can you explain why the United States tends to use mainly strategy (b)? Do you agree with this decision?

5. In a country called Econoland, there are 10 people. Their incomes (in thousands) are \$3, \$6, \$2, \$8, \$4, \$9, \$1, \$5, \$7, and \$5. Construct a table of income quintiles like Table 17-2. Plot a Lorenz curve. Calculate the Gini coefficient defined in Section A.
6. People continue to argue about what form assistance for the poor should take. One school says, "Give people money and let them buy health services and the foods they need." The other school says, "If you give money to the poor, they may spend it on beer and drugs. Your dollar goes further in alleviating malnourishment and disease if you provide the services "in kind" (meaning by directly providing the good or service rather than providing money to buy the good or service.) The dollar that you earn may be yours to spend, but society's income-support dollar is a dollar that society has the right to channel directly to its targets."

The argument of the first school might rest on demand theory: Let each household decide how to maximize its utility on a limited budget. Chapter 5 shows why this argument might be right. But what if the parents' utility includes mainly beer and lottery tickets and no milk or clothing for the children? Might you agree with the second view? From your own personal experience and reading, which of these two arguments would you endorse? Explain your reasoning.

International Trade

18



TO THE CHAMBER OF DEPUTIES:

We are subject to the intolerable competition of a foreign rival, who enjoys such superior facilities for the production of light that he can inundate our national market at reduced price. This rival is no other than the sun. Our petition is to pass a law shutting up all windows, openings, and fissures through which the light of the sun is used to penetrate our dwellings, to the prejudice of the profitable manufacture we have been enabled to bestow on the country. Signed: The Candle Makers

F. Bastiat

A. THE NATURE OF INTERNATIONAL TRADE

As we go about our daily lives, it is easy to overlook the importance of international trade. America ships enormous volumes of food, airplanes, computers, and machinery to other countries; and in return we get vast quantities of oil, footwear, cars, coffee, and other goods and services. While Americans pride themselves on their ingenuity, it is sobering to realize how many of our products—including gunpowder, classical music, clocks, railroads, penicillin, and radar—arose from the inventions of long-forgotten people in faraway places.

What are the economic forces that lie behind international trade? Simply put, trade promotes specialization, and specialization increases productivity.

Over the long run, increased trade and higher productivity raise living standards for all nations. Gradually, countries have realized that opening up their economies to the global trading system is the most secure road to prosperity.

This chapter extends our analysis by examining the principles governing *international trade*, through which nations export and import goods, services, and capital. International economics involves many of the most controversial questions of the day. Should the nation be concerned that so many of its consumer goods are made abroad? Do we gain from free trade, or should we tighten up the rules on trading with Mexico and China? Are workers hurt in competition with “cheap foreign labor”? How should the principles governing trade be extended to intellectual property rights, such as patents and copyrights? The economic stakes are high in finding sound answers to these questions.

International vs. Domestic Trade

In a deep economic sense, trade is trade, whether it involves people within the same nation or people in different countries. There are, however, three important differences between domestic and international trade, and these have important practical and economic consequences:

- 1. Expanded trading opportunities.** The major advantage of international trade is that it expands the scope of trade. If people were forced to consume only what they produced at home, the world would be poorer on both the material and the spiritual planes. Canadians could drink no wine, Americans could eat no bananas, and most of the world would be without jazz and Hollywood movies.
- 2. Sovereign nations.** Trading across frontiers involves people and firms living in different nations. Each nation is a sovereign entity which regulates the flow of people, goods, and finance crossing its borders. This contrasts with domestic trade, where there is a single currency, where trade and money flow freely within the borders, and where people can migrate easily to seek new opportunities. Countries sometimes build barriers to international trade, using tariffs or quotas, to “protect” affected workers or firms from foreign competition.
- 3. International finance.** Most nations have their own currencies. I want to pay for a Japanese car in U.S. dollars, while Toyota wants to be paid in Japanese yen. Dollars are translated into yen by the foreign exchange rate, which is the relative price of different currencies. The international financial system must ensure a smooth flow and exchange of dollars, yen, and other currencies—or else risk a breakdown in trade. The financial aspects of international trade are analyzed in the chapters on macroeconomics.

Trends in Foreign Trade

What are the major components of international trade for the United States? Table 18-1 shows the composition of U.S. foreign trade for 2007. The bulk of trade is in goods, particularly manufactured goods, although trade in services has increased rapidly. The data reveal that the United States exports surprisingly large amounts of primary commodities (such as food)

	International Trade in Goods and Services, 2007 (billions of dollars)	
	Exports	Imports
Goods	1,149	1,965
Food and beverages	84	50
Industrial supplies	316	269
Capital goods	446	284
Motor vehicles	121	204
Consumer goods	146	308
Other goods	36	49
Services	479	372
Travel	97	76
Passenger fares	25	29
Other transportation	52	67
Royalties and license fees	71	28
Other private services	217	135
Military sales and government	17	37
Total goods and services	1,628	2,337

TABLE 18-1. International Trade in Goods and Services

The United States exports a wide array of goods and services from food to intellectual property. In 2007, U.S. imports exceeded exports by around \$700 billion. The United States exports primarily specialized capital goods like machinery. At the same time, it imports many other manufactured goods, like cars and cameras, because other countries specialize in different market niches and enjoy economies of scale.

Source: U.S. Bureau of Economic Analysis, available at www.bea.gov/international/.

and imports large quantities of sophisticated, capital-intensive manufactured goods (like automobiles and computer parts). Moreover, we find a great deal of two-way, or intra-industry, trade. Within a particular industry, the United States exports and imports at the same time because a high degree of product differentiation means that different countries tend to have niches in different parts of a market.

THE REASONS FOR INTERNATIONAL TRADE IN GOODS AND SERVICES

What are the economic factors that lie behind the patterns of international trade? Nations find it beneficial to participate in international trade for several

reasons: diversity in the conditions of production, differences in tastes among nations, and decreasing costs of large-scale production.

Diversity in Natural Resources

Trade may take place because of the diversity in productive possibilities among countries. In part, these differences reflect endowments of natural resources. One country may be blessed with a supply of petroleum, while another may have a large amount of fertile land. Or a mountainous country may generate large amounts of hydroelectric power which it sells to its neighbors, while a country with deep-water harbors may become a shipping center.

Differences in Tastes

A second reason for trade lies in preferences. Even if the conditions of production were identical in all regions, countries might engage in trade if their tastes for goods were different.

For example, suppose that Norway and Sweden both produce fish from the sea and meat from the land in about the same amounts but the Swedes have a great fondness for meat while the Norwegians are partial to fish. A mutually beneficial export of meat from Norway and fish from Sweden would take place. Both countries would gain from this trade; the sum of human happiness is increased, just as when Jack Sprat trades fat meat for his wife's lean.

Differences in Costs

Perhaps the most important reason for trade is differences among countries in production costs. We see vast differences in labor costs among nations. In 2006, for example, China's hourly wage of \$1 was about one-thirtieth of that in Western Europe. Companies looking to compete effectively strive to find those parts of the production chain that can profitably be located in China to use unskilled Chinese workers. When an iPod or mobile phone is labeled "Made in China," that probably means that it was assembled in China, while the design, patents, marketing, and hard drives were produced in other countries.

An important feature in today's world is that some companies or countries enjoy economies of scale; that is, they tend to have lower average costs of production as the volume of output expands. So when a particular country gets a head start in producing a particular product, it can become the high-

volume, low-cost producer. The economies of scale give it a significant cost and technological advantage over other countries, which find it cheaper to buy from the leading producer than to make the product themselves.

Large-scale production is an important advantage in industries with major research-and-development expenses. As the leading aircraft maker in the world, Boeing can spread the enormous cost of designing, developing, and testing a new plane over a large sales volume. That means it can sell planes at a lower price than competitors with a smaller volume. Boeing's only real competitor, Airbus, got off the ground through large subsidies from several European countries to cover its research-and-development costs.

The example of decreasing cost helps explain the important phenomenon of extensive intra-industry trade shown in Table 18-1. Why is it that the United States both imports and exports computers and related equipment? Consider a company such as Intel, which produces high-end semiconductors. Intel has facilities in the United States as well as in China, Malaysia, and the Philippines, and the company often ships products manufactured in one country to be assembled and tested in another country. Similar patterns of intra-industry specialization are seen with cars, steel, textiles, and many other manufactured products.

B. COMPARATIVE ADVANTAGE AMONG NATIONS

THE PRINCIPLE OF COMPARATIVE ADVANTAGE

It is only common sense that countries will produce and export goods for which they are uniquely qualified. But there is a deeper principle underlying *all* trade—in a family, within a nation, and among nations—that goes beyond common sense. The *principle of comparative advantage* holds that a country can benefit from trade even if it is absolutely more efficient (or absolutely less efficient) than other countries in the production of every good. Indeed, trade according to comparative advantage provides mutual benefits to all countries.

Uncommon Sense

Take a world in which there are only two goods, computers and clothing. Suppose that the United States has higher output per worker (or per unit of input) than the rest of the world in making both computers and clothing. But suppose the United States is relatively more efficient in the production of computers than it is in clothing. For example, it might be 50 percent more productive in computers and 10 percent more productive in clothing than other countries. In this case, it would benefit the United States to export that good in which it is relatively more efficient (computers) and import that good in which it is relatively less efficient (clothing).

Or consider a poor country like Mali. How could impoverished Mali, whose workers use handlooms and have productivity that is only a fraction of that of workers in industrialized countries, hope to export any of its textiles? Surprisingly, according to the principle of comparative advantage, Mali can benefit by exporting the goods in which it is *relatively* more efficient (like textiles) and importing those goods which it produces *relatively* less efficiently (like turbines and automobiles).

The principle of comparative advantage holds that each country will benefit if it specializes in the production and export of those goods that it can produce at relatively low cost. Conversely, each country will benefit if it imports those goods which it produces at relatively high cost.

This simple principle provides the unshakable basis for international trade.

Ricardo's Analysis of Comparative Advantage

Let us illustrate the fundamental principles of international trade by considering America and Europe two centuries ago. If labor (or resources, more generally) is absolutely more productive in America than in Europe, does this mean that America will import nothing? And is it economically wise for Europe to "protect" its markets with tariffs or quotas?

These questions were first answered in 1817 by the English economist David Ricardo, who showed that international specialization benefits a nation. He called this result the law of comparative advantage.

For simplicity, Ricardo worked with only two regions and only two goods, and he chose to measure all production costs in terms of labor-hours. We will

American and European Labor Requirements for Production		
Product	Necessary Labor for Production (labor-hours)	
	In America	In Europe
1 unit of food	1	3
1 unit of clothing	2	4

TABLE 18-2. Comparative Advantage Depends Only on Relative Costs

In a hypothetical example, America has lower labor costs in both food and clothing. American labor productivity is between 2 and 3 times Europe's (twice in clothing, thrice in food).

follow his lead here, analyzing food and clothing for Europe and America.¹

Table 18-2 shows the illustrative data. In America, it takes 1 hour of labor to produce a unit of food, while a unit of clothing requires 2 hours of labor. In Europe the cost is 3 hours of labor for food and 4 hours of labor for clothing. We see that America has *absolute advantage* in both goods, for it can produce either one with greater absolute efficiency than can Europe. However, America has *comparative advantage* in food, while Europe has *comparative advantage* in clothing. The reason is that food is *relatively inexpensive* in America compared to Europe, while clothing is *relatively inexpensive* in Europe compared to America.

From these facts, Ricardo proved that both regions will benefit if they specialize in their areas of comparative advantage—that is, if America specializes in the production of food while Europe specializes in the production of clothing. In this situation, America will export food to pay for European clothing, while Europe will export clothing to pay for American food.

To analyze the effects of trade, we must measure the amounts of food and clothing that can be produced and consumed in each region (1) if there is no international trade and (2) if there is free trade with each region specializing in its area of comparative advantage.

¹ An analysis of comparative advantage with many countries and many commodities is presented later in this chapter.

Before Trade. Start by examining what occurs in the absence of any international trade, say, because all trade is illegal or because of a prohibitive tariff. Table 18-2 shows the real wage of the American worker for an hour's work as 1 unit of food or $\frac{1}{2}$ unit of clothing. The European worker earns only $\frac{1}{3}$ unit of food or $\frac{1}{4}$ unit of clothing per hour of work.

Clearly, if perfect competition prevails in each isolated region, the prices of food and clothing will be different in the two places because of the difference in production costs. In America, clothing will be 2 times as expensive as food because it takes twice as much labor to produce a unit of clothing as it does to produce a unit of food. In Europe, clothing will be only $\frac{1}{3}$ as expensive as food.

After Trade. Now suppose that all tariffs are repealed and free trade is allowed. For simplicity, further assume that there are no transportation costs. What is the flow of goods when trade is opened up? Clothing is relatively more expensive in America (with a price ratio of 2 as compared to $\frac{1}{3}$), and food is relatively more expensive in Europe (with a price ratio of $\frac{3}{4}$ as compared to $\frac{1}{2}$). Given these relative prices, and with no tariffs or transportation costs, food will soon be shipped from America to Europe and clothing from Europe to America.

As European clothing penetrates the American market, American clothiers will find prices falling and profits shrinking, and they will begin to shut down their factories. By contrast, European farmers will find that the prices of foodstuffs begin to fall when American products hit the European markets; they will suffer losses, some will go bankrupt, and resources will be withdrawn from farming.

After all the adjustments to international trade have taken place, the prices of clothing and food must be equalized in Europe and America (just as the water in two connecting pipes must come to a common level once you remove the barrier between them). Without further knowledge about the exact supplies and demands, we cannot know the exact level to which prices will move. But we do know that the relative prices of food and clothing must lie somewhere between the European price ratio (which is $\frac{3}{4}$ for the ratio of food to clothing prices) and the American price ratio (which is $\frac{1}{2}$). Let us say that the final ratio is $\frac{2}{3}$, so 2 units of clothing trade for 3 units of food. For simplicity,

we measure prices in American dollars and assume that the free-trade price of food is \$2 per unit, which means that the free-trade price of clothing must be \$3 per unit.

With free trade, the regions have shifted their productive activities. America has withdrawn resources from clothing in favor of food, while Europe has contracted its farm sector and expanded its clothing manufacture. *Under free trade, countries shift production toward their areas of comparative advantage.*

The Economic Gains from Trade

What are the economic effects of opening up the two regions to international trade? America as a whole benefits from the fact that imported clothing costs less than clothing produced at home. Likewise, Europe benefits by specializing in clothing and consuming food that is less expensive than domestically produced food.

We can most easily reckon the gains from trade by calculating the effect of trade upon the real wages of workers. Real wages are measured by the quantity of goods that a worker can buy with an hour's pay. Using Table 18-2, we can see that the real wages after trade will be greater than the real wages before trade for workers in both Europe *and* America. For simplicity, assume that each worker buys 1 unit of clothing and 1 unit of food. Before trade, this bundle of goods costs an American worker 3 hours of work and a European worker 7 hours of work.

After trade has opened up, the price of clothing is \$3 per unit while the price of food is \$2 per unit. An American worker must still work 1 hour to buy a unit of food, because food is domestically produced; but at the price ratio of 2 to 3, the American worker need work only $1\frac{1}{2}$ hours to produce enough to buy 1 unit of European clothing. Therefore the bundle of goods costs the American worker $2\frac{1}{2}$ hours of work when trade is allowed—this represents an increase of 20 percent in the real wage of the American worker.

For European workers, a unit of clothing will still cost 4 hours of labor in a free-trade situation. To obtain a unit of food, however, the European worker need produce only $\frac{1}{3}$ of a unit of clothing (which requires $\frac{1}{3} \times 4$ hours of labor) and then trade that $\frac{1}{3}$ clothing unit for 1 unit of American food. The total European labor needed to obtain the bundle of consumption is then $4 + 2\frac{1}{3} = 6\frac{1}{3}$, which represents an

increase in real wages of about 5 percent over the no-trade situation.

When countries concentrate on their areas of comparative advantage under free trade, each country is better off. Compared to a no-trade situation, workers in each region can obtain a larger quantity of consumer goods for the same amount of work when they specialize in their areas of comparative advantage and trade their own production for goods in which they have a relative disadvantage.

Outsourcing as Another Kind of Trade

Recently, Americans have become concerned about outsourcing (sometimes also called “offshoring”). What exactly is the issue here? *Outsourcing* refers to locating services or production processes abroad. Prominent examples are telemarketing, medical diagnostics, publishing, web development, and engineering. These differ from the more conventional international trade in goods because they relate to services that were expensive to locate in foreign countries in an earlier era, whereas today, with rapid and low-cost communication, such processes can be economically located where costs are lower. Just as low-cost ocean shipping made possible greater international trade in grains in the nineteenth century, low-cost communication makes it possible to have Indian architects work on designs for New York firms today.

Many economists respond to outsourcing by arguing that it is just an extension of the principle of comparative advantage to more sectors. For example, when he was G. W. Bush’s chief economist, Greg Mankiw stated, “I think outsourcing is a growing phenomenon, but it’s something that we should realize is probably a plus for the economy in the long run.” His comment ignited a firestorm of controversy among both Republicans and Democrats, and one political figure called it “Alice in Wonderland economics.”

Most economists tend to agree with Mankiw that outsourcing is another example of comparative advantage at work. But there are policy consequences for governments. A careful analysis by Princeton economist (and adviser to Democratic presidents) Alan Blinder suggested the following advice for the country, and perhaps also for today’s students:

Rich countries such as the United States will have to reorganize the nature of work to exploit their big

advantage in non-tradable services: they are close to where the money is. That will mean, in part, specializing more in the delivery of services where personal presence is either imperative or highly beneficial. Thus, the U.S. work force of the future will likely have more divorce lawyers and fewer attorneys who write routine contracts, more internists and fewer radiologists, more salespeople and fewer typists. The market system is very good at making adjustments like these, even massive ones. It has done so before and will do so again. But it takes time and can move in unpredictable ways.

GRAPHICAL ANALYSIS OF COMPARATIVE ADVANTAGE

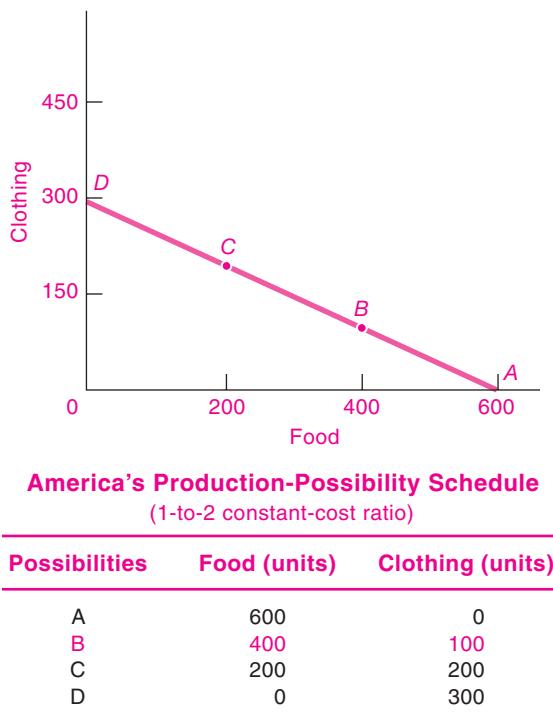
We can use the production-possibility frontier (*PPF*) to expand our analysis of comparative advantage. We will continue with the simple numerical example developed in this chapter, but the theory is equally valid in a competitive world with many different inputs.

America without Trade

Chapter 1 introduced the *PPF*, which shows the combinations of commodities that can be produced with a society’s given resources and technology. Using the production data shown in Table 18-2, and assuming that both Europe and America have 600 units of labor, we can easily derive each region’s *PPF*. The table that accompanies Figure 18-1 shows the possible levels of food and clothing that America can produce with its inputs and technology. Figure 18-1 plots the production possibilities; the green line *DA* shows America’s *PPF*. The *PPF* has a slope of $-\frac{1}{2}$, which represents the terms on which food and clothing can be substituted in production. In competitive markets with no international trade, the price ratio of food to clothing will also be one-half.

So far we have concentrated on production and ignored consumption. Note that if America is isolated from all international trade, it can consume only what it produces. Say that, for the incomes and demands in the marketplace, point *B* in Figure 18-1 marks America’s production and consumption in the absence of trade. Without trade, America produces and consumes 400 units of food and 100 units of clothing.

We can do exactly the same thing for Europe. But Europe’s *PPF* will look different from America’s

**FIGURE 18-1. American Production Data**

The constant-cost line *DA* represents America's domestic production-possibility frontier. America will produce and consume at *B* in the absence of trade.

because Europe has different efficiencies in producing food and clothing. Europe's price ratio is $\frac{3}{4}$, reflecting the relative cost of food and clothing in that region.

Opening Up to Trade

Now allow trade between the two regions. Food can be exchanged for clothing at some price ratio. We call the ratio of export prices to import prices the **terms of trade**. To indicate the trading possibilities, we put the two *PPFs* together in Figure 18-2. America's green *PPF* shows its domestic production possibilities, while Europe's blue *PPF* shows the terms on which it can domestically substitute food and clothing. Note that Europe's *PPF* is drawn closer to the origin than America's because Europe has lower productivities in both industries; it has an absolute disadvantage in the production of both food and clothing.

Europe need not be discouraged by its absolute disadvantage, however, for it is the difference in

relative productivities or comparative advantage that makes trade beneficial. The gains from trade are illustrated by the outer lines in Figure 18-2. If America could trade at Europe's pretrade relative prices, it could produce 600 units of food and move northwest along the outer blue line in Figure 18-2(a)—where the blue line represents the price ratio or terms of trade that are generated by Europe's *PPF*. Similarly, if Europe could trade at America's pretrade prices, Europe could specialize in clothing and move southeast along the green line in Figure 18-2(b)—where the green line is America's pretrade price ratio.

This leads to an important and surprising conclusion: Small countries have the most to gain from international trade. Small countries affect world prices the least and therefore can trade at world prices that are very different from domestic prices. Additionally, countries that are very different from other countries gain most, while large countries have the least to gain. (These points are raised in question 3 at the end of this chapter.)

Equilibrium Price Ratio. Once trade opens up, some set of prices must hold in the world marketplace depending upon the overall market supplies and demands. Without further information we cannot specify the exact price ratio, but we can determine what the price range will be. The prices must lie somewhere between the prices of the two regions. That is, we know that the relative price of food to clothing must lie somewhere in the range between $\frac{1}{2}$ and $\frac{3}{4}$.

The final price ratio will depend upon the relative demands for food and clothing. If food were very much in demand, the food price would be relatively high. If food demand were so high that Europe produced food as well as clothing, the price ratio would be that of Europe's pretrade relative prices, or $\frac{3}{4}$. On the other hand, if clothing demand were so strong that America produced clothing as well as food, the terms of trade would equal America's pretrade price ratio of $\frac{1}{2}$. If each region specializes completely in the area of its comparative advantage, with Europe producing only clothing and America producing only food, the price ratio will lie somewhere between $\frac{1}{2}$ and $\frac{3}{4}$. The exact ratio will depend on the strength of demand.

Assume now that the demands are such that the final price ratio is $\frac{2}{3}$, with 3 units of food selling for

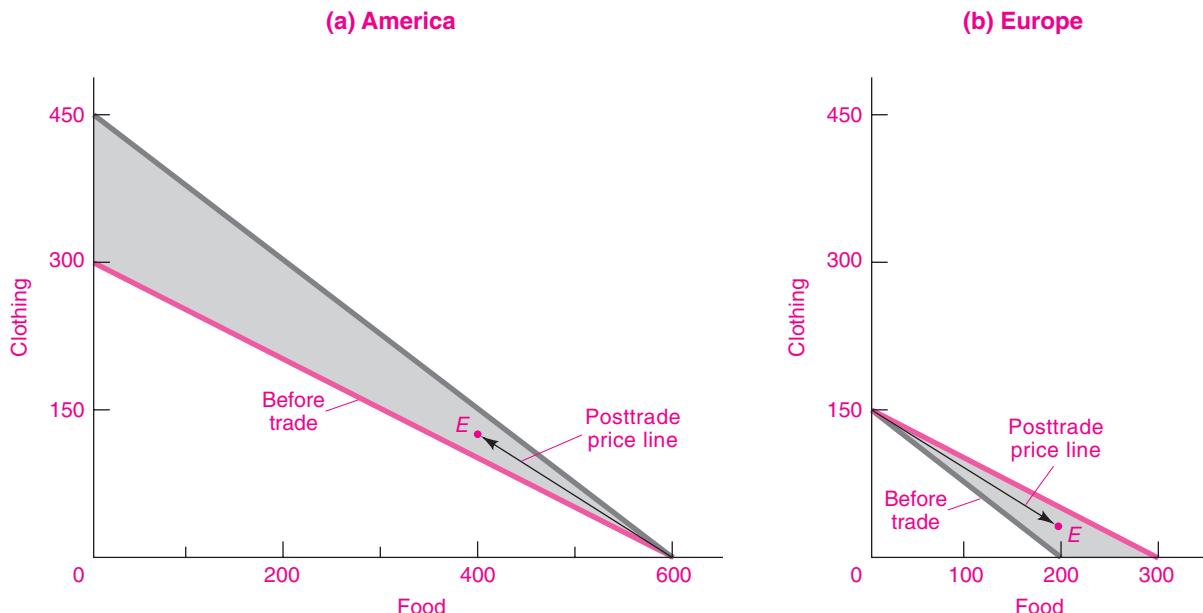


FIGURE 18-2. Comparative Advantage Illustrated

Through trade, both Europe and America improve their available consumption. If no trade is allowed, each region must be satisfied with its own production. It is therefore limited to its production-possibility curve, shown for each region as the line marked “Before trade.” After borders are opened and competition equalizes the relative prices of the two goods, the relative-price line will be as shown by the arrows. If each region is faced with prices given by the arrows, can you see why its consumption possibilities must improve?

2 units of clothing. With this price ratio, each region will then specialize—America in food and Europe in clothing—and export some of its production to pay for imports at the world price ratio of $\frac{1}{2}$.

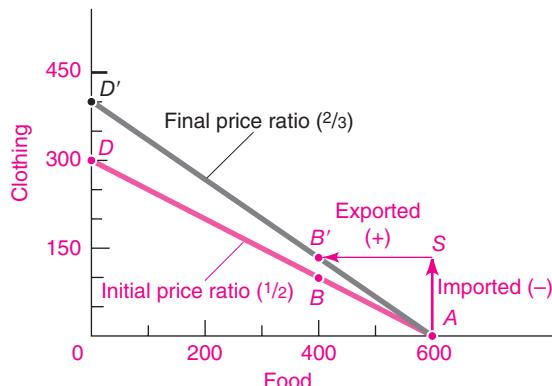
Figure 18-2 illustrates how trade will take place. Each region will face a consumption-possibility curve according to which it can produce, trade, and consume. *The consumption-possibility curve begins at the region's point of complete specialization and then runs out at the world price ratio of $\frac{1}{2}$.* Figure 18-2(a) shows America's consumption possibilities as a thin blue arrow with a slope of $-\frac{1}{2}$ coming out of its complete-specialization point at 600 units of food and no clothing. Similarly, Europe's posttrade consumption possibilities are shown in Figure 18-2(b) by the blue arrow running southeast from its point of complete specialization with a slope of $-\frac{1}{2}$.

The final outcome is shown by the points E in Figure 18-2. At this free-trade equilibrium, Europe specializes in producing clothing and America

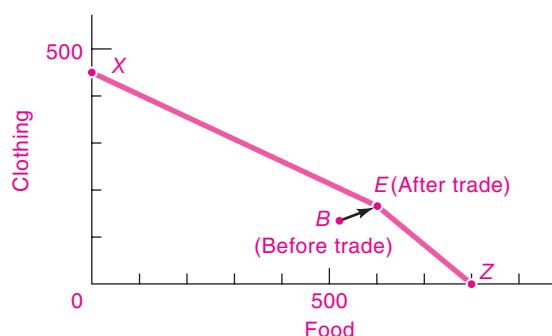
specializes in producing food. Europe exports $13\frac{1}{2}$ units of clothing for 200 units of America's food. Both regions are able to consume more than they could produce alone; both regions have benefited from international trade.

Figure 18-3 illustrates the benefits of trade for America. The green inner line shows the *PPF*, while the blue outer line shows the consumption possibilities at the world price ratio of $\frac{1}{2}$. The green arrows show the amounts exported and imported. America ends up at point B' . Through trade it moves along the blue line $D'A$ just as if a fruitful new invention had pushed out its *PPF*.

The lessons of this analysis are summarized in Figure 18-4. This figure shows the *world production frontier*. The world *PPF* represents the maximum output that can be obtained from the world's resources when goods are produced in the most efficient manner—that is, with the most efficient division of labor and regional specialization.

**FIGURE 18-3. America before and after Trade**

Free trade expands the consumption options of America. The green line DA represents America's production-possibility curve; the blue line $D'A$ is the new production-possibility curve when America is able to trade freely at the price ratio of $\frac{2}{3}$ and, in consequence, to specialize completely in the production of food (at A). The green arrows from S to B' and A to S show the amounts exported (+) and imported (-) by America. As a result of free trade, America ends up at B' , with more of both goods available than would be the case if it consumed only what it produced along DA .

**FIGURE 18-4. Free Trade Allows the World to Move to Its Production-Possibility Frontier**

We show here the effect of free trade from the viewpoint of the world as a whole. Before trade is allowed, each region is on its own national PPF . Because the no-trade equilibrium is inefficient, the world is inside its PPF at point B .

Free trade allows each region to specialize in the goods in which it has comparative advantage. As a result of efficient specialization, the world moves out to the efficiency frontier at point E .

The world PPF is built up from the two regional $PPFs$ in Figure 18-2 by determining the maximum level of world output that can be obtained from the individual regional $PPFs$. For example, the maximum quantity of food that can be produced (with no clothing production) is seen in Figure 18-2 to be 600 units in America and 200 units in Europe, for a world maximum of 800 units. This same point (800 food, 0 clothing) is then plotted in the world PPF in Figure 18-4. Additionally, we can plot the point (0 food, 450 clothing) in the world PPF by inspection of the regional $PPFs$. All the individual points in between can be constructed by a careful calculation of the maximum world outputs that can be produced if the two regions are efficiently specializing in the two goods.

Before opening up borders to trade, the world is at point B . This is an inefficient point—inside the world PPF —because regions have different levels of relative efficiency in different goods. After opening the borders to trade, the world moves to the free-trade equilibrium at E , where countries are specializing in their areas of comparative advantage.

Free trade in competitive markets allows the world to move to the frontier of its production-possibility curve.

EXTENSIONS TO MANY COMMODITIES AND COUNTRIES

The world of international trade consists of more than two regions and two commodities. However, the principles we explained above are essentially unchanged in realistic situations.

Many Commodities

When two regions or countries produce many commodities at constant costs, the goods can be arranged in order according to the comparative advantage or cost of each. For example, the commodities might be microprocessors, computers, aircraft, automobiles, wine, and croissants—all arranged in the comparative-advantage sequence shown in Figure 18-5. As you can see from the figure, of all the commodities, microprocessors are least expensive in America relative to the costs in Europe. Europe has its greatest comparative advantage in croissants. Two decades ago, America was dominant in the commercial-aircraft market, but



FIGURE 18-5. With Many Commodities, There Is a Spectrum of Comparative Advantages

Europe has now gained a substantial market share, so aircraft have been moving right on the line.

We can be virtually certain that the introduction of trade will cause America to produce and export microprocessors, while Europe will produce and export croissants. But where will the dividing line fall? Between aircraft and automobiles? Or wine and croissants? Or will the dividing line fall on one of the commodities rather than between them? Perhaps automobiles will be produced in both places.

You will not be surprised to find that the answer depends upon the demands and supplies of the different goods. We can think of the commodities as beads arranged on a string according to their comparative advantage; the strength of supply and demand will determine where the dividing line between American and European production will fall. An increased demand for microprocessors and computers, for example, would tend to raise the relative prices of American goods. This shift might lead America to specialize so much more in areas of its comparative advantage that it would no longer be profitable to produce in areas of comparative disadvantage, like automobiles.

Many Countries

What about the case of many countries? Introducing many countries need not change our analysis. As far as a single country is concerned, all the other nations can be lumped together into one group as “the rest of the world.” The advantages of trade have no special relationship to national boundaries. The principles already developed apply between groups of countries and, indeed, between regions within the same country. In fact, they are just as applicable to trade between our northern and southern states as to trade between the United States and Canada.

Triangular and Multilateral Trade

With many countries brought into the picture, it will generally be beneficial to engage in *triangular* or *multilateral trade* with many other countries. Bilateral trade between two countries is generally unbalanced.

Consider the simple example of triangular trade flows presented in Figure 18-6, where the arrows show the direction of exports. America buys consumer electronics from Japan, Japan buys oil and primary commodities from developing countries, and developing countries buy computers from America. In reality, trade patterns are more complex than this triangular example.

QUALIFICATIONS AND CONCLUSIONS

We have now completed our look at the elegant theory of comparative advantage. Its conclusions apply for any number of countries and commodities. Moreover, it can be generalized to handle many inputs, changing factor proportions, and diminishing returns. But we cannot conclude without noting two important qualifications to this elegant theory:

1. *Classical assumptions.* From a theoretical point of view, the major defect of comparative-advantage theory lies in its classical assumptions. This theory assumes a smoothly working competitive economy. But trade might lead to worsening environmental problems if there are local or global public goods (see Chapter 14 for a further discussion). Moreover, inefficiencies might arise in the presence of inflexible prices and wages, business cycles, and involuntary unemployment. When there are macroeconomic or microeconomic

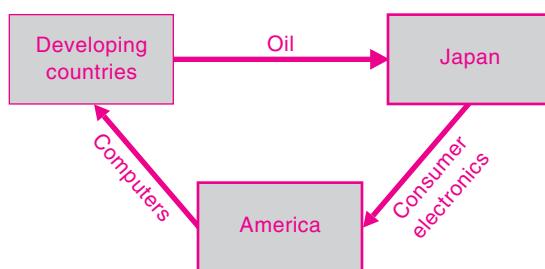


FIGURE 18-6. Triangular Trade Benefits All

In reality, international trade, like domestic trade, is many-sided.

market failures, trade might well push a nation *inside* its PPF. When the economy is in depression or the price system malfunctions because of environmental or other reasons, we cannot be sure that countries will gain from trade.

Given these reservations, there can be little wonder that the theory of comparative advantage sells at a big discount during business downturns. In the Great Depression of the 1930s, as unemployment soared and real outputs fell, nations built high tariff walls at their borders and the volume of foreign trade shrank sharply. Additionally, during the prosperous 1990s, free trade was increasingly attacked by environmental advocates, who saw it as a means of allowing companies to dump pollutants in oceans or in countries with lax regulations. Environmentalists were among the leading critics of the latest attempts to promote freer trade (see the section “Negotiating Free Trade” at the end of this chapter).

2. *Income distribution.* A second proviso concerns the impact on particular people, sectors, or factors of production. We showed above that opening a country to trade will raise a country’s national income. The country can consume more of all goods and services than would be possible if the borders were sealed to trade.

But this does not mean that *everyone* will benefit from trade, as shown by the Stolper-Samuelson theorem. We can illustrate this theorem using an example. Suppose that America has a relatively skilled labor force, while China has a relatively unskilled labor force. Moreover, suppose that skilled labor is used more heavily in aircraft, while unskilled labor is used more heavily in clothing. Now move from a situation of no trade to a situation of free trade. As in the example, we would expect that America will export aircraft and import clothing. The price of aircraft in America would rise, and the price of clothing would fall.

The interesting point is the impact on labor. As a result of the shift in domestic production, the demand for unskilled labor falls because of the decline in clothing prices and production, while the demand for skilled labor rises because of the rise in aircraft prices and production. In a world of flexible wages, this leads to a decline in the wages of unskilled labor and a rise in the wages of skilled labor in America. More generally, free trade tends to increase the prices of factors

that are intensive in exports and to reduce the prices of factors that are intensive in imports. (In a world with inflexible wages, it may lead to unemployment of unskilled workers, as our discussion of macroeconomics shows.)

Recent studies indicate that unskilled workers in high-income countries have suffered reductions in real wages in the last three decades because of the increased imports of goods from low-wage developing countries. Wage losses occur because imports of goods like clothing are produced by unskilled workers in developing countries. In a sense, these workers are close substitutes for the unskilled workers in the clothing industry of high-income countries. The increased international trade in clothing reduces the prices of clothing, and that tends to reduce the wages of unskilled workers in high-income countries.

The theory of comparative advantage shows that other sectors will gain more than the injured sectors will lose. Moreover, over long periods of time, those displaced from low-wage sectors eventually gravitate to higher-wage jobs. But those who are temporarily injured by international trade are genuinely harmed and are vocal advocates for protection and trade barriers.

Notwithstanding its limitations, the theory of comparative advantage is one of the deepest truths in all of economics. Nations that disregard comparative advantage pay a heavy price in terms of their living standards and economic growth.

C. PROTECTIONISM

Go back to the beginning of this chapter and reread the “Petition of the Candle Makers,” written by the French economist Frederic Bastiat to satirize solemn proposals to protect domestic goods from imports. Today, people often regard foreign competition with suspicion, and campaigns to “Buy American” sound patriotic.

Yet economists since the time of Adam Smith have marched to a different drummer. Economists generally believe that free trade promotes a mutually beneficial division of labor among nations; free and open trade allows *each* nation to expand its production and consumption possibilities, raising the

world's living standard. Protectionism prevents the forces of comparative advantage from working to maximum advantage.

This section reviews the economic arguments about protectionism.

SUPPLY-AND-DEMAND ANALYSIS OF TRADE AND TARIFFS

Free Trade vs. No Trade

The theory of comparative advantage can be illuminated through the analysis of supply and demand for goods in foreign trade. Consider the clothing market in America. Assume, for simplicity, that America is a small part of the market and therefore cannot affect the world price of clothing. (This assumption will allow us to analyze supply and demand very easily; the more realistic case in which a country can

affect world prices will be considered later in this chapter.)

Figure 18-7 shows the supply and demand curves for clothing in America. The demand curve of American consumers is drawn as DD and the domestic supply curve of American firms as SS . We assume that the price of clothing is determined in the world market and is equal to \$4 per unit. Although transactions in international trade are carried out in different currencies, for now we can simplify by converting the foreign supply schedule into a dollar supply curve by using the current exchange rate.

No-Trade Equilibrium. Suppose that transportation costs or tariffs for clothing were prohibitive (say, \$100 per unit of clothing). Where would the no-trade equilibrium lie? In this case, the American market for clothing would be at the intersection of *domestic*

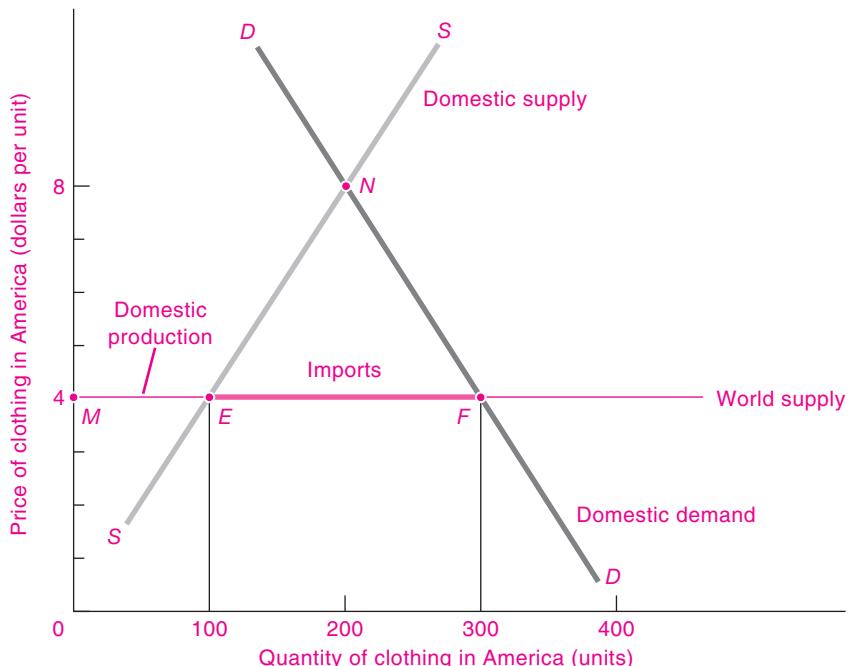


FIGURE 18-7. American Production, Imports, and Consumption under Free Trade

We see here the free-trade equilibrium in the market for clothing. America has a comparative disadvantage in clothing. Therefore, at the no-trade equilibrium at N , America's price would be \$8, while the world price is \$4.

Assuming that American demand does not affect the world price of \$4 per unit, the free-trade equilibrium comes when America produces ME (100 units) and imports the difference between domestic demand and domestic supply, shown as EF (or 200 units).

supply and demand, shown at point *N* in Figure 18-7. At this no-trade point, prices would be relatively high at \$8 per unit, and domestic producers would be meeting all the demand.

Free Trade. Now open the American clothing market to international trade. In the absence of transport costs, tariffs, and quotas, the price in America must be equal to the world price. Why? Because if the American price were above the Chinese price, sharp-eyed entrepreneurs would buy where clothing was cheap (China) and sell where clothing was expensive (America); China would therefore export clothing to America. Once trade flows fully adjusted to supplies and demands, the price in America would equal the world price level. (In a world with transportation and tariff costs, the price in America would equal the world price adjusted for these costs.)

Figure 18-7 illustrates how prices, quantities, and trade flows will be determined under free trade in our clothing example. The horizontal line at \$4 represents the supply curve for imports; it is horizontal, or perfectly price-elastic, because American demand is assumed to be too small to affect the world price of clothing.

Once trade opens up, imports flow into America, lowering the price of clothing to the world price of \$4 per unit. At that level, domestic producers will supply the amount *ME*, or 100 units, while at that price consumers will want to buy 300 units. The difference, shown by the heavy line *EF*, is the amount of clothing imports. Who decided that we would import just this amount of clothing and that domestic producers would supply only 100 units? A Chinese planning agency? A cartel of clothing firms? No, the amount of trade was determined by supply and demand.

Moreover, the level of prices in the no-trade equilibrium determined the direction of the trade flows. America's no-trade prices were higher than China's, so goods flowed into America. Remember this rule: *Under free trade, indeed in markets generally, goods flow uphill from low-price regions to high-price regions.* When markets are opened to free trade, clothing flows uphill from the lower-price Chinese market to the higher-price American market until the price levels are equalized.

Trade Barriers

For centuries, governments have used tariffs and quotas to raise revenues and influence the development

of individual industries. Since the eighteenth century—when the British Parliament attempted to impose tariffs on tea, sugar, and other commodities on its American colonies—tariff policy has proved fertile soil for revolution and political struggle.

We can use supply-and-demand analysis to understand the economic effects of tariffs and quotas. To begin with, note that a **tariff** is a tax levied on imports. A **quota** is a limit on the quantity of imports. The United States has quotas on many products, including textiles, watches, and cheeses.

Table 18-3 shows the average tariff rates for major countries in 2003. Note that tariffs vary widely for different goods in most countries. It would take deep study to understand why tariffs on imports of horses are zero while those on asses are 6.8 percent of value in the United States. On the other hand, it does not take much study to understand why textiles and steel have tight quotas or high tariffs, because these are industries with political clout in Congress or the White House.

Country or region	Average tariff rate, 2003 (%)
Hong Kong (China)	0.0
Switzerland	0.0
Japan	3.3
United States	3.9
Canada	4.2
European Union	4.4
Russia	11.3
China	12.0
Mexico	17.3
Pakistan	17.2
India	33.0
Iran	30.0
Average of major groups:	
Low-income countries	5.9
Middle-income countries	14.1

TABLE 18-3. Average Tariff Rates, 2003

Tariff rates vary widely among regions. The United States and regions like Singapore and Hong Kong (China) have low tariff rates today, although there are exceptions such as for textiles and steel. Countries like India and China continue to maintain protectionist trade barriers.

Source: World Trade Organization and government organizations.

Prohibitive Tariff. The easiest case to analyze is a *prohibitive tariff*—one that is so high that it chokes off all imports. Looking back at Figure 18-7, what would happen if the tariff on clothing were more than \$4 per unit (that is, more than the difference between America’s no-trade price of \$8 and the world price of \$4)? This would be a prohibitive tariff, shutting off all clothing trade. Any importer who buys clothing at the world price of \$4 would sell it in America at the no-trade price of \$8. But this price would not cover the cost of the good plus the tariff. Prohibitive tariffs thus kill off all trade.

Nonprohibitive Tariff. Lower tariffs (less than \$4 per unit of clothing) would injure but not kill off trade. Figure 18-8 shows the equilibrium in the clothing market with a \$2 tariff. Again assuming no transportation costs, a \$2 tariff means that foreign clothing will sell in America for \$6 per unit (equal to the \$4 world price plus the \$2 tariff).

The equilibrium result of a \$2 tariff is that domestic consumption (or quantity demanded) is lowered from 300 units in the free-trade equilibrium to 250 units after the tariff is imposed, the amount of domestic production is raised by 50 units, and the quantity of imports is lowered by 100 units. This example summarizes the economic impact of tariffs:

A tariff will tend to raise price, lower the amounts consumed and imported, and raise domestic production of the covered good.

Quotas. Quotas have the same qualitative effect as tariffs. A prohibitive quota (one that prevents all imports) is equivalent to a prohibitive tariff. The price and quantity would move back to the no-trade equilibrium at *N* in Figure 18-8. A less stringent quota might limit imports to 100 clothing units; this quota would equal the heavy line *HJ* in Figure 18-8. A quota of 100 units would lead to the same equilibrium price and output as did the \$2 tariff.

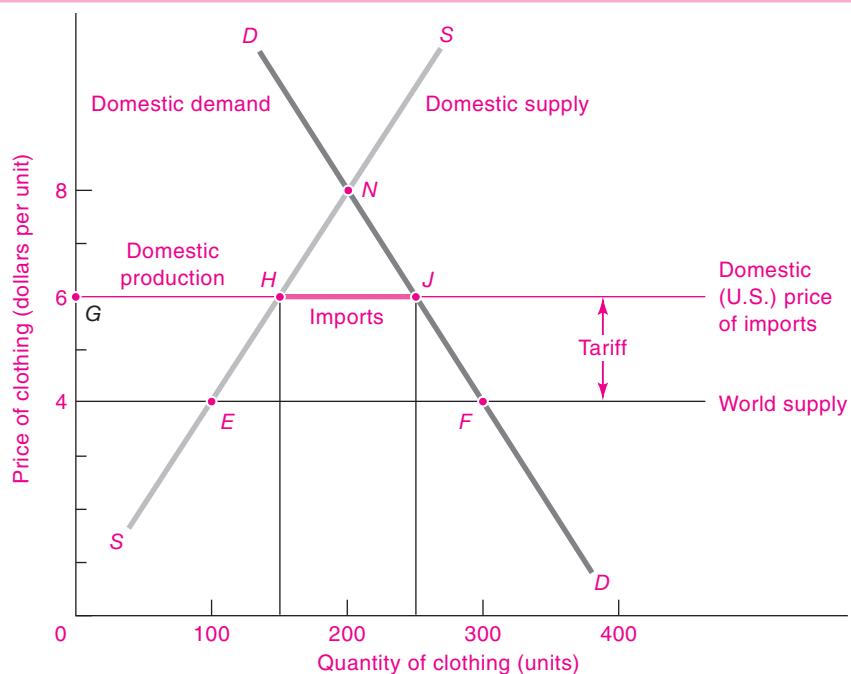


FIGURE 18-8. Effect of a Tariff

A tariff lowers imports and consumption and raises domestic production and price. Starting from the free-trade equilibrium in Fig. 18-7, America now puts a \$2 tariff on clothing imports. The price of Chinese clothing imports rises to \$6 (including the tariff).

The market price rises from \$4 to \$6, so the total amount demanded falls. Imports shrink from 200 to 100 units, while domestic production rises from 100 to 150 units.

Although there is no essential difference between tariffs and quotas, some subtle differences do exist. A tariff gives revenue to the government, perhaps allowing other taxes to be reduced and thereby offsetting some of the harm done to consumers in the importing country. A quota, on the other hand, puts the profit from the resulting price difference into the pocket of the importers or exporters lucky enough to get a permit or import license. They can afford to use the proceeds to wine, dine, or even bribe the officials who give out import licenses.

Because of these differences, economists generally regard tariffs as the lesser evil. However, if a government is determined to impose quotas, it should auction off the scarce import-quota licenses. An auction will ensure that the government rather than the importer gets the revenue from the scarce right to import; in addition, the bureaucracy will not be tempted to allocate quota rights by bribery, friendship, or nepotism.

Transportation Costs. What of transportation costs? The cost of moving bulky and perishable goods has the same effect as tariffs, reducing the extent of beneficial regional specialization. For example, if it costs \$2 per unit to transport clothing from China to the United States, the supply-and-demand equilibrium would look just like Figure 18-8, with the American price \$2 above the Chinese price.

But there is one difference between protection and transportation costs: Transport costs are imposed by nature—by oceans, mountains, and rivers—whereas restrictive tariffs are squarely the responsibility of nations. Indeed, one economist called tariffs “negative railroads.” Imposing a tariff has the same economic impact as throwing sand in the engines of vessels that transport goods to our shores from other lands.

The Economic Costs of Tariffs

What happens when America puts a tariff on clothing, such as the \$2 tariff shown in Figure 18-8? There are three effects: (1) The domestic producers, operating under a price umbrella provided by the tariff, can expand production; (2) consumers are faced with higher prices and therefore reduce their consumption; and (3) the government gains tariff revenue.

Tariffs create economic inefficiencies. When tariffs are imposed, the economic loss to consumers

exceeds the revenue gained by the government plus the extra profits earned by producers.

Diagrammatic Analysis. Figure 18-9 shows the economic cost of a tariff. The supply and demand curves are identical to those in Figure 18-8, but three areas are highlighted. (1) Area *B* is the tariff revenue collected by the government. It is equal to the amount of the tariff times the units of imports and totals \$200. (2) The tariff raises the price in domestic markets from \$4 to \$6, and producers increase their output to 150. Hence total profits rise by \$250, shown by area *LEHM* and equal to \$200 on old units plus an additional \$50 on the 50 new units. (3) Finally, note that a tariff imposes a heavy cost on consumers. The total consumer-surplus loss is given by area *LMJF* and is equal to \$550.

The overall social impact is, then, a gain to producers of \$250, a gain to the government of \$200, and a loss to consumers of \$550. The net social cost (counting each of these dollars equally) is therefore \$100. We can reckon this as equal to the sum of *A* and *C*. The interpretation of these areas is important:

- Area *A* is the net loss that comes because domestic production is more costly than foreign production. When the domestic price rises, businesses are thereby induced to increase the use of relatively costly domestic capacity. They produce output up to the point where the marginal cost is \$6 per unit instead of up to \$4 per unit under free trade. Firms reopen inefficient old factories or work existing factories extra shifts. From an economic point of view, these plants have a comparative disadvantage because the new clothing produced by these factories could be produced more cheaply abroad. The new social cost of this inefficient production is area *A*, equal to \$50.
- In addition, there is a net loss to the country from the higher price, shown by area *C*. This is the loss in consumer surplus that cannot be offset by business profits or tariff revenue. This area represents the economic cost incurred when consumers shift their purchases from low-cost imports to high-cost domestic goods. This area is also equal to \$50.

Hence, the total social loss from the tariff is \$100, calculated either way.

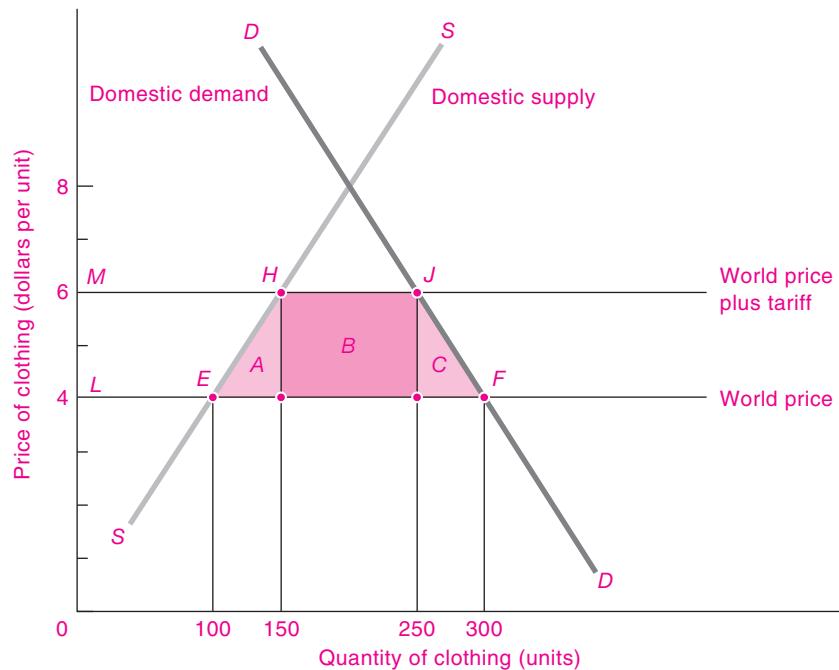


FIGURE 18-9. Economic Cost of a Tariff

Imposing a tariff raises revenues and leads to inefficiency. We see the impact of the tariff as three effects. Rectangle *B* is the tariff revenue gained by the government. Triangle *A* is the excess cost of production by firms producing under the umbrella of the tariff. Triangle *C* is the net loss in consumer surplus from the inefficiently high price. Areas *A* and *C* are the irreducible inefficiencies caused by the tariff.

Figure 18-9 illustrates one feature that is important in understanding the politics and history of tariffs. When a tariff is imposed, part of the economic impact comes because tariffs redistribute income from consumers to the protected domestic producers and workers. In the example shown in Figure 18-9, areas *A* and *C* represent efficiency losses from inefficiently high domestic production and inefficiently low consumption, respectively. Under the simplifying assumptions used above, the efficiency losses sum up to \$100. The redistribution involved is much larger, however, equaling \$200 raised in tariff revenues levied upon consumers of the commodity plus \$250 in higher profits. Consumers will be unhappy about the higher product cost, while domestic producers and workers in those firms will benefit. We can see why battles over import restrictions generally center more on the redistributive gains and losses than on the issues of economic efficiency.

Imposing a tariff has three effects: It encourages inefficiently high domestic production; it raises prices, thus inducing consumers to reduce their purchases of the tariffed good below efficient levels; and it raises revenues for the government. Only the first two of these necessarily impose efficiency costs on the economy.



The Cost of Textile Protection

Let's flesh out this analysis by examining the effects of a particular tariff, one on clothing.

Today, tariffs on imported textiles and apparel are among the highest levied by the United States. How do these high tariffs affect consumers and producers?

To begin with, the tariffs raise domestic clothing prices. Because of the higher prices, many factories, which would otherwise be bankrupt in the face of a declining comparative advantage in textiles, remain open. They are just barely

profitable, but they manage to eke out enough sales to continue domestic production. Domestic employment in textiles exceeds the free-trade situation, although—because of pressure from foreign competition—textile wages are among the lowest of any manufacturing industry.

From an economic point of view, the nation is wasting resources in textiles. These workers, materials, and capital would be more productively used in other sectors—perhaps in aircraft or financial services or Internet commerce. The nation's productive potential is lower because it keeps factors of production working in an industry in which it has lost its comparative advantage.

Consumers, of course, pay for this protection of the textile industry with higher prices. They get less satisfaction from their incomes than they would if they could buy textiles from Korea, China, or Indonesia at prices that exclude the high tariffs. Consumers are induced to cut back on their clothing purchases, channelling funds into food, transportation, and recreation, whose relative prices are lowered by the tariffs.

Finally, the government gets revenues from tariffs on textiles. These revenues can be used to buy public goods or to reduce other taxes, so (unlike the consumer loss or the productive inefficiency) this effect is not a real social burden.

THE ECONOMICS OF PROTECTIONISM

Having examined the impact of tariffs on prices and quantities, we now turn to an analysis of the arguments for and against protectionism. The arguments for tariff or quota protection against the competition of foreign imports take many different forms. Here are the main categories: (1) noneconomic arguments that suggest it is desirable to sacrifice economic welfare in order to subsidize other national objectives, (2) arguments that are based on a misunderstanding of economic logic, and (3) analyses that rely on market power or macroeconomic imperfections.

Noneconomic Goals

If you are ever on a debating team given the assignment of defending free trade, you will strengthen your case at the beginning by conceding that there is more to life than economic welfare. A nation surely should not sacrifice its liberty, culture, and human rights for a few dollars of extra income.

The U.S. semiconductor industry provides a useful example here. In the 1980s, the Defense Department claimed that without an independent semiconductor industry, the military would become excessively dependent on Japanese and other foreign suppliers for chips to use in high-technology weaponry. This led to an agreement to protect the industry. Economists were skeptical about the value of this approach. Their argument did not question the goal of national security. Rather, it focused on the efficiency of the means of achieving the desired result. They thought that protection was more expensive than a policy targeting the domestic industry, perhaps a program to buy a minimum number of high-quality chips.

National security is not the only noneconomic goal in trade policy. Countries may desire to preserve their cultural traditions or environmental conditions. France argued that its citizens need to be protected from “uncivilized” American movies. The fear is that the French film industry could be drowned by the new wave of stunt-filled, high-budget Hollywood thrillers. As a result, France has maintained strict quotas on the number of U.S. movies and television shows that can be imported.

Unsound Grounds for Tariffs

Mercantilism. To Abraham Lincoln has been attributed the remark, “I don’t know much about the tariff. I do know that when I buy a coat from England, I have the coat and England has the money. But when I buy a coat in America, I have the coat and America has the money.”

This reasoning represents an age-old fallacy typical of the so-called mercantilist writers of the seventeenth and eighteenth centuries. They considered a country fortunate which sold more goods than it bought, because such a “favorable” balance of trade meant that gold would flow into the country to pay for its export surplus.

The mercantilist argument confuses means and ends. Accumulating gold or other monies will not improve a country’s living standard. Money is worthwhile not for its own sake but for what it will buy from other countries. Most economists today therefore reject the idea that raising tariffs to run a trade surplus will improve a country’s economic welfare.

Tariffs for Special Interests. The single most important source of pressure for protective tariffs is powerful

special-interest groups. Firms and workers know very well that a tariff on their particular products will help *them* even if it imposes costs on others. Adam Smith understood this point well when he wrote:

To expect freedom of trade is as absurd as to expect Utopia. Not only the prejudices of the public, but what is much more unconquerable, the private interests of many individuals, irresistibly oppose it.

If free trade is so beneficial to the nation as a whole, why do the proponents of protectionism continue to wield such a disproportionate influence on legislatures? The few who benefit gain much from specific protection and therefore devote large sums to lobbying politicians. By contrast, individual consumers are only slightly affected by the tariff on one product; because losses are small and widespread, individuals have little incentive to spend resources expressing an opinion on every tariff case. A century ago, outright bribery was used to buy the votes necessary to pass tariff legislation. Today, powerful political action committees (PACs), financed by labor or business, round up lawyers and drum up support for tariffs or quotas on textiles, lumber, steel, sugar, and other goods.

If political votes were cast in proportion to total economic benefit, nations would legislate most tariffs out of existence. But each dollar of economic interests does not get proportional representation. It is much harder to persuade consumers about the benefits of free trade than it is to organize a few companies or labor unions to argue against “cheap Chinese labor.” In every country, the special interests of protected firms and workers are the tireless enemies of free trade.

A dramatic case is the U.S. quota on sugar, which benefits a few producers while costing American consumers over \$1 billion a year. The average consumer is probably unaware that the sugar quota costs about a penny a day per person, so there is little incentive to lobby for free trade in sugar.

Competition from Cheap Foreign Labor. Of all the arguments for protection, the most persistent is that free trade exposes U.S. workers to competition from low-wage foreign labor. The only way to preserve high U.S. wages, so the argument goes, is to protect domestic workers by keeping out or putting high

tariffs on goods produced in low-wage countries. An extreme version of this contention is that under free trade U.S. wages would decline to the low level of foreign wages. This point was trumpeted by presidential candidate Ross Perot during the debates over the North American Free Trade Agreement (NAFTA) when he argued:

Philosophically, [NAFTA] is wonderful, but realistically it will be bad for our country. That thing is going to create a giant sucking sound in the United States at a time when we need jobs coming in, not jobs going out. Mexican wages will come up to \$7½ an hour and our wages will come down to \$7½ an hour.

This argument sounds plausible, but it is all wrong because it ignores the principle of comparative advantage. The reason American workers have higher wages is that they are on average more productive. If America's wage is 5 times that in Mexico, it is because the marginal product of American workers is on average 5 times that of Mexican workers. Trade flows according to comparative advantage, not wage rates or absolute advantage.

Having shown that the nation gains from importing the goods produced by “cheap foreign labor” in which it has a comparative disadvantage, we should not ignore the impacts that trade may have on particular firms and workers. Remember the Stolper-Samuelson theorem explained above. If America has a comparative disadvantage in industries like textiles or toys, and these industries are intensive in unskilled labor, reducing trade barriers will tend to reduce the wages of unskilled labor in America. There may also be temporary effects on workers whose wages drop while they look for alternative jobs. The difficulties of displaced workers will be greater when the overall economy is depressed or when the local labor markets have high unemployment. Over the long run, labor markets will reallocate workers from declining to advancing industries, but the transition may be costly for many people.

In summary:

The cheap-foreign-labor argument is flawed because it ignores the theory of comparative advantage. A country will benefit from trade even though its wages are far above those of its trading partners. High wages come from high efficiency, not from tariff protection.

Retaliatory Tariffs. While many people would agree that a world of free trade would be the best of all possible worlds, they note that this is not the world we live in. They reason, “As long as other countries impose import restrictions or otherwise discriminate against our products, we have no choice but to play the protection game in self-defense. We’ll go along with free trade only as long as it is fair trade. But we insist on a level playing field.” On several occasions in the 1990s, the United States went to the brink of trade wars with Japan and China, threatening high tariffs if the other country did not stop some objectionable trade practice.

Those who advocate this approach argue that it can beat down the walls of protectionism in other countries. This rationale was described in an analysis of protection in the *Economic Report of the President*:

Intervention in international trade . . . even though costly to the U.S. economy in the short run, may, however, be justified if it serves the strategic purpose of increasing the cost of interventionist policies by foreign governments. Thus, there is a potential role for carefully targeted measures . . . aimed at convincing other countries to reduce their trade distortions.

While potentially valid, this argument should be used with great caution. Just as threatening war leads to armed conflict as often as to arms control, protectionist bluffs may end up hurting the bluffer as well as the opponent. Historical studies show that retaliatory tariffs usually lead other nations to raise their tariffs still higher and are rarely an effective bargaining chip for multilateral tariff reduction.

Import Relief. In the United States and other countries, firms and workers that are injured by foreign competition attempt to get protection in the form of tariffs or quotas. Today, relatively little direct tariff business is conducted on the floor of Congress. Congress realized that tariff politics was too hot to handle and has set up specialized agencies to investigate and rule on complaints. Generally, a petition for relief is analyzed by the U.S. Department of Commerce and the U.S. International Trade Commission. Relief measures include the following actions:

- The *escape clause* was popular in earlier periods. It allows temporary import relief (tariffs, quotas, or export quotas negotiated with other countries) when an industry has been “injured” by imports. Injury occurs when the output, employment, and

profits in a domestic industry have fallen while imports have risen.

- *Antidumping tariffs* are levied when foreign countries sell in the United States at prices below average costs or at prices lower than those in the home market. When dumping is found, a “dumping duty” is placed on the imported good.
- *Countervailing duties* are imposed to offset the cost advantage for imports that arises when foreigners subsidize exports to the United States. They have become the most popular form of import relief and have been pursued in hundreds of cases.

What is the justification for such measures? Import relief sounds reasonable, but it actually is completely counter to the theory of comparative advantage. That theory says that an industry which cannot compete with foreign firms ought to be injured by imports. *From an economic vantage point, less productive industries are actually being killed off by the competition of more productive domestic industries.*

This sounds ruthless indeed. No industry willingly dies. No region gladly undergoes conversion to new industries. Often the shift from old to new industries involves considerable unemployment and hardship. The weak industry and region feel they are being singled out to carry the burden of progress.

Potentially Valid Arguments for Protection

Finally, we can consider three arguments for protection that may have true economic merit:

- Tariffs may shift the terms of trade in a country’s favor.
- Temporary tariff protection for an “infant industry” with growth potential may be efficient in the long run.
- A tariff may under certain conditions help reduce unemployment.

The Terms-of-Trade or Optimal-Tariff Argument. One valid argument for imposing tariffs is that doing so will shift the terms of trade in a country’s favor and against foreign countries. The phrase *terms of trade* refers to the ratio of export prices to import prices. The idea is that when a large country levies tariffs on its imports, the reduced demand for the good in world markets will lower the equilibrium price and thereby reduce the pretariff cost of the

good to the country. Such a change will improve the country's terms of trade and increase domestic real income. The set of tariffs that maximizes domestic real income is called the *optimal tariff*.

The terms-of-trade argument goes back over 150 years to the free-trade proponent John Stuart Mill. It is the only argument for tariffs that is valid under conditions of full employment and perfect competition. Suppose that the U.S. imposes an "optimal" tariff on imported oil. The tariff will increase the price of domestic oil and will reduce the world demand for oil. The world market price of oil will therefore be bid down. So part of the tariff actually falls on the oil producer. (We can see that a very small country could not use this argument, since it cannot affect world prices.)

Have we not therefore found a theoretically secure argument for tariffs? The answer would be yes if we could forget that this is a "beggar-thy-neighbor" policy and could ignore the reactions of other countries. But other countries are likely to react. After all, if the United States were to impose an optimal tariff of 30 percent on its imports, why should the European Union and Japan not put 30 or 40 percent tariffs on their imports? In the end, as every country calculated and imposed its own nationalistic optimal tariff, the overall level of tariffs might spiral upward in the tariff version of an arms race.

Ultimately, such a situation would surely not represent an improvement of either world or individual economic welfare. When all countries impose optimal tariffs, it is likely that *everyone's* economic welfare will decline as the impediments to free trade become larger. All countries are likely to benefit if all countries abolish trade barriers.

Tariffs for Infant Industries. In his famous *Report on Manufactures* (1791), Alexander Hamilton proposed to encourage the growth of manufacturing by protecting "infant industries" from foreign competition. According to this doctrine, which received the cautious support of free-trade economists like John Stuart Mill and Alfred Marshall, there are lines of production in which a country could have a comparative advantage if only they could get started.

Such infant industries would not be able to survive the rough treatment by larger bullies in the global marketplace. With some temporary nurturing, however, they might grow up to enjoy economies of

mass production, a pool of skilled labor, inventions well adapted to the local economy, and the technological efficiency typical of many mature industries. Although protection will raise prices to the consumer at first, the mature industry would become so efficient that cost and price would actually fall. A tariff is justified if the benefit to consumers at that later date would be more than enough to make up for the higher prices during the period of protection.

This argument must be weighed cautiously. Historical studies have turned up some genuine cases of protected infant industries that grew up to stand on their own feet. And studies of successful newly industrialized countries (such as Singapore and Taiwan) show that they have often protected their manufacturing industries from imports during the early stages of industrialization. But subsidies will be a more efficient and transparent way of nurturing young industries. In fact, the history of tariffs reveals many cases like steel, sugar, and textiles in which perpetually protected infants have not shed their diapers after these many years.



Brazil's Tragic Protection of Its Computer Industry

Brazil offers a striking example of the pitfalls of protectionism. In 1984, Brazil passed

a law actually banning most foreign computers. The idea was to provide a protected environment in which Brazil's own infant computer industry could develop. The law was vigorously enforced by special "computer police" who would search corporate offices and classrooms looking for illegal imported computers.

The results were startling. Technologically, Brazilian-made computers were years behind the fast-moving world market, and consumers paid 2 or 3 times the world price—when they could get them at all. At the same time, because Brazilian computers were so expensive, they could not compete on the world market, so Brazilian computer companies could not take advantage of economies of scale by selling to other countries. The high price of computers hurt competitiveness in the rest of the economy as well. "We are effectively very backward because of this senseless nationalism," said Zelia Cardoso de Mello, Brazil's economy minister in 1990. "The computer problem effectively blocked Brazilian industry from modernizing."

The combination of pressure from Brazilian consumers and businesses and U.S. demands for open markets

forced Brazil to drop the ban on imported computers in 1992. Within a year, electronics stores in São Paulo and Rio de Janeiro were filled with imported laptop computers, laser printers, and cellular telephones, and Brazilian companies could begin to exploit the computer revolution. Each country and each generation learns anew the lessons of comparative advantage.

Tariffs and Unemployment. Historically, a powerful motive for protection has been the desire to increase employment during a period of recession or stagnation. Protection creates jobs by raising the price of imports and diverting demand toward domestic production; Figure 18-8 demonstrates this effect. As domestic demand increases, firms will hire more workers and unemployment will fall. This too is a beggar-thy-neighbor policy, for it raises domestic demand at the expense of output and employment in other countries.

However, while economic protection may raise employment, it does not constitute an effective program to pursue high employment, efficiency, and stable prices. Macroeconomic analysis shows that there are better ways of reducing unemployment than by imposing import protection. By the appropriate use of monetary and fiscal policy, a country can increase output and lower unemployment. Moreover, the use of general macroeconomic policies will allow workers displaced from low-productivity jobs in industries losing their comparative advantage to move to high-productivity jobs in industries enjoying a comparative advantage.

This lesson was amply demonstrated in the 1990s. From 1991 to 1999, the United States created 16 million net new jobs while maintaining open markets and low tariffs; its trade deficit increased sharply during this period. By contrast, the countries of Europe created virtually no new jobs while moving toward a position of trade surpluses.

Tariffs and import protection are an inefficient way to create jobs or to lower unemployment. A more effective way to increase productive employment is through domestic monetary and fiscal policy.

Other Barriers to Trade

While this chapter has mainly spoken of tariffs, most points apply equally well to any other impediments to

trade. Quotas have much the same effects as tariffs, for they prevent the comparative advantages of different countries from determining prices and outputs in the marketplace. In recent years, countries have negotiated quotas with other countries. The United States, for example, forced Japan to put “voluntary” export quotas on automobiles and negotiated similar export quotas on televisions, shoes, and steel.

We should also mention the so-called nontariff barriers (or NTBs). These consist of informal restrictions or regulations that make it difficult for countries to sell their goods in foreign markets. For example, American firms complained that Japanese regulations shut them out of the telecommunications, tobacco, and construction industries.

How important are the nontariff barriers relative to tariffs? Economic studies indicate that nontariff barriers were actually more important than tariffs during the 1960s; in recent years, they have effectively doubled the protection found in the tariff codes. In a sense, nontariff barriers have been substitutes for more conventional tariffs as the latter have been reduced.

MULTILATERAL TRADE NEGOTIATIONS

Given the tug-of-war between the economic benefits of free trade and the political appeal of protection, which force has prevailed? The history of U.S. tariffs, shown in Figure 18-10, has been bumpy. For most of American history, the United States was a high-tariff nation. The pinnacle of protectionism came after the infamous Smoot-Hawley tariff of 1930, which was opposed by virtually every American economist yet sailed through Congress.

The trade barriers erected during the Great Depression helped raise prices and exacerbated economic distress. In the trade wars of the 1930s, countries attempted to raise employment and output by raising trade barriers at the expense of their neighbors. Nations soon learned that at the end of the tariff-retaliation game, all were losers.

Negotiating Free Trade

At the end of World War II, the international community established a number of institutions to promote peace and economic prosperity through cooperative policies.

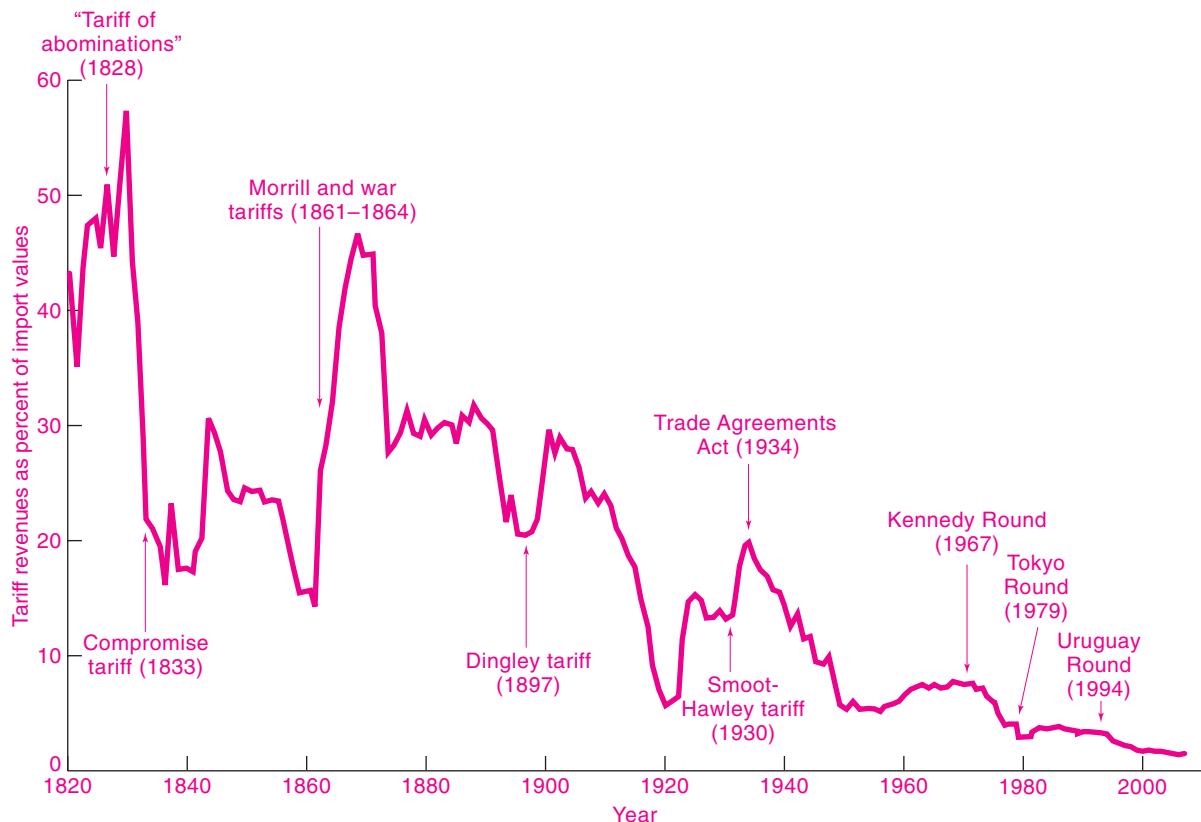


FIGURE 18-10. America Was Historically a High-Tariff Nation

Tariffs were high for most of our nation's history, but trade negotiations since the 1930s have lowered tariffs significantly.

Multilateral Agreements. One of the most successful multilateral agreements was the General Agreement on Tariffs and Trade (GATT). Its provisions were incorporated into the World Trade Organization (WTO) at the beginning of 1995. Their charters speak of raising living standards through “substantial reduction of tariffs and other barriers to trade and the elimination of discriminatory treatment in international commerce.” As of 2008, the WTO had 153 member countries, which accounted for 90 percent of international trade.

Among the principles underlying the WTO are (1) countries should work to lower trade barriers; (2) all trade barriers should be applied on a non-discriminatory basis across nations (i.e., all nations should enjoy “most-favored-nation” status); (3) when a country increases its tariffs above agreed-upon

levels, it must compensate its trading partners for the economic injury; and (4) trade conflicts should be settled by consultations and arbitration.

Multilateral trade negotiations successfully lowered trade barriers in the half-century following World War II. The latest successful negotiations were the Uruguay Round, which included 123 countries and was completed in 1994. In 2001, countries launched a new round in Doha, Qatar. Among the items on the agenda are agriculture, intellectual property rights, and the environment. The new negotiations have been controversial both among developing countries, which believe that the rich countries are protecting agriculture too heavily, and among antiglobalization groups, which argue that growing trade is hurting the environment. In the face of deep divisions, the Doha Round has made no progress as of 2008.

Regional Approaches. Over the last few years, governments have taken a number of steps to promote free trade or to broaden regional markets. Among the most important were the following.

The most controversial proposal for lowering trade barriers was the North American Free Trade Agreement (NAFTA), which was hotly debated and passed by Congress by a close vote in 1993. Mexico is the third-largest trading partner of the United States, and most U.S.-Mexico trade is in manufactured goods. NAFTA not only allows goods to pass tariff-free across the borders but also liberalizes regulations on investments by the United States and Canada in Mexico. Proponents of the plan argued that it would allow a more efficient pattern of specialization and would enable U.S. firms to compete more effectively against firms in other countries; opponents, particularly labor groups, argued that it would increase the supply of goods produced by low-skilled labor and thereby depress the wages of workers in the affected industries.

Economists caution, however, that regional trading agreements like NAFTA can cause inefficiency if they exclude potential trading countries. They point to the stagnation in the Caribbean countries, which were excluded from the free-trade provisions of NAFTA, as a cautionary example of the dangers of the regional approach.

The most far-reaching trade accord has been the movement toward a single market among the major

European countries. Since World War II, the nations of the European Union (EU) have developed a common market with minimal barriers to international trade or movement of factors of production. The first step involved eliminating all internal tariff and regulatory barriers to trade and labor and capital flows. The most recent step was the introduction of a common currency (the Euro) for most of the members of the EU. European unification is one of history's most eloquent tributes to the power of an idea—the idea that free and open trade promotes economic efficiency and technological advance.

Appraisal

After World War II, policymakers around the world believed firmly that free trade was essential for world prosperity. These convictions translated into several successful agreements to lower tariffs, as Figure 18-10 shows. The free-trade philosophy of economists and market-oriented policymakers has been severely tested by periods of high unemployment, by exchange-rate disturbances, and recently by antiglobalization forces. Nevertheless, most countries have continued the trend toward increased openness and outward orientation.

Economic studies generally show that countries have benefited from lower trade barriers as trade flows and living standards have grown. But the struggle to preserve open markets is constantly tested as the political and economic environment changes.



SUMMARY

A. The Nature of International Trade

1. Specialization, division of labor, and trade increase productivity and consumption possibilities. The gains from trade hold among nations as well as within a nation. Engaging in international exchange is more efficient than relying only on domestic production. International trade differs from domestic trade because it broadens the market, because trade takes place among sovereign nations, and because countries usually have their own monies which must be converted using foreign exchange rates.
2. Diversity is the fundamental reason that nations engage in international trade. Within this general principle,

we see that trade occurs (a) because of differences in the conditions of production, (b) because of decreasing costs (or economies of scale), and (c) because of diversity in tastes.

B. Comparative Advantage among Nations

3. Recall that trade occurs because of differences in the conditions of production or diversity in tastes. The foundation of international trade is the Ricardian principle of comparative advantage. The principle of comparative advantage holds that each country will benefit if it specializes in the production and export of those goods that it can produce at relatively low cost.

Conversely, each country will also benefit if it imports those goods which it produces at relatively high cost. This principle holds even if one region is absolutely more or less productive than another in all commodities. As long as there are differences in *relative* or *comparative* efficiencies among countries, every country must enjoy a comparative advantage or a comparative disadvantage in the production of some goods.

4. The law of comparative advantage predicts more than just the geographic pattern of specialization and the direction of trade. It also demonstrates that countries are made better off and that real wages (or, more generally, total national income) are improved by trade and the resulting enlarged world production. Quotas and tariffs, designed to “protect” workers or industries, will lower a nation’s total income and consumption possibilities.
5. Even with many goods or many countries, the same principles of comparative advantage apply. With many commodities, we can arrange products along a continuum of comparative advantage, from relatively more efficient to relatively less efficient. With many countries, trade may be triangular or multilateral, with countries having large bilateral (or two-sided) surpluses or deficits with other individual countries.

C. Protectionism

6. Completely free trade equalizes prices of tradeable goods at home with those in world markets. Under trade, goods flow uphill from low-price to high-price markets.

7. A tariff raises the domestic prices of imported goods, leading to a decline in consumption and imports along with an increase in domestic production. Quotas have very similar effects and may, in addition, lower government revenues.
8. A tariff causes economic waste. The economy suffers losses from decreased domestic consumption and from the wasting of resources on goods lacking comparative advantage. The losses generally exceed government revenues from the tariff.
9. Most arguments for tariffs simply rationalize special benefits to particular pressure groups and cannot withstand economic analysis. Three arguments that can stand up to careful scrutiny are the following: (a) The terms-of-trade or optimal tariff can in principle raise the real income of a large country at the expense of its trading partners. (b) In a situation of less-than-full employment, tariffs might push an economy toward fuller employment, but monetary or fiscal policies could attain the same employment goal with fewer inefficiencies than this beggar-thy-neighbor policy. (c) Sometimes, infant industries may need temporary protection in order to realize their true long-run comparative advantages.
10. The principle of comparative advantage must be qualified if markets malfunction because of unemployment or exchange-market disturbances. Moreover, individual sectors or factors may be injured by trade if imports lower their returns. Opening up to trade may hurt the factors that are most embodied in imported goods.

CONCEPTS FOR REVIEW

Principles of International Trade

- absolute and comparative advantage (or disadvantage)
- principle of comparative advantage
- economic gains from trade
- triangular and multilateral trade
- world vs. national PPFs

consumption vs. production possibilities with trade

Stolper-Samuelson theorem

Economics of Protectionism

- price equilibrium with and without trade

tariff, quota, nontariff barriers

effects of tariffs on price, imports, and domestic production

mercantilist, cheap-foreign-labor, and retaliatory arguments

the optimal tariff, unemployment, and infant-industry exceptions

FURTHER READING AND INTERNET WEBSITES

Further Reading

The theory of comparative advantage was discovered and discussed by David Ricardo in *Principles of Political Economy and Taxation* (1819, various publishers).

This is online at several sites, including www.econlib.org/library/Ricardo/ricP.html. A classic review of the debate about free trade is Jagdish Bhagwati, *Protectionism* (MIT Press, Cambridge, Mass., 1990). Some of the best popular writing

on international economics is found in *The Economist*, which is also available at www.economist.com.

Mankiw's remarks on outsourcing, as well as some reactions, can be found at www.cnn.com/2004/US/02/12/bush.outsourcing/. Blinder's article, "Offshoring: The Next Industrial Revolution?" appeared in *Foreign Affairs*, March–April 2006, and is available at www.foreignaffairs.org/.

Websites

The World Bank (www.worldbank.org) has information on its programs and publications at its site, as does the International Monetary Fund, or IMF (www.imf.org). The United Nations website has links to most international

institutions and their databases (www.unsystem.org). Another good source of information about high-income countries is the Organisation for Economic Cooperation and Development, or OECD (www.oecd.org). U.S. trade data are available at www.census.gov.

You can find information on many countries through their statistical offices. A compendium of national agencies is available at www.census.gov/main/www/stat_int.html.

One of the best sources for policy writing on international economics is www.iie.com/homepage.htm, the website of the Peterson Institute for International Economics.

QUESTIONS FOR DISCUSSION

1. State whether or not each of the following is correct and explain your reasoning. If the quotation is incorrect, provide a corrected statement.
 - a. "We Mexicans can never compete profitably with the Northern colossus. Her factories are too efficient, she has too many computers and machine tools, and her engineering skills are too advanced. We need tariffs, or we can export nothing!"
 - b. "If American workers are subjected to the unbridled competition of cheap Mexican labor, our real wages must necessarily fall drastically."
 - c. "The principle of comparative advantage applies equally well to families, cities, and states as it does to nations and continents."
 - d. The quotation from Ross Perot on page 356.
2. Reconstruct Figure 18-1 and its accompanying table to show the production data for Europe; assume that Europe has 600 units of labor and that labor productivities are those given in Table 18-2.
3. What if the data in Table 18-2 changed from (1, 2; 3, 4) to (1, 2; 2, 4)? Show that all trade is killed off. Use this to explain the adage "*Vive la différence!*" (freely translated as "Let diversity thrive!"). Why do the largest gains in trade flow to small countries whose pretrade prices are very different from prevailing world prices?
4. *Follow-up to question 3:* Suppose that the data in Table 18-2 pertain to a newly industrialized country (NIC) and America. What are the gains from trade between the two countries? Now suppose that NIC adopts American technology and has production possibilities identical to those in the American column of Table 18-2. What will happen to international trade? What will happen to NIC's living standards and real wages? What will happen to America's living standards? Is there a lesson here for the impact of converging economies on trade and welfare?
5. A U.S. senator wrote the following: "Trade is supposed to raise the incomes of all nations involved—or at least that is what Adam Smith and David Ricardo taught us. If our economic decline has been caused by the economic growth of our competitors, then these philosophers—and the entire discipline of economics they founded—have been taking us on a 200-year ride." Explain why the first sentence is correct. Also explain why the second sentence does not follow from the first. Can you give an example of how economic growth of Country J could lower the standard of living in Country A? (*Hint:* The answer to question 4 will help uncover the fallacy in the quotation.)
6. Modern protectionists have used the following arguments for protecting domestic industries against foreign competition:
 - a. In some situations, a country can improve its standard of living by imposing protection if no one else retaliates.
 - b. Wages in China are a tiny fraction of those in the United States. Unless we limit the imports of Chinese manufactures, we face a future in which our trade deficit continues to rise under the onslaught of competition from low-wage workers.
 - c. A country might be willing to accept a small drop in its living standard to preserve certain industries that it deems necessary for national security, such as supercomputers or oil, by protecting them from foreign competition.
 - d. *For those who have studied macroeconomics:* If inflexible wages and prices or an inappropriate exchange rate leads to recession and high unemployment,

tariffs might increase output and lower the unemployment rate.

In each case, relate the argument to one of the traditional defenses of protectionism. State the conditions under which it is valid, and decide whether you agree with it.

7. The United States has had quotas on steel, shipping, automobiles, textiles, and many other products. Economists estimate that by auctioning off the quota rights,

the Treasury would gain at least \$10 billion annually. Use Figure 18-9 to analyze the economics of quotas as follows: Assume that the government imposes a quota of 100 on imports, allocating the quota rights to importing countries on the basis of last year's imports. What would be the equilibrium price and quantity of clothing? What would be the efficiency losses from quotas? Who would get revenue rectangle *B*? What would be the effect of auctioning off the quota rights?

PART FIVE

Macroeconomics: Economic Growth and Business Cycles

Overview of Macroeconomics

19



The whole purpose of the economy is production of goods or services for consumption now or in the future. I think the burden of proof should always be on those who would produce less rather than more, on those who would leave idle people or machines or land that could be used. It is amazing how many reasons can be found to justify such waste: fear of inflation, balance-of-payments deficits, unbalanced budgets, excessive national debt, loss of confidence in the dollar.

James Tobin,
National Economic Policy

Are jobs plentiful or hard to find? Are real wages and living standards growing rapidly, or are consumers struggling to make ends meet as price inflation reduces real wages? Is there a period of financial exuberance with stock prices rising rapidly? Or is the central bank using monetary policy to fight off the effects of falling housing prices and a financial crisis? What are the impacts of globalization and foreign trade on domestic employment and output? These questions are central to macroeconomics, which is the subject of the following chapters.

Macroeconomics is the study of the behavior of the economy as a whole. It examines the forces that affect firms, consumers, and workers in the aggregate. It contrasts with **microeconomics**, which studies individual prices, quantities, and markets.

Two central themes will run through our survey of macroeconomics:

- The short-term fluctuations in output, employment, financial conditions, and prices that we call the *business cycle*
- The longer-term trends in output and living standards known as *economic growth*

The development of macroeconomics was one of the major breakthroughs of twentieth-century economics, leading to a much better understanding of how to combat periodic economic crises and how to stimulate long-term economic growth. In response to the Great Depression, John Maynard Keynes developed his revolutionary theory, which helped explain the forces producing economic fluctuations and suggested how governments can

control the worst excesses of the business cycle. At the same time, economists have endeavored to understand the mechanics of long-term economic growth.

Macroeconomic issues dominated the U.S. political and economic agenda for much of the last century. In the 1930s, when production, employment, and prices collapsed in the United States and across much of the industrial world, economists and political leaders wrestled with the calamity of the Great Depression. During the Vietnam War in the 1960s and the energy crises of the 1970s, the burning issue was “stagflation,” a combination of slow growth and rising prices. The 1990s witnessed a period of rapid growth, falling unemployment, and stable prices—years when everything went right, labeled by some as “the fabulous decade.” Then asset-market bubbles burst twice in the first decade of the 2000s. The first shock was a sharp decline in the prices of technology stocks in 2000, and this was followed by a sharp decline in housing prices after 2007. The 2007–2009 housing-price decline produced a profound financial crisis and led to a deep and long recession.

Sometimes, macroeconomic failures raise life-and-death questions for countries and even for ideologies. The communist leaders of the former Soviet Union proclaimed that they would overtake the West economically. History proved that to be a hollow promise, as Russia, a country teeming with natural resources and military might, was unable to produce adequate butter for its citizens along with the guns for its imperial armies. Eventually, macroeconomic failures brought down the communist regimes of the Soviet Union and Eastern Europe and convinced people of the economic superiority of private markets as the best approach to encouraging rapid economic growth.

This chapter will serve as an introduction to macroeconomics. It presents the major concepts and shows how they apply to key historical and policy questions of recent years. But this introduction is only a first course to whet the appetite. Not until you have mastered all the chapters in Parts Five through Seven can you fully enjoy the rich macroeconomic banquet that has been a source of both inspiration for economic policy and continued controversy among macroeconomists.

A. KEY CONCEPTS OF MACROECONOMICS

THE BIRTH OF MACROECONOMICS

The 1930s marked the first stirrings of the science of macroeconomics, founded by John Maynard Keynes as he tried to understand the economic mechanism that produced the Great Depression. After World War II, reflecting both the increasing influence of Keynesian views and the fear of another depression, the U.S. Congress formally proclaimed federal responsibility for macroeconomic performance. It enacted the landmark Employment Act of 1946, which stated:

The Congress hereby declares that it is the continuing policy and responsibility of the federal government to use all practicable means consistent with its needs and obligations . . . to promote maximum employment, production, and purchasing power.

For the first time, Congress affirmed the government’s role in promoting output growth, fostering employment, and maintaining price stability. The Employment Act usefully frames the three central questions of macroeconomics:

1. *Why do output and employment sometimes fall, and how can unemployment be reduced?* All market economies show patterns of expansion and contraction known as *business cycles*. The latest business-cycle recession in the United States occurred after a severe financial-market crisis that began in 2007. Housing and stock prices fell sharply, and banks tightened credit and lending. As a result, output and employment fell sharply. Political leaders around the world used the tools of monetary and fiscal policy to reduce unemployment and stimulate economic activity.

From time to time countries experience high unemployment that persists for long periods, sometimes as long as a decade. Such a period occurred in the United States during the Great Depression, which began in 1929. In the following years, unemployment rose to almost one-quarter of the workforce, while industrial production fell by one-half. One of the deepest

and most prolonged economic downturns of the modern era came in Japan, which experienced declining prices and was unable to shake off high unemployment and slow economic growth after 1990.

Macroeconomics studies the sources of persistent unemployment and high inflation. Having considered the symptoms, macroeconomists suggest possible remedies, such as using monetary policy to alter interest rates and credit conditions or using fiscal instruments such as taxes and spending. The lives and fortunes of millions of people depend upon whether economists find correct diagnoses for major macroeconomic ailments—and upon whether governments apply the right medicine at the right time.

2. *What are the sources of price inflation, and how can it be kept under control?* A market economy uses prices as a yardstick to measure economic values and conduct business. When prices are rising—a phenomenon we call *inflation*—the price yardstick loses its value. During periods of high inflation, people may get confused about relative prices and make mistakes in their spending and investment decisions. Tax burdens may rise. Households on fixed incomes find that inflation is eating away at their real incomes.

Macroeconomic policy has increasingly emphasized low and stable inflation as a key goal. Many countries set “inflation targets” for their economic policy, with targets often being in the range from 1 to 3 percent per year. Except for brief spikes, the United States has succeeded in containing inflation over the last two decades, with an average inflation rate of 3 percent per year for the consumer price index. Many countries have not been so successful. Formerly socialist countries like Russia and many Latin American and developing countries experienced inflation rates of 50, 100, or 1000 percent per year in the last two decades. The inflationary record in the last few years was in troubled Zimbabwe, where inflation was around 20,000,000 percent per year in 2008. A chicken that cost 10 thousand Zimbabwean dollars at the beginning of the year would cost 10 trillion Zimbabwean dollars at the end! Why was the United States able to contain the inflationary tiger, while Zimbabwe failed to do so? Macroeconomics can suggest the proper

role of monetary and fiscal policies, of exchange-rate systems, and of an independent central bank in containing inflation.

3. *How can a nation increase its rate of economic growth?*

The single most important goal of macroeconomics concerns a nation’s long-term economic growth. This refers to the growth in the per capita output of a country. Such growth is the central factor in determining the growth in real wages and living standards. Most countries of North America and Western Europe have enjoyed rapid economic growth for two centuries, and residents in these countries have high average incomes. Over the last five decades, Asian countries such as Japan, South Korea, and Taiwan produced dramatic gains in living standards for their peoples. China’s growth has similarly been outstanding in recent years. A few countries, particularly those of sub-Saharan Africa, have suffered declining per capita output and living standards.

Nations want to know the ingredients in a successful growth recipe. Economic historians have found that the key factors in long-term economic growth include reliance on well-regulated private markets for most economic activity, stable macroeconomic policy, high rates of saving and investment, openness to international trade, and accountable and noncorrupt governing institutions.

All economies face inevitable tradeoffs among these goals. Increasing the rate of growth of output over the long run may require greater investment in education and capital, but higher investment requires lower current consumption of items like food, clothing, and recreation. Additionally, policymakers are sometimes forced to rein in the economy through macroeconomic policies when it grows too fast in order to prevent rising inflation or when financial conditions exhibit irrational exuberance.

There are no magic formulas for ensuring low and stable inflation, high employment, and rapid growth. Macroeconomists have vigorous debates about both the goals and the appropriate policies for reaching the goals. But sound macroeconomic policies are essential if a country wishes to achieve its economic objectives in the most effective manner.



The Patron Saint of Macroeconomics

Every discussion of macroeconomic policy must begin with John Maynard Keynes. Keynes (1883–1946) was a many-sided genius who won eminence in the fields of mathematics, philosophy, and literature. In addition, he found time to run a large insurance company, advise the British treasury, help govern the Bank of England, edit a world-famous economics journal, collect modern art and rare books, start a repertory theater, and marry a leading Russian ballerina. He was also an investor who knew how to make money by shrewd speculation, both for himself and for his college, King's College, Cambridge.

His principal contribution, however, was his invention of a new way of looking at macroeconomics and macroeconomic policy. Before Keynes, most economists and policymakers accepted the highs and lows of business cycles as being as inevitable as the tides. These long-held views left them helpless in the face of the Great Depression of the 1930s. But Keynes took an enormous intellectual leap in his 1936 book, *The General Theory of Employment, Interest, and Money*. He made a twofold argument: First, he argued that it is possible for high unemployment and underutilized capacity to persist in market economies. In addition, he argued that government fiscal and monetary policies can affect output and thereby reduce unemployment and shorten economic downturns.

These propositions had an explosive impact when Keynes first introduced them, engendering much controversy and dispute. In the years after World War II, Keynesian economics came to dominate macroeconomics and government policy. Since then, new developments incorporating supply factors, expectations, and alternative views of wage and price dynamics have undermined the earlier Keynesian consensus. While few economists now believe that government action can eliminate business cycles, as Keynesian economics once seemed to promise, neither economics nor economic policy has been the same since Keynes's great discovery.

OBJECTIVES AND INSTRUMENTS OF MACROECONOMICS

Having surveyed the principal issues of macroeconomics, we now turn to a discussion of the major goals and instruments of macroeconomic policy. How do economists evaluate the success of an economy's overall performance? What are the tools that

Objectives
Output: High level and rapid growth of output
Employment: High level of employment with low involuntary unemployment
Stable prices
Instruments
Monetary policy: Buying and selling bonds, regulating financial institutions
Fiscal policy: Government expenditures Taxation

TABLE 19-1. Goals and Instruments of Macroeconomic Policy

The top part of the table displays the major goals of macroeconomic policy. The lower half shows the major instruments or policy measures available to modern economies. Policymakers change the instruments of policy to affect the pace and direction of economic activity.

governments can use to pursue their economic goals? Table 19-1 lists the major objectives and instruments of macroeconomic policy.

Measuring Economic Success

The major macroeconomic goals are a high level and rapid growth of output, low unemployment, and stable prices. We will use this section both to define the major macroeconomic terms and to discuss their importance. A more detailed treatment of the data of macroeconomics is postponed to the next chapter. Some key data are provided in the appendix to this chapter.

Output. The ultimate objective of economic activity is to provide the goods and services that the population desires. What could be more important for an economy than to produce ample shelter, food, education, and recreation for its people?

The most comprehensive measure of the total output in an economy is the **gross domestic product** (GDP). GDP is the measure of the market value of all



FIGURE 19-1. Growth Rate of U.S. Real Gross Domestic Product, 1929–2008

Real GDP is the most comprehensive measure of an economy's output. This figure shows the rate of growth from one year to the next. Note the string of negative growth rates in the Great Depression of the 1930s. Also, we see the Great Moderation of the last few years, in which output was less volatile than in earlier periods.

Source: U.S. Bureau of Economic Analysis at www.bea.gov. Shaded regions are major economic downturns.

final goods and services—beer, cars, rock concerts, donkey rides, and so on—produced in a country during a year. There are two ways to measure GDP. *Nominal GDP* is measured in actual market prices. *Real GDP* is calculated in constant or invariant prices (where we measure the number of cars times the prices of cars in a given year such as 2000).

Real GDP is the most closely watched measure of output; it serves as the carefully monitored pulse of a nation's economy. Figure 19-1 shows the growth rate of real GDP in the United States since 1929. The growth rate is defined as

$$\begin{aligned} \text{\% growth rate of real GDP in year } t \\ = 100 \times \frac{\text{GDP}_t - \text{GDP}_{t-1}}{\text{GDP}_{t-1}} \end{aligned}$$

For example, real GDP in 2006 was \$11,294.8 billion and in 2007 was \$11,523.9 billion (both in

2000 prices). A calculator will show that the growth of real GDP in 2007 was 2.0 percent over the year. It is worthwhile making sure you can replicate this calculation. Note the sharp economic decline during the Great Depression of the 1930s, the boom during World War II, and the recessions in 1974, 1982, 1991, and 2008.

Despite the short-term fluctuations seen in business cycles, advanced economies generally exhibit a steady long-term growth in real GDP and an improvement in living standards; this process is known as *economic growth*. The American economy has proved itself a powerful engine of progress over a period of more than a century, as shown by the growth in potential output.

Potential GDP represents the maximum sustainable level of output that the economy can produce. When an economy is operating at its

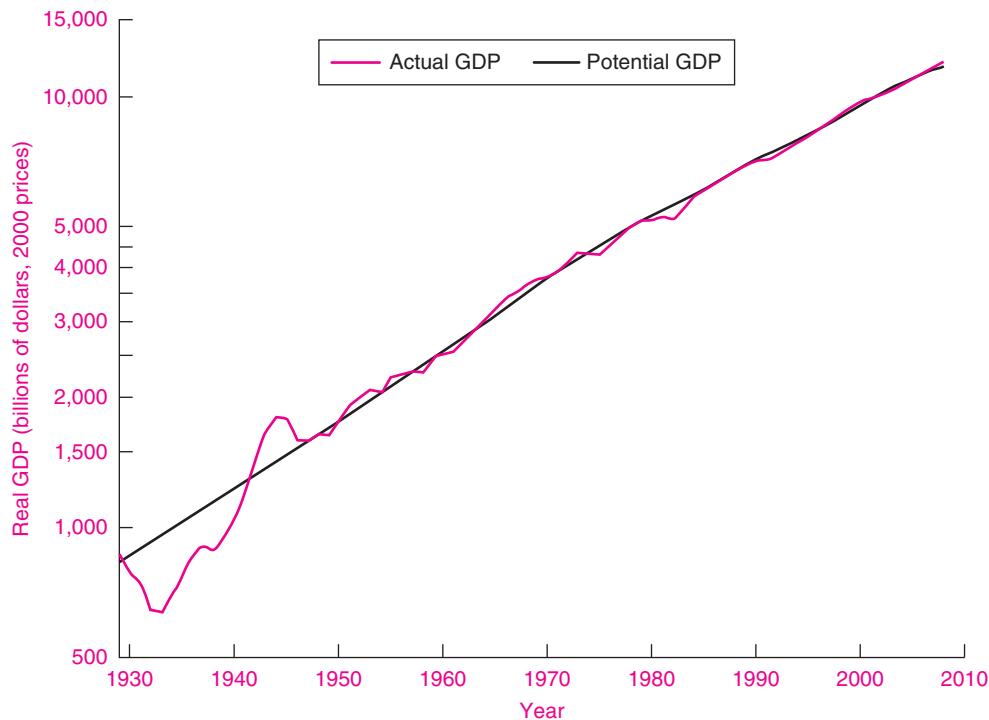


FIGURE 19-2. Actual and Potential GDP in the United States

Business cycles occur when actual output departs from its potential. The smooth blue line shows potential or trend output over the period 1929–2008. Potential output has grown about 3.4 percent annually. Note the large gap between actual and potential output during the Great Depression of the 1930s.

Source: U.S. Bureau of Economic Analysis, Congressional Budget Office, and authors' estimates. Note that actual GDP is directly estimated from underlying data while potential output is an analytical concept derived from actual GDP and unemployment data.

potential, there are high levels of utilization of the labor force and the capital stock. When output rises above potential output, price inflation tends to rise, while a below-potential level of output leads to high unemployment.

Potential output is determined by the economy's productive capacity, which depends upon the inputs available (capital, labor, land, etc.) and the economy's technological efficiency. Potential GDP tends to grow steadily because inputs like labor and capital and the level of technology change quite slowly over time. By contrast, actual GDP is subject to large business-cycle swings if spending patterns change sharply.

During business downturns, actual GDP falls below its potential, and unemployment rises. In 1982, for example, the U.S. economy produced about

\$400 billion less than its potential output. This represented \$5000 lost per family during a single year. A *recession* is a period of significant decline in total output, income, and employment, usually lasting more than a few months and marked by widespread contractions in many sectors of the economy. A severe and protracted downturn is called a *depression*. Output can be temporarily above its potential during booms and wartime as capacity limits are strained, but the high utilization rates may bring rising inflation and are usually brought to an end by monetary or fiscal policy.

Figure 19-2 shows the estimated potential and actual output for the period 1929–2008. Note how large the gap between actual and potential output was during the Great Depression of the 1930s.

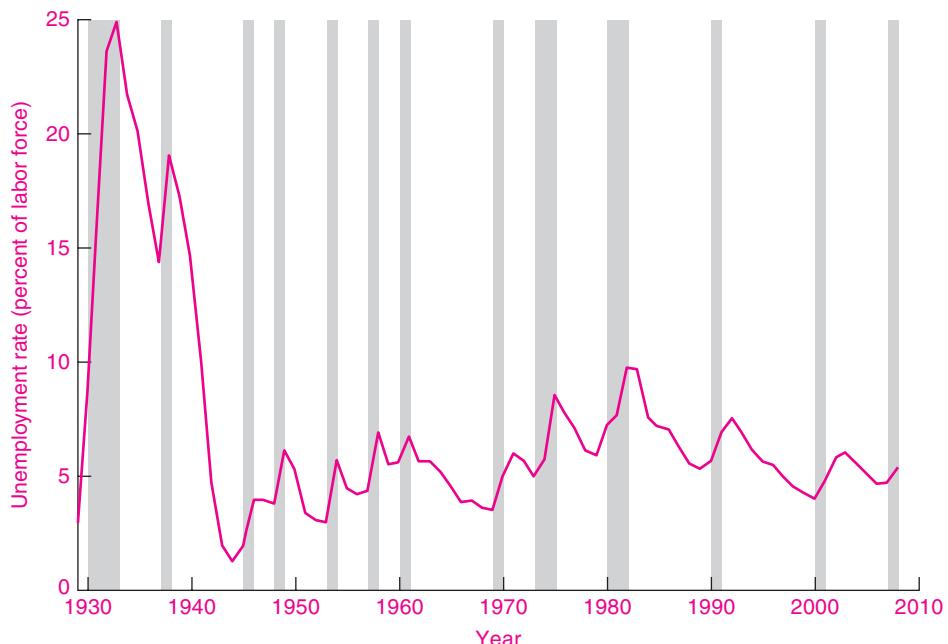


FIGURE 19-3. Unemployment Rises in Recessions, Falls during Expansions

The unemployment rate measures the fraction of the labor force that is looking for work but cannot find work. Unemployment rises in business-cycle downturns and falls during expansions. Shaded regions are NBER recessions.

Source: U.S. Bureau of Labor Statistics at www.bea.gov.

High Employment, Low Unemployment. Of all the macroeconomic indicators, employment and unemployment are most directly felt by individuals. People want to be able to get high-paying jobs without searching or waiting too long, and they want to have job security and good benefits. In macroeconomic terms, these are the objectives of *high employment*, which is the counterpart of *low unemployment*. Figure 19-3 shows trends in unemployment over the last eight decades. The **unemployment rate** on the vertical axis is the percentage of the labor force that is unemployed. The labor force includes all employed persons and those unemployed individuals who are seeking jobs. It excludes those without work who are not looking for jobs.

The unemployment rate tends to reflect the state of the business cycle: when output is falling, the demand for labor falls and the unemployment rate rises. Unemployment reached epidemic proportions in the Great Depression of the 1930s, when as much as one-quarter of the workforce was idled. Since World War II, unemployment in the United States

has fluctuated but has avoided the high rates associated with depressions.

Price Stability. The third macroeconomic objective is *price stability*. This is defined as a low and stable inflation rate.

To track prices, government statisticians construct **price indexes**, or measures of the overall price level. An important example is the **consumer price index** (CPI), which measures the trend in the average price of goods and services bought by consumers. We will generally denote the overall price level by the letter *P*.

Economists measure price stability by looking at **inflation**, or the **rate of inflation**. The inflation rate is the percentage change in the overall level of prices from one year to the next. For example, the CPI was 201.6 in 2006 and 207.3 in 2007. The inflation-rate calculation is just like the growth-rate calculation above:

$$\text{Rate of inflation in year } t = 100 \times \frac{P_t - P_{t-1}}{P_{t-1}}$$

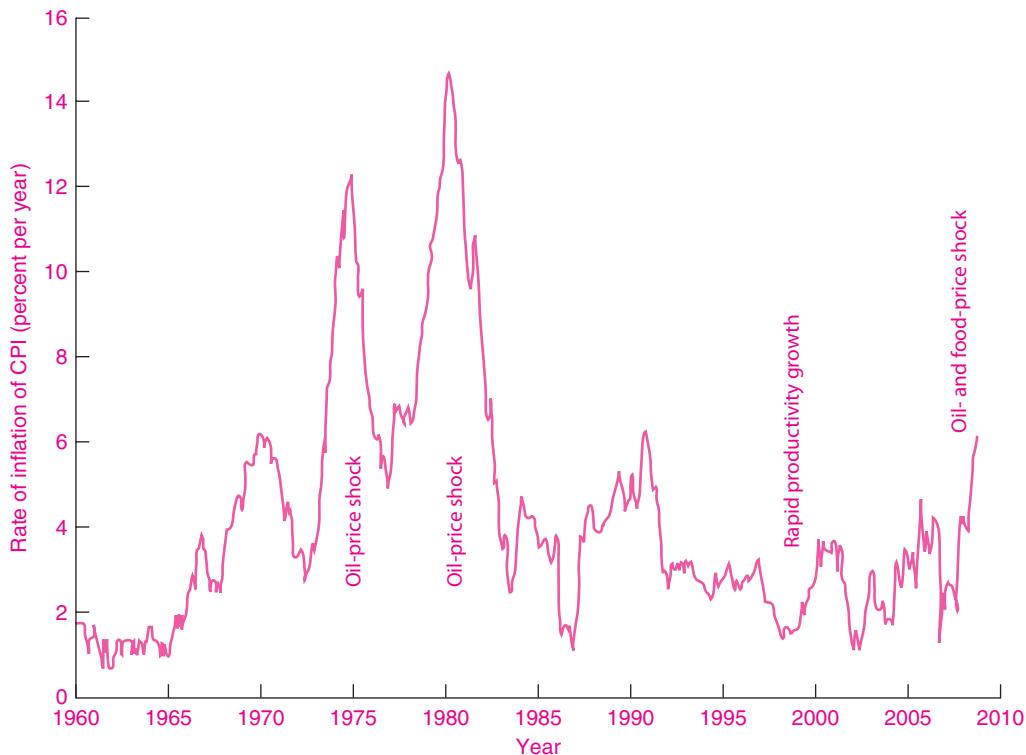


FIGURE 19-4. U.S. Consumer Price Inflation, 1960–2008

The rate of inflation measures the rate of change of prices from one year to the next; here we see the rate of inflation as measured by the consumer price index (CPI). Most inflationary episodes have been associated with shocks to oil or food prices. Note that inflation has moved in a narrow corridor since the mid-1980s.

Source: U.S. Bureau of Labor Statistics. Data show rate of inflation from 12 months earlier.

We thus calculate the inflation rate for 2007 as

$$\begin{aligned}\text{Rate of inflation in 2007} &= 100 \times \frac{207.3 - 201.6}{201.6} \\ &= 2.8\% \text{ per year}\end{aligned}$$

Figure 19-4 shows the inflation rate for the CPI from 1960 to 2008. Since the end of the inflationary period in the early 1980s, inflation has averaged 3 percent per year through 2008.

A *deflation* occurs when prices decline (which means that the rate of inflation is negative). At the other extreme is a *hyperinflation*, a rise in the price level of a thousand or a million percent a year. In such situations, as in Weimar Germany in the 1920s, Brazil in the 1980s, Russia in the 1990s, or Zimbabwe

in recent years, prices are virtually meaningless and the price system breaks down.

Price stability is important because a smoothly functioning market system requires that prices accurately convey information about relative scarcities. History has shown that high inflation imposes many costs—some visible and some hidden—on an economy. With high inflation, taxes become highly variable, the real values of people's pensions are eroded, and people spend real resources to avoid depreciating rubles or pesos. But declining prices (deflation) are also costly. Hence, most nations seek the golden mean of slowly rising prices as the best way of encouraging the price system to function efficiently.

To summarize:

The goals of macroeconomic policy are:

1. A high and growing level of national output
2. High employment with low unemployment
3. A stable or gently rising price level

The Tools of Macroeconomic Policy

Put yourself in the shoes of the chief economist advising the government. Unemployment is rising and GDP is falling. Or perhaps the burst of a speculative bubble in housing prices has led to massive defaults, banking losses, and a credit crunch. Or your country has a balance-of-payments crisis, with a large trade deficit and a foreign-exchange rate that is in free fall. What policies will help reduce inflation or unemployment, speed economic growth, or correct a trade imbalance?

Governments have certain instruments that they can use to affect macroeconomic activity. A *policy instrument* is an economic variable under the control of government that can affect one or more of the macroeconomic goals. By changing monetary, fiscal, and other policies, governments can avoid the worst excesses of the business cycle or increase the growth rate of potential output. The major instruments of macroeconomic policy are listed in the bottom half of Table 19-1.

Fiscal Policy. **Fiscal policy** denotes the use of taxes and government expenditures. *Government expenditures* come in two distinct forms. First there are government purchases. These comprise spending on goods and services—purchases of tanks, construction of roads, salaries for judges, and so forth. In addition, there are government transfer payments, which increase the incomes of targeted groups such as the elderly or the unemployed. Government spending determines the relative size of the public and private sectors, that is, how much of our GDP is consumed collectively rather than privately. From a macroeconomic perspective, government expenditures also affect the overall level of spending in the economy and thereby influence the level of GDP.

The other part of fiscal policy, *taxation*, affects the overall economy in two ways. To begin with, taxes affect people's incomes. By leaving households with more or less disposable or spendable income, taxes affect the amount people spend on goods and

services as well as the amount of private saving. Private consumption and saving have important effects on investment and output in the short and long run.

In addition, taxes affect the prices of goods and factors of production and thereby affect incentives and behavior. The United States has often employed special tax provisions (such as an investment tax credit or accelerated depreciation) as ways of increasing investment and boosting economic growth. Many provisions of the tax code have an important impact on economic activity through their effect on the incentives to work and to save.

Monetary Policy. The second major instrument of macroeconomic policy is **monetary policy**, which the government conducts through managing the nation's money, credit, and banking system. You may have read how our central bank, the Federal Reserve System, affects the economy by determining short-term interest rates. How does the Federal Reserve or any other central bank actually accomplish this? It does so primarily by setting short-run interest-rate targets and through buying and selling government securities to attain those targets. Through its operations, the Federal Reserve influences many financial and economic variables, such as interest rates, stock prices, housing prices, and foreign exchange rates. These financial variables affect spending on investment, particularly in housing, business investment, consumer durables, and exports and imports.

Historically, the Fed has raised interest rates when inflation threatened to rise too high. This led to reduced investment and consumption, causing a decline in GDP and lower inflation. In the most recent slowdown, which started in 2007, the Fed acted quickly to lower interest rates, provide credit, and extend its lending facilities outside traditional banking institutions.

The central bank is a key macroeconomic institution for every country. Japan, Britain, Russia, and the countries of the European Union all have powerful central banks. In an "open economy"—that is, one whose borders are open to goods, services, and financial flows—the exchange-rate system is also a central part of monetary policy.

Monetary policy is the tool that countries most often rely on to stabilize the business cycle, although it becomes less potent in deep recessions. The exact way that central banks can affect economic activity

will be thoroughly analyzed in the chapters on monetary policy.

Summary:

A nation has two major kinds of policies that can be used to pursue its macroeconomic goals—fiscal policy and monetary policy.

1. Fiscal policy consists of government expenditure and taxation. Government expenditure influences the relative size of collective spending and private consumption. Taxation subtracts from incomes, reduces private spending, and affects private saving. In addition, it affects investment and potential output. Fiscal policy is primarily used to affect long-term economic growth through its impact on national saving and investment; it is also used to stimulate spending in deep or sharp recessions.
2. Monetary policy, conducted by the central bank, determines short-run interest rates. It thereby affects credit conditions, including asset prices such as stock and bond prices and exchange rates. Changes in interest rates, along with other financial conditions, affect spending in sectors such as business investment, housing, and foreign trade. Monetary policy has an important effect on both actual GDP and potential GDP.

INTERNATIONAL LINKAGES

No nation is an island unto itself. Nations increasingly participate in the world economy and are linked together through trade and finance—this is the phenomenon called *globalization*. As the costs of transportation and communication have declined, international linkages have become tighter than they were a generation ago. International trade has replaced empire-building and military conquest as the surest road to national wealth and influence.

The trade linkages of imports and exports of goods and services are seen when the United States imports cars from Japan or exports computers to Mexico. Financial linkages come in activities such as foreigners' buying U.S. bonds for their sovereign debt funds or Americans' diversifying their pension funds with emerging-market stocks.

Nations keep a close watch on their international transactions. One particularly important measure is the *balance on current account*. This represents the numerical difference between the value of exports

and the value of imports, along with some other adjustments. (The current account is closely related to *net exports*, which is the difference between the value of exports and the value of imports of goods and services.) When exports exceed imports, the difference is a surplus, while a negative balance is a deficit. In 2007, exports totaled \$2463 billion, while total imports and net transfers were \$3194 billion; the difference was the U.S. current-account deficit of \$731 billion.

For most of the twentieth century, the United States had a surplus in its foreign trade, exporting more than it imported. But trading patterns changed dramatically in the last quarter-century. As saving in the United States declined and foreign saving increased, a substantial part of foreign saving flowed to the United States. The counterpart of foreigners saving in the United States was that the current account turned sharply to deficit. As foreign investment in the nation increased, the United States by 2008 owed on balance around \$2½ trillion to foreigners. Some economists worry that the large foreign debt poses major risks for the United States—risks that we will analyze in later chapters.

As economies become more closely linked, international economic policy becomes more important, particularly in small open economies. But remember that international trade and finance are not ends in themselves. Rather, international exchange serves the ultimate goal of improving living standards.

The major areas of concern are trade policies and international financial management. *Trade policies* consist of tariffs, quotas, and other regulations that restrict or encourage imports and exports. Most trade policies have little effect on short-run macroeconomic performance, but from time to time, as was the case in the 1930s, restrictions on international trade are so severe that they cause major economic dislocations, inflations, or recessions.

A second set of policies is *international financial management*. A country's international trade is influenced by its foreign exchange rate, which represents the price of its own currency in terms of the currencies of other nations. Foreign exchange systems are an integral part of monetary policy. In small open economies, managing the exchange rate is the single most important macroeconomic policy.

The international economy is an intricate web of trading and financial connections among countries.

When the international economic system runs smoothly, it contributes to rapid economic growth; when trading systems break down, production and incomes suffer throughout the world. Countries therefore consider the impacts of trade policies and international financial policies on their domestic objectives of high output, high employment, and price stability.

B. AGGREGATE SUPPLY AND DEMAND

The economic history of nations can be seen in their macroeconomic performance. Economists have developed aggregate supply-and-demand analysis to help explain the major trends in output and prices. We begin by explaining this important tool of macroeconomics and then use it to understand some important historical events.

INSIDE THE MACROECONOMY: AGGREGATE SUPPLY AND DEMAND

Definitions of Aggregate Supply and Demand

How do different forces interact to determine overall economic activity? Figure 19-5 shows the relationships among the different variables inside the macroeconomy. It separates variables into two categories: those affecting aggregate supply and those affecting aggregate demand. While the division is simplified, dividing variables into these two categories helps us understand what determines the levels of output, prices, and unemployment.

The lower part of Figure 19-5 shows the forces affecting aggregate supply. **Aggregate supply** refers to the total quantity of goods and services that the nation's businesses willingly produce and sell in a given period. Aggregate supply (often written AS) depends upon the price level, the productive capacity of the economy, and the level of costs.

In general, businesses would like to sell everything they can produce at high prices. Under some circumstances, prices and spending levels may be depressed, so businesses might find they have excess capacity. Under other conditions, such as during a

wartime boom, factories may be operating at capacity as businesses scramble to produce enough to meet all their orders.

We see, then, that aggregate supply depends on the price level that businesses can charge as well as on the economy's capacity or potential output. Potential output in turn is determined by the availability of productive inputs (labor and capital being the most important) and the managerial and technical efficiency with which those inputs are combined.

National output and the overall price level are determined by the twin blades of the scissors of aggregate supply and demand. The second blade is **aggregate demand**, which refers to the total amount that different sectors in the economy willingly spend in a given period. Aggregate demand (often written *AD*) equals total spending on goods and services. It depends on the level of prices, as well as on monetary policy, fiscal policy, and other factors.

The components of aggregate demand include *consumption* (the cars, food, and other consumption goods bought by consumers); *investment* (construction of houses and factories as well as business equipment); *government purchases* (such as spending on teachers and missiles); and *net exports* (the difference between exports and imports). Aggregate demand is affected by the prices at which the goods are offered, by exogenous forces like wars and weather, and by government policies.

Using both blades of the scissors of aggregate supply and demand, we achieve the resulting equilibrium, as is shown in the right-hand circle of Figure 19-5. National output and the price level settle at that level where demanders willingly buy what businesses willingly sell. The resulting output and price level determine employment, unemployment, and international trade.

Aggregate Supply and Demand Curves

Aggregate supply and demand curves are often used to help analyze macroeconomic conditions. Recall that in Chapter 3 we used market supply and demand curves to analyze the prices and quantities of individual products. An analogous graphical apparatus can help us understand how monetary policy or technological change acts through aggregate supply and demand to determine national output and the price level.

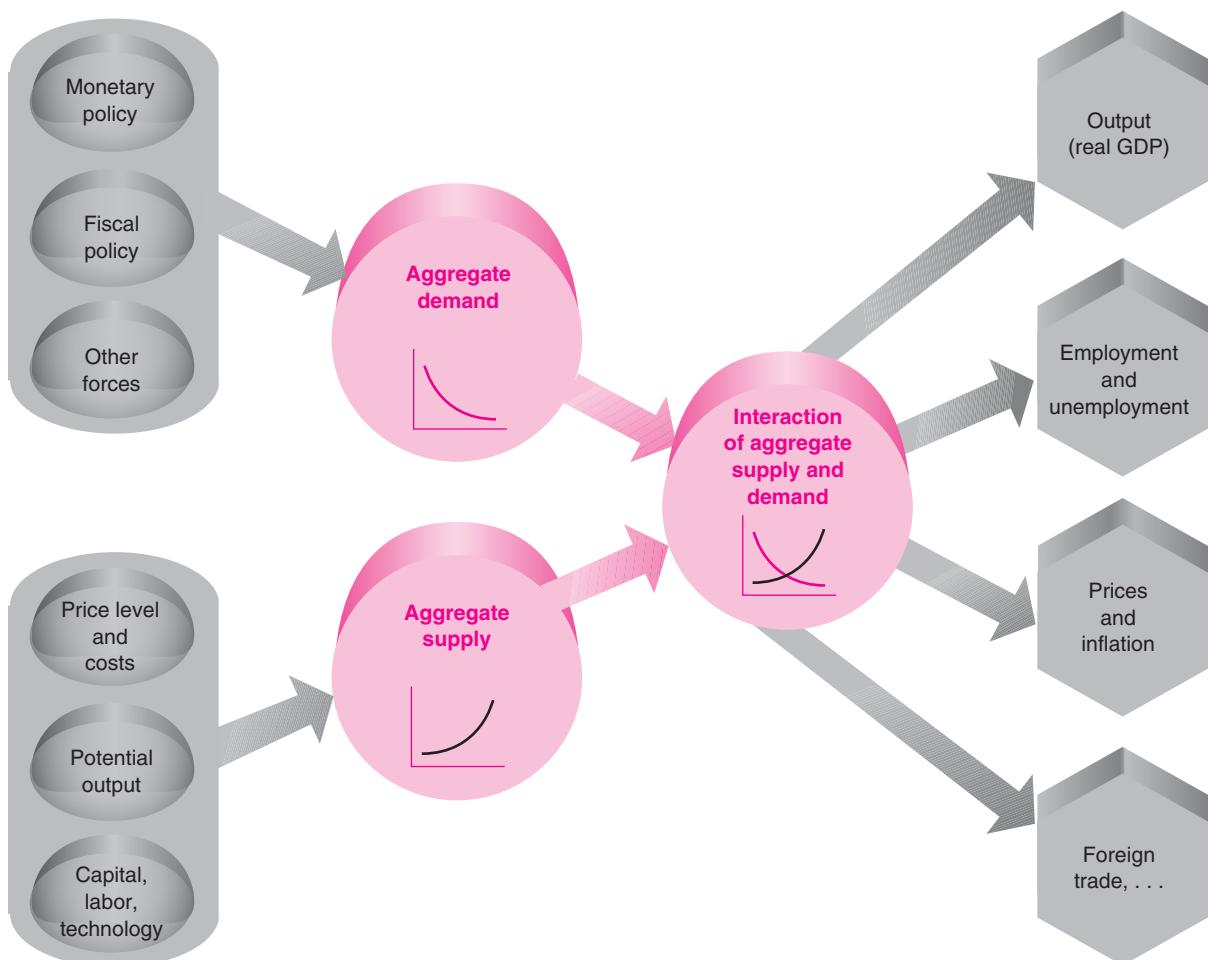


FIGURE 19-5. Aggregate Supply and Demand Determine the Major Macroeconomic Variables

This key diagram shows the major factors affecting overall economic activity. On the left are the major variables determining aggregate supply and demand; these include policy variables, like monetary and fiscal policies, along with stocks of capital and labor. In the center, aggregate supply and demand interact. The chief outcomes are shown on the right in hexagons: output, employment, the price level, and international trade.

Figure 19-6 shows the aggregate supply and demand schedules for the output of an entire economy. On the horizontal axis is the total output (real GDP) of the economy. On the vertical axis is the overall price level (as measured by the “price of GDP”). We use the symbol Q for real output and P for the price level.

The downward-sloping curve is the **aggregate demand schedule**, or AD curve. It represents what

everyone in the economy—consumers, businesses, foreigners, and governments—would buy at different aggregate price levels (with other factors affecting aggregate demand held constant). From the curve, we see that at an overall price level of 150, total spending would be \$3000 billion (per year). If the price level rises to 200, total spending would fall to \$2300 billion.

The upward-sloping curve is the **aggregate supply schedule**, or AS curve. This curve represents the

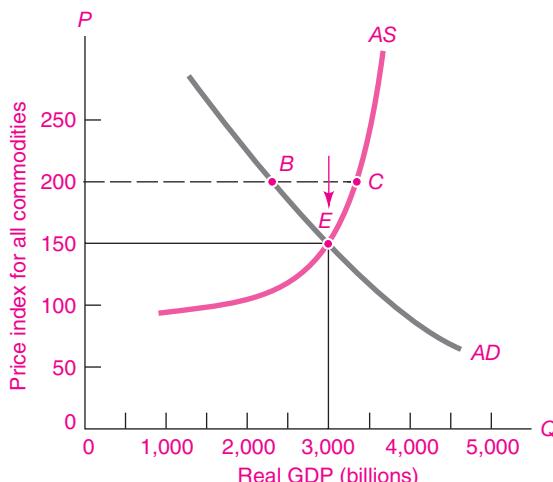


FIGURE 19-6. Aggregate Price and Output Are Determined by the Interaction of Aggregate Supply and Demand

The *AD* curve represents the quantity of total spending at different price levels, with other factors held constant. The *AS* curve shows what firms will produce and sell at different price levels, other things equal.

National output and the overall price level are determined at the intersection of the aggregate demand and supply curves, at point *E*. This equilibrium occurs at an overall price level where firms willingly produce and sell what consumers and other demanders willingly buy.

quantity of goods and services that businesses are willing to produce and sell at each price level (with other determinants of aggregate supply held constant). According to the curve, businesses will want to sell \$3000 billion at a price level of 150; they will want to sell a higher quantity, \$3300 billion, if prices rise to 200. As the level of total output demanded rises, businesses will want to sell more goods and services at a higher price level.



Warning on AS and AD Curves

Before proceeding, here is one important word of caution: Do not confuse the macroeconomic *AD* and *AS* curves with the microeconomic *DD* and *SS* curves. The microeconomic supply and demand curves show the quantities and prices of individual commodities, with such things as national income and other goods' prices held as given. By contrast, the aggregate supply and demand curves show the

determination of total output and the overall price level, with such things as the money supply, fiscal policy, and the capital stock held constant.

Aggregate supply and demand explain how *total taxes* affect aggregate demand, national output, and the overall price level. Microeconomic supply and demand might consider the way increases in *gasoline taxes* affect purchases of gasoline, holding income constant. The two sets of curves have a superficial resemblance, but they explain very different phenomena.

Note as well that we have drawn the *AS* curve as upward-sloping and the *AD* curve as downward-sloping. We explain the reasons for these slopes in later chapters.

Macroeconomic Equilibrium. We now see how aggregate output and the price level adjust or equilibrate to bring aggregate supply and aggregate demand into balance. That is, we use the *AS* and *AD* concepts to see how *equilibrium values of price and quantity* are determined or to find the *P* and *Q* that satisfy the buyers and sellers all taken together. For the *AS* and *AD* curves shown in Figure 19-6, the overall economy is in equilibrium at point *E*. Only at that point, where the level of output is *Q* = 3000 and *P* = 150, are spenders and sellers satisfied. Only at point *E* are demanders willing to buy exactly the amount that businesses are willing to produce and sell.

How does the economy reach its equilibrium? Indeed, what do we mean by equilibrium? A **macroeconomic equilibrium** is a combination of overall price and quantity at which all buyers and sellers are satisfied with their overall purchases, sales, and prices.

Figure 19-6 illustrates the concept. If the price level were higher than equilibrium, say, at *P* = 200, businesses would want to sell more than purchasers would want to buy; businesses would desire to sell quantity *C*, while buyers would want to purchase only amount *B*. Goods would pile up on the shelves as firms produced more than consumers bought. Because of the excess aggregate supply of goods, firms would cut production and shave their prices. The overall price level would begin to decline or rise less rapidly. As the price level declined from its original too high level, the gap between desired total spending and desired total sales would narrow. Eventually, prices would decline to the point where overall demand and production were in balance. At the macroeconomic equilibrium, there would be

neither excess supply nor excess demand—and no pressure to change the overall price level.

MACROECONOMIC HISTORY: 1900–2008

We can use the aggregate supply-and-demand apparatus to analyze recent American macroeconomic history. We focus on the economic expansion during the Vietnam War, the deep recession caused by the monetary contraction of the early 1980s, and the phenomenal record of economic growth during the twentieth century. This chapter's appendix also provides data on major macroeconomic variables.

Wartime Boom. The American economy entered the 1960s having experienced multiple recessions (see Figure 19-3). President John Kennedy brought Keynesian economics to Washington. His economic advisers recommended expansionary policies, and Congress enacted measures to stimulate the economy, particularly cuts in personal and corporate taxes in 1963 and 1964. GDP grew rapidly during this period, unemployment declined, and inflation was contained. By 1965, the economy was at its potential output.

Unfortunately, the government underestimated the magnitude of the buildup for the Vietnam War; defense spending grew by 55 percent from 1965 to 1968. Even when it became clear that a major inflationary boom was under way, President Johnson postponed painful fiscal steps to slow the economy. Tax increases and civilian expenditure cuts came only in 1968, which was too late to prevent inflationary pressures from overheating the economy. The Federal Reserve accommodated the expansion with rapid money growth and low interest rates. As a result, the economy grew very rapidly over the period 1966–1970. Under the pressure of low unemployment and high factory utilization, inflation began to rise, inaugurating the “Great Inflation” that lasted from 1966 through 1981.

Figure 19-7 illustrates the events of this period. The tax cuts and defense expenditures shifted the aggregate demand curve to the right from AD to AD' , with the equilibrium shifting from E to E' . Output and employment rose sharply, and inflation rose as output exceeded capacity limits. Economists learned

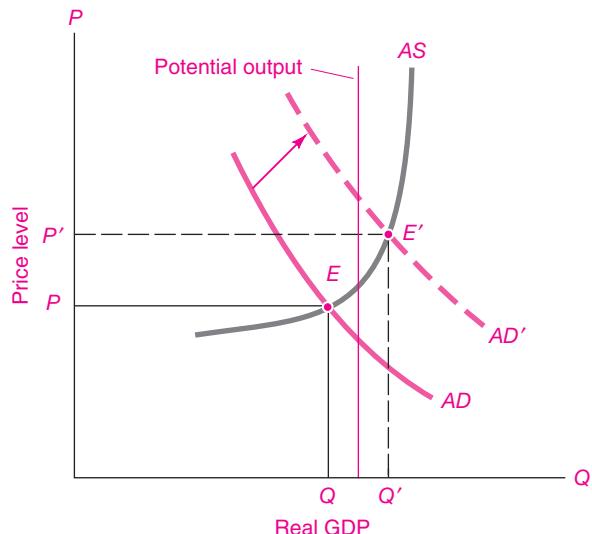


FIGURE 19-7. Wartime Boom Is Propelled by Increasing Aggregate Demand

During wartime, increased military spending increases aggregate spending, moving aggregate demand from AD to AD' , with equilibrium output increasing from E to E' . When output rises far above potential output, the price level moves up sharply from P to P' , and wartime inflation ensues.

that it was easier to stimulate the economy than to persuade policymakers to raise taxes to slow the economy when inflation threatened. This lesson led many to question the wisdom of using fiscal policies to stabilize the economy.

Tight Money, 1979–1982. The 1970s were a time of troubles, with rising oil prices, grain shortages, a sharp increase in import prices, union militancy, and accelerating wages. Price inflation became embedded in the U.S. and many other economies. As Figure 19-4 on page 374 shows, inflation rose to double-digit levels in the 1978–1980 period.

Double-digit inflation was unacceptable. In response, the Federal Reserve, under the leadership of economist Paul Volcker, prescribed the strong medicine of tight money to slow the inflation. Interest rates rose sharply in 1979 and 1980, the stock market fell, and credit was hard to find. The Fed's tight-money policy slowed spending by

consumers and businesses. Particularly hard-hit were interest-sensitive components of aggregate demand. After 1979, housing construction, automobile purchases, business investment, and net exports declined sharply.

We can picture how tight money reduced aggregate demand in Figure 19-7 simply by reversing the arrow. That is, tight monetary policy reduced spending and produced a leftward and downward shift of the aggregate demand curve—exactly the opposite of the effect of the tax cuts and defense buildup during the 1960s.

The effects of the tight money were twofold. First, output moved below its potential and unemployment rose sharply (see Figure 19-3 on page 373). Second, tight money and high unemployment produced a dramatic decline in inflation, from an average of 12 percent per year in the 1978–1980 period to an average of around 4 percent per year in the subsequent period (see Figure 19-4). Tight monetary policies succeeded in bringing an end to the Great Inflation, but the nation paid through higher unemployment and lower output during the period of tight money.

The Growth Century. The final act in our macroeconomic drama concerns the growth of output and prices over the entire period since 1900. Output has grown by a factor of 34 since the beginning of the twentieth century. How can we explain this phenomenal increase?

A careful look at American economic growth reveals that the growth rate during the twentieth century averaged $3\frac{1}{3}$ percent per year. Part of this growth was due to growth in the scale of production as inputs of capital, labor, and even land grew sharply over this period. Just as important were improvements in efficiency due to new products (such as automobiles) and new processes (such as electronic computing). Other, less visible factors also contributed to economic growth, such as improved management techniques and improved services (including such innovations as the assembly line and overnight delivery).

Many economists believe that the measured growth understates true growth because our official statistics tend to miss the contribution to living standards from new products and improvements in product quality. For example, with the introduction

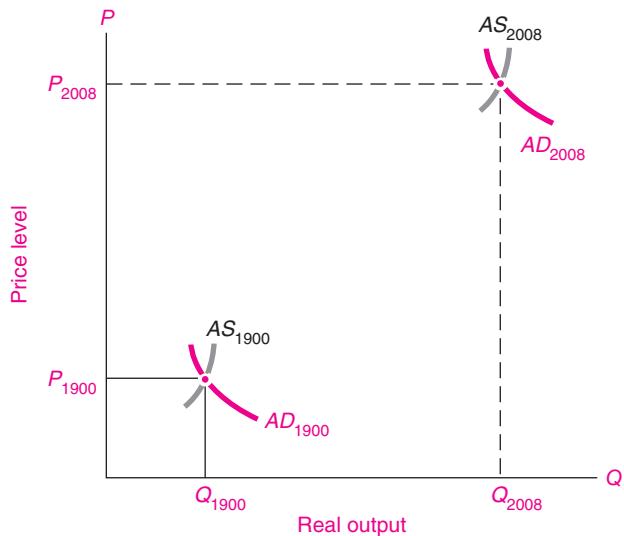


FIGURE 19-8. Growth in Potential Output Determines Long-Run Economic Performance

Over the twentieth century, increases in labor, capital, and efficiency led to a vast increase in the economy's productive potential, shifting aggregate supply far to the right. In the long run, aggregate supply is the primary determinant of output growth.

of the indoor toilet, millions of people no longer had to struggle through the winter snows to relieve themselves in outhouses, yet this increased comfort never showed up in measured gross domestic product.

How can we picture the tremendous rise in output in our AS-AD apparatus? Figure 19-8 shows the way. The increase in inputs and improvements in efficiency led to a massive rightward shift of the AS curve from AS_{1900} to AS_{2008} . Production costs also increased sharply. For example, average earnings rose from \$0.15 per hour in 1900 to over \$30 per hour in 2008. These cost increases shifted the AS curve upward. The overall effect, then, was the increase in both output and prices shown in Figure 19-8.

The Role of Macroeconomic Policy

Macroeconomic policy played a central role in the improved business-cycle conditions of the last half-century. The discovery and application of macroeconomics, along with a good appreciation of the role and limitations of monetary and fiscal policy, reduced business-cycle volatility and led to the

Great Moderation. The application of fiscal policy, and especially monetary policy, helped lower unemployment and ensured largely stable prices over the last two decades. When the United States faced a major shock to its financial system in 2007–2009, central bankers remembered *and understood* the lessons of the Great Depression. They knew that financial fears are contagious, that bank collapses can lead to bank runs, and that instability breeds more instability. Knowledge of macroeconomic history and theory,

and the intervention of the central bank as a lender of last resort, can cushion a banking shock and prevent bank crises from turning into deep depressions.

There is no miracle cure for macroeconomic shocks, however. When a steep decline in output and employment hit the United States in 2007–2009, monetary and fiscal policies were launched to soften the blow, but they could not completely offset it. Up to now, the knowledge is available to prevent depressions, but not to banish recessions.



SUMMARY

A. Key Concepts of Macroeconomics

1. Macroeconomics is the study of the behavior of the entire economy: It analyzes long-run growth as well as the cyclical movements in total output, unemployment and inflation, and international trade and finance. This contrasts with microeconomics, which studies the behavior of individual markets, prices, and outputs.
2. The United States proclaimed its macroeconomic goals in the Employment Act of 1946, which declared that federal policy was “to promote maximum employment, production, and purchasing power.” Since then, the nation’s priorities among these three goals have shifted. But all market economies still face three central macroeconomic questions: (a) Why do output and employment sometimes fall, and how can unemployment be reduced? (b) What are the sources of price inflation, and how can it be kept under control? (c) How can a nation increase its rate of economic growth?
3. In addition to these perplexing questions is the hard fact that there are inevitable conflicts or tradeoffs among these goals: Rapid growth in future living standards may mean reducing consumption today, and curbing inflation may involve a temporary period of high unemployment.
4. Economists evaluate the success of an economy’s overall performance by how well it attains these objectives: (a) high levels and rapid growth of output (measured by real gross domestic product) and consumption; (b) a low unemployment rate and high employment, with an ample supply of good jobs; (c) low and stable inflation.
5. Before the science of macroeconomics was developed, countries tended to drift around in the shifting macroeconomic currents without a rudder. Today, there

are numerous instruments with which governments can steer the economy: (a) Fiscal policy (government spending and taxation) helps determine the allocation of resources between private and collective goods, affects people’s incomes and consumption, and provides incentives for investment and other economic decisions. (b) Monetary policy—particularly the setting of short-term interest rates by the central bank—affects all interest rates, asset prices, credit conditions, and exchange rates. The most heavily affected sectors are housing, business investment, consumer durables, and net exports.

6. The nation is but a small part of an increasingly integrated global economy in which countries are linked together through trade of goods and services and through financial flows. A smoothly running international economic system contributes to rapid economic growth, but the international economy can throw sand in the engine of growth when trade flows are interrupted or the international financial mechanism breaks down. Dealing with international trade and finance is high on the agenda of all countries.

B. Aggregate Supply and Demand

7. The central concepts for understanding the determination of national output and the price level are aggregate supply (AS) and aggregate demand (AD). Aggregate demand consists of the total spending in an economy by households, businesses, governments, and foreigners. It represents the total output that would be willingly bought at each price level, given the monetary and fiscal policies and other factors affecting demand. Aggregate supply describes how much output businesses would willingly produce and sell given prices, costs, and market conditions.

8. AS and AD curves have the same shapes as the familiar supply and demand curves analyzed in microeconomics. But beware of potential confusions of microeconomic and aggregate supply and demand.
9. The overall macroeconomic equilibrium, determining both aggregate price and output, comes where the AS and AD curves intersect. At the equilibrium price level, purchasers willingly buy what businesses willingly sell. Equilibrium output can depart from full employment or potential output.
10. Recent American history shows an irregular cycle of aggregate demand and supply shocks and policy reactions. In the mid-1960s, war-bloated deficits plus easy money led to a rapid increase in aggregate demand. The result was a sharp upturn in prices and inflation. At the end of the 1970s, economic policy-makers reacted to the rising inflation by tightening monetary policy and raising interest rates. The result lowered spending on interest-sensitive demands such as housing, investment, and net exports. Since the mid-1980s, the U.S. economy has experienced a period of low inflation and infrequent and, until recently, mild recessions.
11. Over the long run, the growth of potential output increased aggregate supply enormously and led to steady growth in output and living standards.

CONCEPTS FOR REVIEW

Major Macroeconomic Concepts

macroeconomics vs. microeconomics
gross domestic product (GDP), actual and potential
employment, unemployment,
unemployment rate

inflation, deflation
consumer price index (CPI)
net exports
fiscal policy (government expenditures, taxation)
monetary policy

Aggregate Supply and Demand

aggregate supply, aggregate demand
AS curve, AD curve
equilibrium of AS and AD
sources of long-run economic growth

FURTHER READING AND INTERNET WEBSITES

Further Reading

The great classic of macroeconomics is John Maynard Keynes, *The General Theory of Employment, Interest, and Money* (Harcourt, New York, first published in 1935). Keynes was one of the most graceful writers among economists. An online edition of *The General Theory* is available at www.marxists.org/reference/subject/economics/keynes/general-theory/.

There are many good intermediate textbooks on macroeconomics. You may consult these when you want to dig more deeply into specific topics.

Websites

Macroeconomic issues are a central theme of analysis in *Economic Report of the President*. Various years are available

online at www.access.gpo.gov/eop. Another good source on macroeconomic issues is the Congressional Budget Office, which issues periodic reports on the economy and the state of the budget at www.cbo.gov.

Research organizations often contain excellent online discussions of current macroeconomic issues. See especially the websites of the Brookings Institution, www.brookings.org, and the American Enterprise Institute, www.aei.org.

Some excellent blogs containing macroeconomics are the following: A blog of leading European and some American economists contains much interesting economic commentary at www.voxeu.org; the *International Herald Tribune* has a fine group of expert writers at blogs.iht.com/tribtalk/business/globalization.

QUESTIONS FOR DISCUSSION

1. What are the major objectives of macroeconomics? Write a brief definition of each of these objectives. Explain carefully why each objective is important.
2. Using the data from the appendix to this chapter, calculate the following:
 - a. The inflation rate in 1981 and 2007
 - b. The growth rate of real GDP in 1982 and 1984
 - c. The average inflation rate from 1970 to 1980 and from 2000 to 2007
 - d. The average growth rate of real GDP from 1929 to 2008

[Hint: The formulas in the text give the technique for calculating 1-year growth rates. Growth rates for multiple years use the following formula:

$$g_t^{(n)} = 100 \times \left[\left(\frac{X_t}{X_{t-n}} \right)^{1/n} - 1 \right]$$

where $g_t^{(n)}$ is the average annual growth rate of the variable X for the n years between year $(t - n)$ and year t . For example, assume that the CPI in $(t - 2)$ is 100.0 while the CPI in year t is 106.09. Then the average rate of inflation is $100 \times \left[\left(\frac{106.09}{100.0} \right)^{1/2} - 1 \right] = 3$ percent per year.]

3. What would be the effect of each of the following on aggregate demand or on aggregate supply, as indicated (always holding other things constant)?
 - a. A large cut in personal and business taxes (on AD)
 - b. An arms-reduction agreement reducing defense spending (on AD)
 - c. An increase in potential output (on AS)
 - d. A monetary loosening that lowers interest rates (on AD)
4. For each of the events listed in question 3, use the $AS-AD$ apparatus to show the effect on output and on the overall price level.
5. Put yourself in the shoes of an economic policymaker. The economy is in equilibrium with $P = 100$ and $Q = 3000 =$ potential GDP. You refuse to “accommodate” inflation; that is, you want to keep prices absolutely stable at $P = 100$, no matter what happens to output. You can use monetary and fiscal policies to affect aggregate demand, but you cannot affect aggregate supply in the short run. How would you respond to:
 - a. A surprise increase in investment spending
 - b. A sharp food-price increase following catastrophic flooding of the Mississippi River
 - c. A productivity decline that reduces potential output

- d. A sharp decrease in net exports that followed a deep depression in East Asia
6. In 1981–1983, the Reagan administration implemented a fiscal policy that reduced taxes and increased government spending.
 - a. Explain why this policy would tend to increase aggregate demand. Show the impact on output and prices assuming only an AD shift.
 - b. The supply-side school holds that tax cuts would affect aggregate supply mainly by increasing potential output. Assuming that the Reagan fiscal measures affected AS as well as AD , show the impact on output and the price level. Explain why the impact of the Reagan fiscal policies on output is unambiguous while the impact on prices is unclear.
7. The Clinton economic package as passed by Congress in 1993 had the effect of tightening fiscal policy by raising taxes and lowering spending. Show the effect of this policy (a) assuming that there is no counteracting monetary policy and (b) assuming that monetary policy completely neutralized the impact on GDP and that the lower deficit leads to higher investment and higher growth of potential output.
8. The United States experienced a major economic downturn in the early 1980s. Consider the data on real GDP and the price level in Table 19-2.
 - a. For the years 1981 to 1985, calculate the rate of growth of real GDP and the rate of inflation. Can you determine in which year there was a steep business downturn or recession?
 - b. In an $AS-AD$ diagram like Figure 19-6 (page 379), draw a set of AS and AD curves that trace out the price and output equilibria shown in the table. How would you explain the recession that you have identified?

Year	Real GDP (\$, billion, 2000 prices)	Price level* (2000 = 100)
1980	5,161.7	54.1
1981	5,291.7	59.1
1982	5,189.3	62.7
1983	5,423.8	65.2
1984	5,813.6	67.7
1985	6,053.7	69.7

*Note that the price index shown is the price index for GDP, which measures the price trend for all components of GDP.

TABLE 19-2.



Appendix 19

MACROECONOMIC DATA FOR THE UNITED STATES

Year	Nominal GDP (\$, billion)	Real GDP, 2000 prices (\$, billion)	Unemployment rate (%)	CPI 1982–1984 = 100	Inflation rate (CPI) (% per year)	Federal budget surplus (+) or deficit (-) (\$, billion)	Net exports (\$, billion)
1929	103.6	865.2	3.2	17.1	0.0	1.0	0.4
1933	56.4	635.5	24.9	13.0	-5.2	-0.9	0.1
1939	92.2	950.7	17.2	13.9	-1.4	-2.1	0.8
1945	223.1	1,786.3	1.9	18.0	2.2	-29.0	-0.8
1948	269.2	1,643.2	3.8	24.0	7.4	3.6	5.5
1950	293.8	1,777.2	5.2	24.1	1.1	5.5	0.7
1960	526.4	2,501.8	5.5	29.6	1.5	7.2	4.2
1970	1,038.5	3,771.9	5.0	38.8	5.7	-15.2	4.0
1971	1,127.1	3,898.7	6.0	40.5	4.1	-28.4	0.6
1972	1,238.3	4,104.9	5.6	41.8	3.2	-24.4	-3.4
1973	1,382.7	4,341.4	4.9	44.4	6.1	-11.3	4.1
1974	1,500.0	4,319.5	5.6	49.3	10.4	-13.8	-0.8
1975	1,638.3	4,311.2	8.5	53.8	8.7	-69.0	16.0
1976	1,825.3	4,540.9	7.7	56.9	5.6	-51.7	-1.6
1977	2,030.9	4,750.6	7.1	60.6	6.3	-44.1	-23.1
1978	2,294.7	5,015.0	6.1	65.2	7.4	-26.5	-25.4
1979	2,563.3	5,173.5	5.9	72.6	10.7	-11.3	-22.5
1980	2,789.5	5,161.7	7.2	82.4	12.7	-53.6	-13.1
1981	3,128.4	5,291.7	7.6	90.9	9.9	-53.3	-12.5
1982	3,255.0	5,189.3	9.7	96.5	6.0	-131.9	-20.0
1983	3,536.7	5,423.8	9.6	99.6	3.1	-173.0	-51.7
1984	3,933.2	5,813.6	7.5	103.9	4.3	-168.1	-102.7
1985	4,220.3	6,053.8	7.2	107.6	3.5	-175.0	-115.2
1986	4,462.8	6,263.6	7.0	109.7	1.9	-190.8	-132.7
1987	4,739.5	6,475.1	6.2	113.6	3.5	-145.0	-145.2
1988	5,103.8	6,742.7	5.5	118.3	4.0	-134.5	-110.4
1989	5,484.4	6,981.4	5.3	123.9	4.7	-130.1	-88.2
1990	5,803.1	7,112.5	5.6	130.7	5.3	-172.0	-78.0
1991	5,995.9	7,100.5	6.9	136.2	4.1	-213.7	-27.5
1992	6,337.7	7,336.6	7.5	140.3	3.0	-297.4	-33.2
1993	6,657.4	7,532.7	6.9	144.5	2.9	-273.5	-65.0
1994	7,072.2	7,835.5	6.1	148.2	2.6	-212.3	-93.6
1995	7,397.7	8,031.7	5.6	152.4	2.8	-197.0	-91.4
1996	7,816.9	8,328.9	5.4	156.9	2.9	-141.8	-96.2
1997	8,304.3	8,703.5	4.9	160.5	2.3	-55.8	-101.6
1998	8,747.0	9,066.9	4.5	163.0	1.5	38.8	-159.9
1999	9,268.4	9,470.4	4.2	166.6	2.2	103.6	-260.5
2000	9,817.0	9,817.0	4.0	172.2	3.3	189.5	-379.5
2001	10,128.0	9,890.7	4.7	177.0	2.8	46.7	-367.0
2002	10,469.6	10,048.9	5.8	179.9	1.6	-247.9	-424.4
2003	10,960.8	10,301.1	6.0	184.0	2.3	-372.1	-499.4
2004	11,685.9	10,675.7	5.5	188.9	2.6	-370.6	-615.4
2005	12,433.9	11,003.5	5.1	195.3	3.3	-318.3	-714.6
2006	13,194.7	11,319.4	4.6	201.6	3.2	-220.0	-762.0
2007	13,807.6	11,523.9	4.6	207.3	2.8	-399.4	-707.8
2008	14,304.4	11,666.0	5.8	215.2	4.1	-456.5	-727.9

TABLE 19A-1.

Table 19A-1 contains some of the major macroeconomic data discussed in this chapter. Major data can be obtained through government websites at www.fedstats.gov, www.bea.gov, or www.bls.gov.

CHAPTER

20

Measuring Economic Activity



When you can measure what you are speaking about, and express it in numbers, you know something about it; when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science.

Lord Kelvin

The single most important concept in macroeconomics is the gross domestic product (GDP), which measures the total value of goods and services produced in a country during a year. GDP is part of the *national income and product accounts* (or *national accounts*), which are a body of statistics that enables policymakers to determine whether the economy is contracting or expanding and whether a severe recession or inflation threatens. When economists want to determine the level of economic development of a country, they look at its GDP per capita.

While the GDP and the rest of the national accounts may seem to be arcane concepts, they are truly among the great inventions of modern times. Much as a satellite in space can survey the weather across an entire continent, so can the GDP give an overall picture of the state of the economy. In this chapter, we explain how economists measure GDP and other major macroeconomic indicators.

GROSS DOMESTIC PRODUCT: THE YARDSTICK OF AN ECONOMY'S PERFORMANCE

What is the *gross domestic product*? GDP is the name we give to the total market value of the final goods and services produced within a nation during a given year. It is the figure you get when you apply the measuring rod of money to the diverse goods and services—from apples to zithers—that a country produces with its land, labor, and capital resources. GDP equals the total production of consumption and investment goods, government purchases, and net exports to other lands.

The gross domestic product (GDP) is the most comprehensive measure of a nation's total output of goods and services. It is the sum of the dollar values of consumption (C), gross investment (I), government purchases of goods and services (G), and

net exports (X) produced within a nation during a given year.

In symbols:

$$\text{GDP} = C + I + G + X$$

GDP is used for many purposes, but the most important one is to measure the overall performance of an economy. If you were to ask an economic historian what happened during the Great Depression, the best short answer would be:

Between 1929 and 1933, GDP fell from \$104 billion to \$56 billion. This sharp decline in the dollar value of goods and services produced by the American economy caused high unemployment, hardship, a steep stock market decline, bankruptcies, bank failures, riots, and political turmoil.

Similarly, if you were to ask a macroeconomist about the second half of the twentieth century, she might reply:

The second half of the twentieth century was a unique economic period. During those years, the affluent regions of the North—consisting of Japan, the United States, and Western Europe—experienced the most rapid growth of output per capita in recorded history. From the end of World War II until 2000, for example, real GDP per capita in the United States expanded by almost 250 percent.

We now discuss the elements of the national income and product accounts. We start by showing different ways of measuring GDP and distinguishing real from nominal GDP. We then analyze the major components of GDP. We conclude with a discussion of the measurement of the general price level and the rate of inflation.

Two Measures of National Product: Goods Flow and Earnings Flow

How do economists actually measure GDP? One of the major surprises is that we can measure GDP in two entirely independent ways. As Figure 20-1 shows, GDP can be measured either as a flow of products or as a sum of earnings.

To demonstrate the different ways of measuring GDP, we begin by considering an oversimplified world in which there is no government, no foreign

trade, and no investment. For the moment, our little economy produces only *consumption goods*, which are items that are purchased by households to satisfy their wants. (Important note: Our first example is oversimplified to show the basic ideas. In the realistic examples that follow, we will add investment, government, and the foreign sector.)

Flow-of-Product Approach. Each year the public consumes a wide variety of final goods and services: goods such as apples, computer software, and blue jeans; services such as health care and haircuts. We include only *final goods*—goods ultimately bought and used by consumers. Households spend their incomes for these consumer goods, as is shown in the upper loop of Figure 20-1. Add together all the consumption dollars spent on these final goods, and you will arrive at this simplified economy's total GDP.

Thus, in our simple economy, you can easily calculate national income or product as the sum of the annual flow of final goods and services: (price of blue jeans \times number of blue jeans) plus (price of apples \times number of apples) and so forth for all other final goods. The gross domestic product is defined as the total money value of the flow of final products produced by the nation.

National accountants use market prices as weights in valuing different commodities because market prices reflect the relative economic value of diverse goods and services. That is, the relative prices of different goods reflect how much consumers value their last (or marginal) units of consumption of these goods.

Earnings or Income Approach. The second and equivalent way to calculate GDP is the income accounts (also called the earnings or cost approach). Look at the lower loop in Figure 20-1. Through it flow all the costs of doing business; these costs include the wages paid to labor, the rents paid to land, the profits paid to capital, and so forth. But these business costs are also the earnings that households receive from firms. By measuring the annual flow of these earnings or incomes, statisticians will again arrive at the GDP.

Hence, a second way to calculate GDP is as the total of factor earnings (wages, interest, rents, and

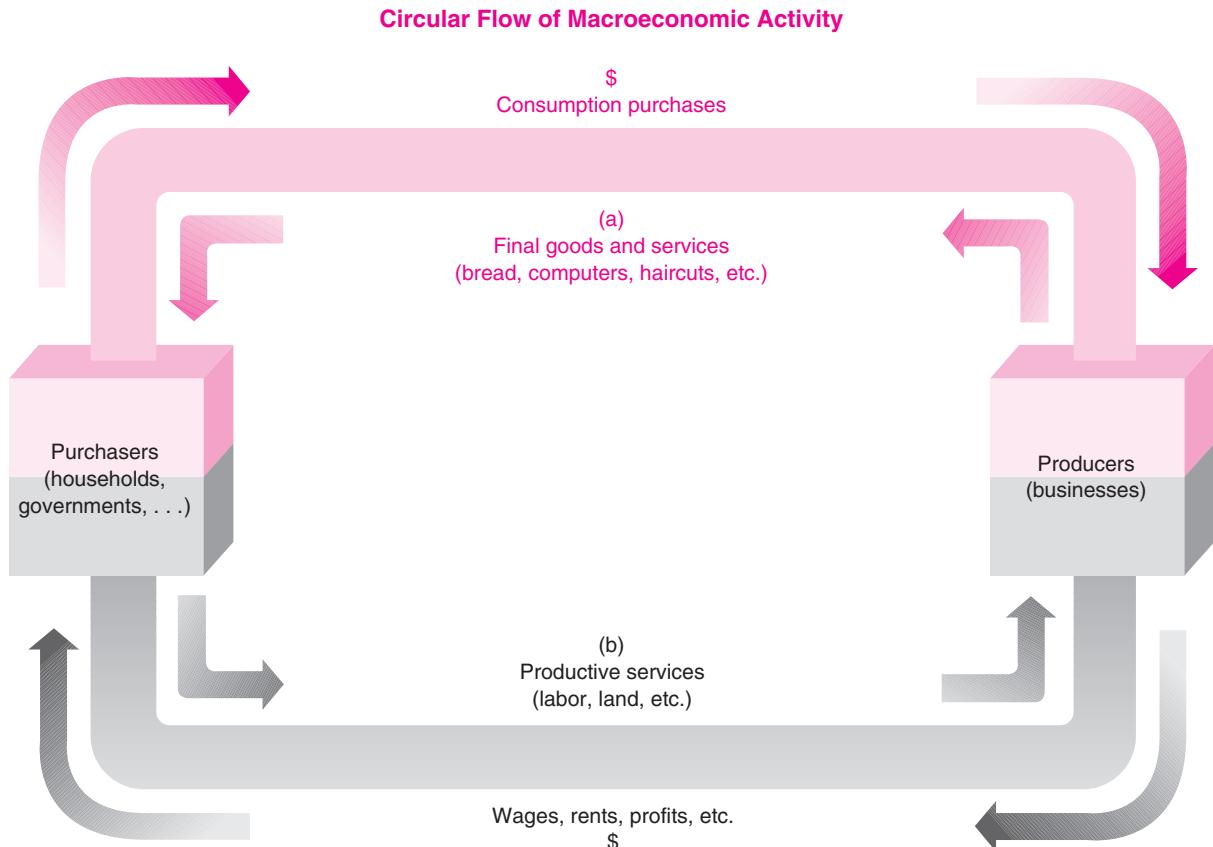


FIGURE 20-1. Gross Domestic Product Can Be Measured Either as (a) a Flow of Final Products or, Equivalently, as (b) a Flow of Earnings or Incomes

In the upper loop, purchasers buy final goods and services. The total dollar flow of their spending each year is one measure of gross domestic product. The lower loop measures the annual flow of costs of output: the earnings that businesses pay out in wages, rent, interest, dividends, and profits.

The two measures of GDP must always be identical. Note that this figure is the macroeconomic counterpart of Fig. 2-1, which presented the circular flow of supply and demand.

profits) that are the costs of producing society's final products.

Equivalence of the Two Approaches. Now we have calculated GDP by the upper-loop flow-of-product approach and by the lower-loop earnings-flow approach. Which is the better approach? The surprise is that *they are exactly the same*.

We can see why the product and earnings approaches are identical by examining a simple barbershop economy. Say the barbers have no expenses

other than labor. If they sell 10 haircuts at \$8 each, GDP is \$80. But the barbers' earnings (in wages and profits) are also exactly \$80. Hence, the GDP here is identical whether measured as a flow of products (\$80 worth of haircuts) or as a flow of costs and incomes (\$80 worth of wages and profits).

In fact, the two approaches are identical because we have included “profits” in the lower loop along with other incomes. What exactly is profit? Profit is what remains from the sale of a product after you have paid the other factor costs—wages, interest, and rents. It

(a) Income Statement of Typical Farm			
Output in Farming		Earnings	
Sales of goods (corn, apples, etc.)	\$1,000	Costs of production:	
		Wages	\$ 800
		Rents	100
		Interest	25
		Profits (residual)	75
Total	\$1,000	Total	\$1,000

(b) National Product Account (millions of dollars)			
Upper-Loop Flow of Product		Lower-Loop Flow of Earnings	
Final output ($10 \times 1,000$)	\$10,000	Costs or earnings:	
		Wages (10×800)	\$ 8,000
		Rents (10×100)	1,000
		Interest (10×25)	250
		Profits (10×75)	750
GDP total	\$10,000	GDP total	\$10,000

TABLE 20-1. Construction of National Product Accounts from Business Accounts

Part (a) shows the income statement of a typical farm. The left side shows the value of production, while the right side shows the farm's costs. Part (b) then adds up or aggregates the 10 million identical farms to obtain total GDP. Note that GDP from the product side exactly equals GDP from the earnings side.

is the residual that adjusts automatically to make the lower loop's costs or earnings exactly match the upper loop's value of goods and services.

To sum up:

GDP, or gross domestic product, can be measured in two different ways: (1) as the flow of spending on final products, or (2) as the total costs or incomes of inputs. Both approaches yield exactly the same measure of GDP.

National Accounts Derived from Business Accounts

You might wonder where on earth economists find all the data for the national accounts. In practice, government economists draw on a wide array of sources, including surveys, income-tax returns, retail-sales statistics, and employment data.

The most important source of data is business accounts. An *account* for a firm or nation is a numerical record of all flows (outputs, costs, etc.) during a

given period. We can show the relationship between business accounts and national accounts by constructing the accounts for an economy made up only of farms. The top half of Table 20-1 shows the results of a year's farming operations for a single, typical farm. We put sales of final products on the left-hand side and the various costs of production on the right. The bottom half of Table 20-1 shows how to construct the GDP accounts for our simple agrarian economy in which all final products are produced on 10 million identical farms. The national accounts simply add together or *aggregate* the outputs and costs of the 10 million identical farms to get the two different measures of GDP.

The Problem of “Double Counting”

We defined GDP as the total production of final goods and services. A *final product* is one that is produced and sold for consumption or investment. GDP excludes *intermediate goods*—goods that are used up to produce other goods. GDP therefore includes

bread but not flour, and home computers but not computer chips.

For the flow-of-product calculation of GDP, excluding intermediate products poses no major complications. We simply include the bread and home computers in GDP but avoid including the flour and yeast that went into the bread or the chips and plastic that went into the computers. If you look again at the upper loop in Figure 20-1, you will see that bread and computers appear in the flow of products, but you will not find any flour or computer chips.

What has happened to products like flour and computer chips? They are intermediate products and are simply cycling around inside the block marked "Producers." If they are not bought by consumers, they never show up as final products in GDP.

"Value Added" in the Lower Loop. A new statistician who is being trained to make GDP measurements might be puzzled, saying:

I can see that, if you are careful, your upper-loop product approach to GDP will avoid including intermediate products. But aren't you in some trouble when you use the lower-loop cost or earnings approach?

After all, when we gather income statements from the accounts of firms, won't we pick up what grain

merchants pay to wheat farmers, what bakers pay to grain merchants, and what grocers pay to bakers?

Won't this result in double counting or even triple counting of items going through several productive stages?

These are good questions, but there is an ingenious technique that resolves the problem. In making lower-loop earnings measurements, statisticians are very careful to include in GDP only a firm's value added. **Value added** is the difference between a firm's sales and its purchases of materials and services from other firms.

In other words, in calculating the GDP earnings or value added by a firm, the statistician includes all costs except for payments made to other businesses. Hence business costs in the form of wages, salaries, interest payments, and dividends are included in value added, but purchases of wheat or steel or electricity are excluded from value added. Why are all the purchases from other firms excluded from value added to obtain GDP? Because those purchases will get properly counted in GDP in the values added by other firms.

Table 20-2 uses the stages of bread production to illustrate how careful adherence to the value-added approach enables us to subtract purchases of intermediate goods that show up in the income statements

Bread Receipts, Costs, and Value Added (cents per loaf)				
Stage of production	(1) Sales receipts	(2) <i>Less: Cost of intermediate products</i>		(3) Value added (wages, profits, etc.) (3) = (1) - (2)
Wheat	23	0	=	23
Flour	53	23	=	30
Baked dough	110	53	=	57
Final product: bread	190	110	=	80
Total	376	186		190
			(sum of value added)	

TABLE 20-2. GDP Sums Up Value Added at Each Production Stage

To avoid double counting of intermediate products, we calculate value added at each stage of production. This involves subtracting all the costs of materials and intermediate products bought from other businesses from total sales. Note that every blue intermediate-product item both appears in column (1) and is subtracted in the next stage of production in column (2). (By how much would we overestimate GDP if we counted all receipts, not just value added? The overestimate would be 186 cents per loaf.)

of farmers, millers, bakers, and grocers. The final calculation shows the desired equality between (1) final sales of bread and (2) total earnings, calculated as the sum of all values added in all the different stages of bread production.

Value-added approach: To avoid double counting, we take care to include only final goods in GDP and to exclude the intermediate goods that are used up in making the final goods. By measuring the value added at each stage, taking care to subtract expenditures on the intermediate goods bought from other firms, the lower-loop earnings approach properly avoids all double counting and records wages, interest, rents, and profits exactly one time.

DETAILS OF THE NATIONAL ACCOUNTS

Now that we have an overview of the national income and product accounts, we will proceed, in the rest of this chapter, on a whirlwind tour of the various sectors. Before we start on the journey, look at Table 20-3 to get an idea of where we are going. This table shows a summary set of accounts for both the product and the income sides. If you understand the structure of the table and the definitions of the terms in it, you will be well on your way to understanding GDP and its family of components.

Real vs. Nominal GDP: “Deflating” GDP by a Price Index

We define GDP as the dollar value of goods and services. In measuring the dollar value, we use the measuring rod of *market prices* for the different goods and services. But prices change over time, as inflation generally sends prices upward year after year. Who would want to measure things with a rubber yardstick—one that stretches in your hands from day to day—rather than a rigid and invariant yardstick?

The problem of changing prices is one of the problems economists have to solve when they use money as their measuring rod. Clearly, we want a measure of the nation’s output and income that uses an invariant yardstick. Economists can replace the elastic yardstick with a reliable one by removing the price-increase component so as to create a real or quantity index of national output.

Here is the basic idea: We can measure the GDP for a particular year using the actual market prices of that year; this gives us the **nominal GDP**, or GDP at current prices. But we are usually more interested in determining what has happened to the **real GDP**, which is an index of the volume or quantity of goods and services produced. Real GDP is calculated by tracking the volume or quantity of production after removing the influence of changing prices or inflation. Hence, nominal GDP is calculated using changing prices, while real GDP represents the change in the volume of total output after price changes are removed.

Product Approach	Earnings Approach
Components of gross domestic product: Consumption (C) + Gross private domestic investment (I) + Government purchases (G) + Net exports (X)	Earnings or income approach to gross domestic product: Compensation of labor (wages, salaries, and supplements) + Corporate profits + Other property income (rent, interest, proprietors’ income) + Depreciation + Net production taxes
Equals: Gross domestic product	Equals: Gross domestic product

TABLE 20-3. Overview of the National Income and Product Accounts

This table presents the major components of the two sides of the national accounts. The left side shows the components of the product approach (or upper loop); the symbols C , I , G , and X are often used to represent these four items of GDP. The right side shows the components of the earnings or income approach (or lower loop). Each approach will ultimately add up to exactly the same GDP.

Date	(1) Nominal GDP (current \$, billion)	(2) Index number of prices (GDP deflator, 1929 = 1)	(3) Real GDP (\$, billion, 1929 prices) (3) = $\frac{(1)}{(2)}$
1929	104	1.00	$\frac{104}{1.00} = 104$
1933	56	0.74	$\frac{56}{0.74} = 76$

TABLE 20-4. Real (or Inflation-Corrected) GDP Is Obtained by Dividing Nominal GDP by the GDP Deflator

Using the price index of column (2), we deflate column (1) to get real GDP in column (3). (Riddle: Can you show that 1929's real GDP was \$77 billion in terms of 1933 prices?

Hint: With 1933 as a base of 1, 1929's price index is 1.35.)

The difference between nominal GDP and real GDP is the **price of GDP**, sometimes called the **GDP deflator**.

A simple example will illustrate the general idea. Say that a country produces 1000 bushels of corn in year 1 and 1010 bushels in year 2. The price of a bushel is \$1 in year 1 and \$2 in year 2. We can calculate nominal GDP (PQ) as $\$1 \times 1000 = \1000 in year 1 and $\$2 \times 1010 = \2020 in year 2. Nominal GDP therefore grew by 102 percent between the two years.

But the actual amount of output did not grow anywhere near that rapidly. To find real output, we need to consider what happened to prices. One common approach is to use the first year as the base year. The *base year* is the year in which we measure prices. We can, for index purposes, set the price index for the first year (the base year) at $P_1 = 1$. This means that output will be measured in prices of the base year. From the data in the previous paragraph, we see that the GDP deflator is $P_2 = \$2/\$1 = 2$ in year 2. Real GDP (Q) is equal to nominal GDP (PQ) divided by the GDP deflator (P). Hence real GDP was equal to $\$1000/1 = \1000 in year 1 and $\$2020/2 = \1010 in year 2. Thus the growth in real GDP, which corrects for the change in prices, is 1 percent and equals the growth in the output of corn, as it should.

A 1929–1933 comparison will illustrate the deflation process for an actual historical episode. Table 20-4 gives nominal GDP figures of \$104 billion for 1929 and

\$56 billion for 1933. This represents a 46 percent drop in nominal GDP from 1929 to 1933. But the government estimates that prices on average dropped about 26 percent over this period. If we choose 1929 as our base year, with the GDP deflator of 1 in that year, this means that the 1933 price index was 0.74. So our \$56 billion of GDP in 1933 was really worth much more than half the \$104 billion GDP of 1929. Table 20-4 shows that, in terms of 1929 prices, or dollars of 1929 purchasing power, real GDP fell to \$76 billion. Hence, part of the near-halving shown by the nominal GDP was due to the rapidly declining price level, or deflation, during the Great Depression.

The green line in Figure 20-2 shows the growth of nominal GDP since 1929, expressed in the actual dollars and prices that were current in each historical year. Then, for comparison, the real GDP, expressed in 2000 dollars, is shown in blue. Clearly, much of the increase in nominal GDP over the last eight decades was due to inflation in the price units of our money yardstick.

Table 20-4 shows the simplest way of calculating real GDP and the GDP deflator. Sometimes these calculations give misleading results, particularly when the relative prices and quantities of important goods are changing rapidly. For example, over the last three decades, computer prices have been falling very sharply while the quantity of computers produced has risen rapidly (we return to this issue in our discussion of price indexes below).

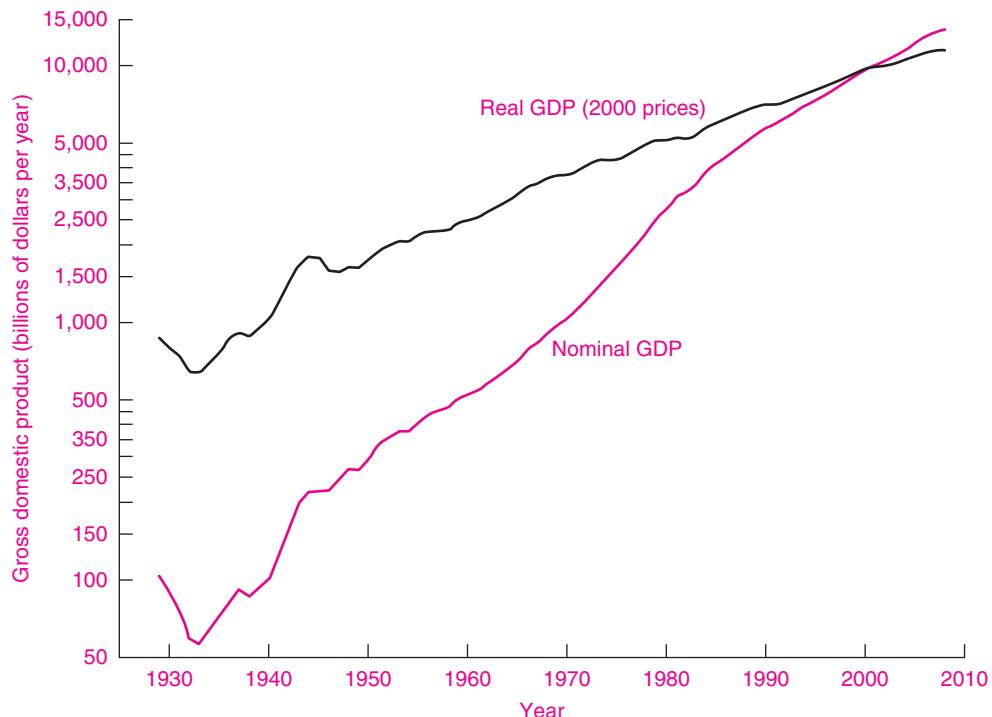


FIGURE 20-2. Nominal GDP Grows Faster than Real GDP because of Price Inflation

The rise in nominal GDP exaggerates the rise in output. Why? Because growth in nominal GDP includes increases in prices as well as growth in output. To obtain an accurate measure of real output, we must correct GDP for price changes.

Source: U.S. Bureau of Economic Analysis.

When relative prices of different goods are changing very rapidly, using prices of a fixed year will give a misleading estimate of real GDP growth. To correct for this bias, statisticians use a procedure known as *chain weighting*. Instead of the relative weights on each good being kept fixed (say, by the use of weights for a given year, like 1990), the weights of the different goods and services change each year to reflect the changes in spending patterns in the economy. Today, the official U.S. government measures of real GDP and the GDP price index rely upon chain weights. The technical names for these constructs are “real GDP in chained dollars” and the “chain-type price index for GDP.” As a shorthand, we generally refer to these as real GDP and the GDP price index.

Further Details on Chain Weights. The details of using chain weights are somewhat involved, but we can get the basic idea using a simple example. The

calculation of chain weights involves linking the output or price series together by multiplying the growth rates from one period to another. An example for a haircut economy will show how this works. Say that the value of the haircuts was \$300 in 2003. Further suppose that the quantity of haircuts increased by 1 percent from 2003 to 2004 and by 2 percent from 2004 to 2005. The value of real GDP (in chained 2003 dollars) would be \$300 in 2003, then $\$300 \times 1.01 = \303 in 2004, and then $\$303 \times 1.02 = \309.06 in 2005. With many different goods and services, we would add together the growth rates of the different components of apples, bananas, catamarans, and so on, and weight the growth rates by the expenditure or output shares of the different goods.

To summarize:

Nominal GDP (PQ) represents the total money value of final goods and services produced in a given year, where the values are expressed in terms of the

market prices of each year. Real GDP (Q) removes price changes from nominal GDP and calculates GDP in terms of the quantities of goods and services. The following equations provide the link between nominal GDP, real GDP, and the GDP price index:

$$Q = \text{real GDP} = \frac{\text{nominal GDP}}{\text{GDP price index}} = \frac{PQ}{P}$$

To correct for rapidly changing relative prices, the U.S. national accounts use chain weights to construct real GDP and price indexes.

Consumption

The first important part of GDP is consumption, or “personal consumption expenditures.” Consumption is by far the largest component of GDP, equaling about two-thirds of the total in recent years. Figure 20-3 shows the fraction of GDP devoted to

consumption over the last eight decades. Consumption expenditures are divided into three categories: durable goods such as automobiles, nondurable goods such as food, and services such as medical care. The most rapidly growing sector is services.

Investment and Capital Formation

So far, our analysis has banished all capital. In real life, however, nations devote part of their output to production of capital—durable items that increase future production. Increasing capital requires the sacrifice of current consumption to increase future consumption. Instead of eating more pizza now, people build new pizza ovens to make it possible to produce more pizza for future consumption.

In the accounts, **investment** consists of the additions to the nation’s capital stock of buildings, equipment, software, and inventories during a year. The



FIGURE 20-3. Share of Consumption in National Output Has Risen Recently

The share of consumption in total GDP rose during the Great Depression as investment prospects soured, then shrank sharply during World War II when the war effort displaced civilian needs. In recent years, consumption has grown more rapidly than total output as the national saving rate and government purchases have declined.

Source: U.S. Bureau of Economic Analysis.

national accounts include mainly tangible capital (such as buildings and computers) and omit most intangible capital (such as research-and-development or educational expenses).



Real Investment versus Financial Investment

Economists define “investment” (or sometimes *real investment*) as production of durable capital goods. In common usage, “investment” often denotes using money to buy General Motors stock or to open a savings account. For clarity, economists call this *financial investment*. Try not to confuse these two different uses of the word “investment.”

If I take \$1000 from my safe and buy some stocks, this is not what macroeconomists call investment. I have simply exchanged one financial asset for another. Investment takes place when a durable capital good is produced.

How does investment fit into the national accounts? Economic statisticians recognize that if people are using part of society’s production for capital formation, such outputs must be included in the upper-loop flow of GDP. Investments represent additions to the stock of durable capital that increase production possibilities in the future. So we must modify our original definition to read:

Gross domestic product is the sum of all final products. Along with consumption goods and services, we must also include gross investment.

Net vs. Gross Investment. Our revised definition includes “gross investment” along with consumption. What does the word “gross” mean in this context? It indicates that investment includes all investment goods produced. Gross investment is not adjusted for **depreciation**, which measures the amount of capital that has been used up in a year. Thus gross investment includes all the machines, factories, and houses built during a year—even though some were produced simply to replace old capital goods that burned down or were thrown on the scrap heap.

If you want to get a measure of the increase in society’s capital, gross investment is not a sensible

measure. Because it does not subtract depreciation, gross investment is too large a number—too gross.

An analogy to population will make clear the importance of considering depreciation. If you want to measure the increase in the size of the population, you cannot simply count the number of births, for this would clearly exaggerate the net change in population. To get population growth, you must also subtract the number of deaths.

The same point holds for capital. To find the net increase in capital, you must start with gross investment and subtract the deaths of capital in the form of depreciation, or the amount of capital used up.

Thus to estimate the increase in the capital stock we measure *net investment*. Net investment is always births of capital (gross investment) less deaths of capital (capital depreciation):

Net investment equals gross investment minus depreciation.

Government Purchases

Some of our national output is purchased by federal, state, and local governments, and these purchases are clearly part of our GDP. Some government purchases are consumption-type goods (like food for the military), while some are investment-type items (such as schools or roads). In measuring government’s contribution to GDP, we simply add all these government purchases to the flow of private consumption, private investment, and, as we will see later, net exports.

Hence, all the government payroll expenditures on its employees plus the costs of goods it buys from private industry (lasers, roads, and airplanes) are included in this third category of flow of products, called “government consumption expenditures and gross investment.” This category equals the contribution of federal, state, and local governments to GDP.

Exclusion of Transfer Payments. Does this mean that every dollar of government expenditure is included in GDP? Definitely not. GDP includes only government purchases; it excludes spending on transfer payments.

Government **transfer payments** are payments to individuals that are not made in exchange for goods or services supplied. Examples of government transfers include unemployment insurance, veterans’ benefits, and old-age or disability payments. These

payments meet important social purposes. But they are not purchases of current goods or services, and they are therefore omitted from GDP.

Thus if you teach in the local public school and receive a salary from the government, your salary is a factor payment and your services are included in GDP. If you receive a social security benefit as a retired worker, that payment is a transfer payment and is excluded from GDP. Similarly, government interest payments are treated as transfers and are excluded from GDP.

Finally, do not confuse the way the national accounts measure government spending on goods and services (G) with the official government budget. When the Treasury measures its expenditures, it includes purchases of goods and services (G) *plus* transfers.

Taxes. In using the flow-of-product approach to compute GDP, we need not worry about how the government finances its spending. It does not matter whether the government pays for its goods and services by taxing, by printing money, or by borrowing. Wherever the dollars come from, the statistician computes the governmental component of GDP as the actual cost to the government of the goods and services.

But while it is fine to ignore taxes in the flow-of-product approach, we must account for taxes in the earnings or cost approach to GDP. Consider wages, for example. Part of my wage is turned over to the government through personal income taxes. These direct taxes definitely do get included in the wage component of business expenses, and the same holds for direct taxes (personal or corporate) on interest, rent, and profits.

Or consider the sales tax and other indirect taxes that manufacturers and retailers have to pay on a loaf of bread (or on the wheat, flour, and dough stages). Suppose these indirect taxes total 10 cents per loaf, and suppose wages, profit, and other value-added items cost the bread industry 90 cents. What will the bread sell for in the product approach? For 90 cents? Surely not. The bread will sell for \$1, equal to 90 cents of factor costs plus 10 cents of indirect taxes.

Thus the cost approach to GDP includes both indirect and direct taxes as elements of the cost of producing final output.

Net Exports

The United States is an open economy engaged in importing and exporting goods and services. The last component of GDP—and an increasingly important one in recent years—is **net exports**, the difference between exports and imports of goods and services.

How do we draw the line between our GDP and other countries' GDPs? The U.S. GDP represents all goods and services produced within the boundaries of the United States. Production differs from sales in the United States in two respects. First, some of our production (Iowa wheat and Boeing aircraft) is bought by foreigners and shipped abroad, and these items constitute our *exports*. Second, some of what we consume at home (Mexican oil and Japanese cars) is produced abroad, and such items are American *imports*.

A Numerical Example. We can use a simple farming economy to understand how the national accounts work. Suppose that Agrovia produces 100 bushels of corn and 7 bushels are imported. Of these, 87 bushels are consumed (in C), 10 go for government purchases to feed the army (as G), and 6 go into domestic investment as increases in inventories (I). In addition, 4 bushels are exported, so net exports (X) are $4 - 7$, or minus 3.

What, then, is the composition of the GDP of Agrovia? It is the following:

$$\begin{aligned} \text{GDP} &= 87 \text{ of } C + 10 \text{ of } G + 6 \text{ of } I - 3 \text{ of } X \\ &= 100 \text{ bushels} \end{aligned}$$

Gross Domestic Product, Net Domestic Product, and Gross National Product

Although GDP is the most widely used measure of national output in the United States, two other concepts are frequently cited: net domestic product and gross national product.

Recall that GDP includes *gross* investment, which is net investment plus depreciation. A little thought suggests that including depreciation is rather like including wheat as well as bread. A better measure would include only *net* investment in total output. By subtracting depreciation from GDP we obtain **net domestic product** (NDP). If NDP is a sounder measure of a nation's output than GDP, why do national accountants focus on GDP? They do so because

- 1. GDP from the product side is the sum of four major components:**
 - Personal consumption expenditures on goods and services (C)
 - Gross private domestic investment (I)
 - Government consumption expenditures and gross investment (G)
 - Net exports of goods and services (X), or exports minus imports
- 2. GDP from the cost side is the sum of the following major components:**
 - Compensation (wages, salaries, and supplements)
 - Property income (corporate profits, proprietors' incomes, interest, and rents)
 - Production taxes and depreciation of capital

(Remember to use the value-added technique to prevent double counting of intermediate goods bought from other firms.)
- 3. The product and cost measures of GDP are identical** (by adherence to the rules of value-added bookkeeping and the definition of profit as a residual).
- 4. Net domestic product (NDP) equals GDP minus depreciation.**

TABLE 20-5. Key Concepts of the National Income and Product Accounts

depreciation is somewhat difficult to estimate, whereas gross investment can be estimated fairly accurately.

An alternative measure of national output, widely used until recently, is **gross national product** (GNP). What is the difference between GNP and GDP? GNP is the total output produced with labor or capital *owned by U.S. residents*, while GDP is the output produced with labor and capital *located inside the United States*.

For example, some of the U.S. GDP is produced in Honda plants that are owned by Japanese corporations operating in the U.S. The profits from these plants are included in U.S. GDP but not in U.S. GNP because Honda is a Japanese company. Similarly, when an American economist flies to Japan to give a paid lecture on baseball economics, payment for that lecture would be included in Japanese GDP and in American GNP. For the United States, GDP is very close to GNP, but these may differ substantially for very open economies.

To summarize:

Net domestic product (NDP) equals the total final output produced within a nation during a year, where output includes net investment, or gross investment less depreciation:

$$\text{NDP} = \text{GDP} - \text{depreciation}$$

Gross national product (GNP) is the total final output produced with inputs owned by the residents of a country during a year.

Table 20-5 provides a comprehensive definition of important components of GDP.

GDP and NDP: A Look at Numbers

Armed with an understanding of the concepts, we can turn to look at the actual data in the important Table 20-6.

Flow-of-Product Approach. Look first at the left side of Table 20-6. It gives the upper-loop, flow-of-product approach to GDP. Each of the four major components appears there, along with the dollar total for each component for 2007. Of these, C and G and their obvious subclassifications require little discussion.

Gross private domestic investment does require one comment. Its total (\$2130 billion) includes all new business investment, residential construction, and increase in inventory of goods. This gross total is the amount before a subtraction for depreciation of capital. After subtracting \$1721 billion of depreciation from gross investment, we obtain \$410 billion of net investment.

Finally, note the large negative entry for net exports, -\$708 billion. This negative entry represents the fact that in 2007 the United States imported \$708 billion more in goods and services than it exported.

Adding up the four components on the left gives the total GDP of \$13,808 billion. This is the harvest

Gross Domestic Product, 2007 (billions of current dollars)				
Production Approach		Earnings or Cost Approach		
1. Personal consumption expenditures		9,710	1. Compensation of employees	7,812
Durable goods	1,083		2. Proprietors' income	1,056
Nondurable goods	2,833		3. Rental income	40
Services	5,794		4. Net interest	664
2. Gross private domestic investment		2,130	5. Corporate profits (with adjustments)	1,642
Fixed investment			6. Depreciation	1,721
Nonresidential	1,504		7. Production taxes, statistical	
Residential	630		discrepancy, and miscellaneous	872
Change in private inventories	−4			
3. Net exports of goods and services		−708		
Exports	1,662			
Imports	2,370			
4. Government consumption expenditures and gross investment		2,675		
Federal	979			
State and local	1,696			
Gross domestic product		13,808	Gross domestic product	13,808

TABLE 20-6. The Two Ways of Looking at the GDP Accounts, in Actual Numbers

The left side measures flow of products (at market prices). The right side measures flow of costs (factor earnings and depreciation).

Source: U.S. Bureau of Economic Analysis.

we have been working for: the money measure of the American economy's overall performance for 2007.

Flow-of-Cost Approach. Now turn to the right-hand side of the table, which gives the lower-loop, flow-of-cost approach. Here we have all *costs of production* plus *taxes and depreciation*.

Compensation of employees represents wages, salaries, and other employee supplements. Net interest is a similar item.

Rent income of persons includes rents received by landlords. In addition, if you own your own home, you are treated as *paying rent to yourself*. This is one of many "imputations" (or derived data) in the national accounts. It makes sense if we really want to measure the housing services the American people are enjoying and do not want the estimate to change when people decide to own a home rather than rent one.

Production taxes are included as a separate item along with some small adjustments, including the

inevitable "statistical discrepancy," which reflects the fact that the officials never have every bit of needed data.¹

Depreciation on capital goods that were used up must appear as an expense in GDP, just like other expenses. Profit is a residual—what is left over after all other costs have been subtracted from total sales. There are two kinds of profits: profit of corporations and net earnings of unincorporated enterprises.

Income of unincorporated enterprises consists of earnings of partnerships and single-ownership businesses. This includes much farm and professional

¹ Statisticians work with incomplete reports and fill in data gaps by estimation. Just as measurements in a chemistry lab differ from the ideal, so do errors creep into both upper- and lower-loop GDP estimates. These are balanced by an item called the "statistical discrepancy." Along with the civil servants who are heads of units called "Wages," "Interest," and so forth, there actually used to be someone with the title "Head of the Statistical Discrepancy." If data were perfect, that individual would have been out of a job.

income. Finally, corporate profits before taxes are shown.

On the right side, the flow-of-cost approach gives us the same \$13,808 billion of GDP as does the flow-of-product approach. The right and left sides do agree.

From GDP to Disposable Income

The basic GDP accounts are of interest not only for themselves but also because of their importance for understanding how consumers and businesses behave. Some further distinctions will help illuminate the way the nation's books are kept.

National Income. To help us understand the division of total income among the different factors of production, we construct data on *national income* (*NI*). *NI* represents the total incomes received by labor, capital, and land. It is constructed by subtracting depreciation from GDP. National income equals total compensation of labor, rental income, net interest, income of proprietors, and corporate profits.

The relationship between GDP and national income is shown in the first two bars of Figure 20-4. The left-hand bar shows GDP, while the second bar shows the subtractions required to obtain *NI*.

From GDP to National Income to Disposable Income

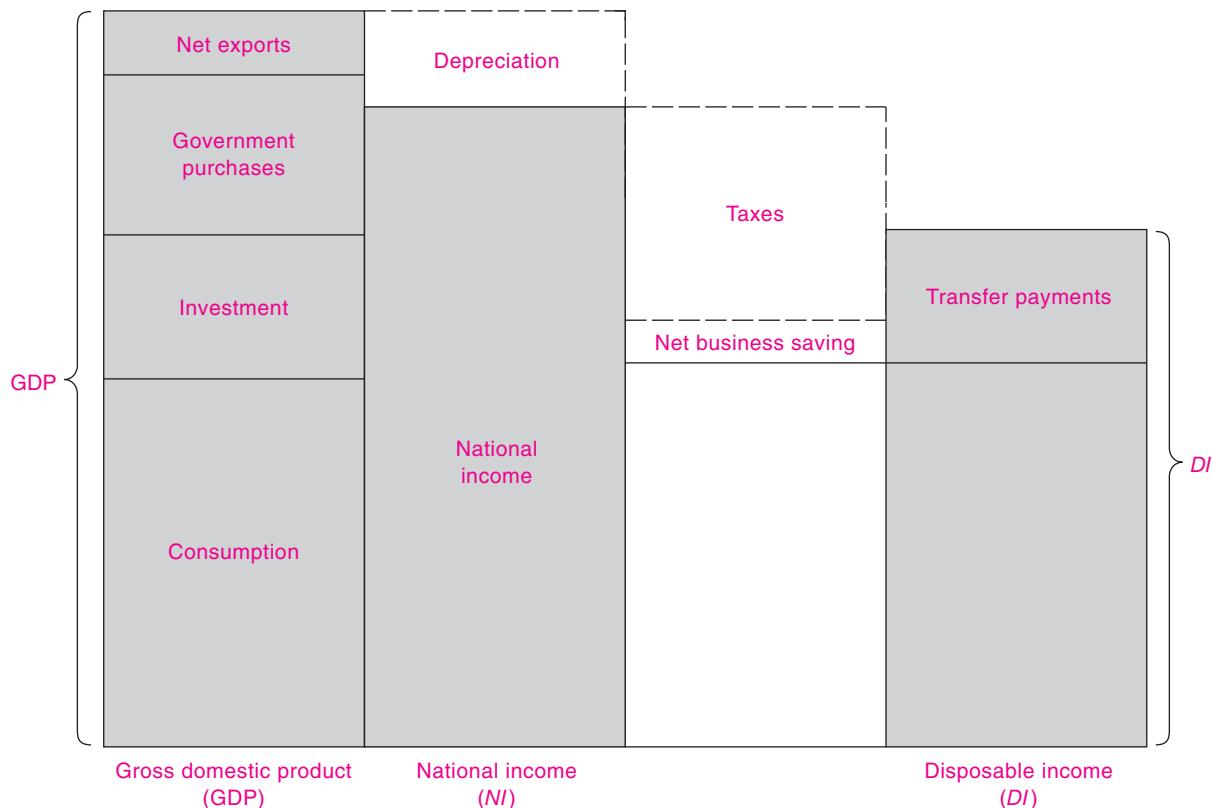


FIGURE 20-4. Starting with GDP, We Can Calculate National Income (NI) and Disposable Personal Income (DI)

Important income concepts are (1) GDP, which is total gross income to all factors; (2) national income, which is the sum of factor incomes and is obtained by subtracting depreciation from GDP; and (3) disposable personal income, which measures the total incomes of the household sector, including transfer payments but subtracting taxes.

Disposable Income. A second important concept asks, How many dollars per year do households actually have available to spend? The concept of disposable personal income (usually called **disposable income**, or *DI*) answers this question. To get disposable income, you calculate the market and transfer incomes received by households and subtract personal taxes.

Figure 20-4 shows the calculation of *DI*. We begin with national income in the second bar. We then subtract all taxes and further subtract net business saving. (Net business saving is profits after depreciation less dividends.) Finally, we add back the transfer payments that households receive from governments. This constitutes *DI*, shown as the right-hand bar in Figure 20-4. Disposable income is what actually gets into the hands of consumers to dispose of as they please. (This discussion omits some minor items such as the statistical discrepancy and net foreign factor incomes that are usually close to zero.)

As we will see in the next chapters, *DI* is what people divide between (1) consumption spending and (2) personal saving.

Saving and Investment

As we have seen, output can be either consumed or invested. Investment is an essential economic activity because it increases the capital stock available for future production. One of the most important points about national accounting is the identity between saving and investment. We will show that, under the accounting rules described above, *measured saving is exactly equal to measured investment*. This equality is an *identity*, which means that it must hold by definition.

In the simplest case, assume for the moment that there is no government or foreign sector. Investment is that part of national output which is not consumed. Saving is that part of national income which is not consumed. But since national income and output are equal, this means that saving equals investment. In symbols:

$$I = \text{product-approach GDP minus } C$$

$$S = \text{earnings-approach GDP minus } C$$

However, both approaches always give the same measure of GDP, so

$$I = S: \text{the identity between measured saving and investment}$$

That is the simplest case. We also need to consider the complete case which brings businesses, government, and net exports into the picture. On the saving side, *total or national saving (S^T)* is composed of *private saving* by households and businesses (S^P) along with *government saving (S^G)*. Government saving equals the government's budget surplus or the difference between tax revenues and expenditures.

On the investment side, total or *national investment (I^T)* starts with *gross private domestic investment (I)* but also adds *net foreign investment*, which is approximately the same as net exports (X). Hence, the complete saving-investment identity is given by²

$$\begin{aligned} \text{National investment} &= \text{private investment} + \text{net exports} \\ &= \text{private saving} + \text{government saving} = \text{national saving} \end{aligned}$$

or

$$I^T = I + X = S^P + S^G = S^T$$

National saving equals national investment by definition. The components of investment are private domestic investment and foreign investment (or net exports). The sources of saving are private saving (by households and businesses) and government saving (the government budget surplus). Private investment plus net exports equals private saving plus the budget surplus. These identities must hold always, whatever the state of the business cycle.

BEYOND THE NATIONAL ACCOUNTS

Advocates of the existing economic and social system often argue that market economies have produced a growth in real output never before seen in human history. "Look how GDP has grown because of the genius of free markets," say the admirers of capitalism.

But critics point out the deficiencies of GDP. GDP includes many questionable entries and omits many

² For this discussion, we consider only private investment and therefore treat all government purchases as consumption. In most national accounts today, government purchases are divided between consumption and tangible investments. If we include government investment, then this amount will add to both national investment and the government surplus.

valuable economic activities. As one dissenter said, “Don’t speak to me of all your production and your dollars, your gross domestic product. To me, GDP stands for gross domestic pollution!”

What are we to think? Isn’t it true that GDP includes government production of bombs and missiles along with salaries paid to prison guards? Doesn’t an increase in crime boost sales of home alarms, which adds to the GDP? Doesn’t cutting our irreplaceable redwoods show up as a positive output in our national accounts? Doesn’t GDP fail to account for environmental degradation such as acid rain and global warming?

In recent years, economists have begun developing new measures to correct the major defects of the standard GDP numbers and better reflect the true satisfaction-producing outputs of our economy. The new approaches attempt to extend the boundaries of the traditional accounts by including important nonmarket activities as well as correcting for harmful activities that are included as part of national output. Let’s consider some of the omitted pluses and minuses.

Omitted Nonmarket Activities. Recall that the standard accounts include primarily market activities. Much useful economic activity takes place outside the market. For example, college students are investing in human capital. The national accounts record the tuition, but they omit the opportunity costs of earnings forgone. Studies indicate that inclusion of nonmarket investments in education and other areas would more than double the national saving rate.

Similarly, many household activities produce valuable “near-market” goods and services such as meals, laundering, and child-care services. Recent estimates of the value of unpaid household work indicate that it might be half as large as total market consumption. Perhaps the largest omission from the market accounts is the value of leisure time. On average, Americans spend as much of their time on utility-producing leisure activities as they do on money-producing work activities. Yet the value of leisure time is excluded from our official national statistics.

You might wonder about the underground economy, which covers a wide variety of market activities that are not reported to the government. These include activities like gambling, prostitution, drug

dealing, work done by illegal immigrants, bartering of services, and smuggling. Actually, much underground activity is intentionally excluded because national output excludes illegal activities—these are by social consensus “bads” and not “goods.” A swelling cocaine trade will not enter into GDP. For legal but unreported activities, like unreported tips, the Commerce Department makes estimates on the basis of surveys and audits by the Internal Revenue Service.

Omitted Environmental Damage. In addition to omitting activities, sometimes GDP omits some of the harmful side effects of economic activity. An important example is the omission of environmental damages. For example, suppose the residents of Suburbia buy 10 million kilowatt-hours of electricity to cool their houses, paying Utility Co. 10 cents per kilowatt-hour. That \$1 million covers the labor costs, plant costs, and fuel costs. But suppose the company damages the neighborhood with pollution in the process of producing electricity. It incurs no monetary costs for this externality. Our measure of output should not only add in the value of the electricity (which GDP does) but also subtract the environmental damage caused by the pollution (which GDP does not).

Suppose that in addition to 10 cents of direct costs, there are 2 cents per kilowatt-hour of environmental damages to human health. These are the “external costs” of pollution not paid by Utility Co., and they total \$200,000. To correct for this hidden cost in a set of augmented accounts, we should subtract \$200,000 of “pollution bads” from the \$1,000,000 flow of “electricity goods.” In fact, government statisticians do *not* subtract pollution costs in the economic accounts.

Economists have made considerable progress in developing *augmented national accounts*, which are designed to include activities beyond the traditional definitions of the national accounts. The general principle of augmented accounting is to include as much of economic activity as is feasible, whether or not that activity takes place in the market. Examples of augmented accounts include estimates of the value of research and development, nonmarket investments in human capital, the value of unpaid production in the home, the value of forests, and the value of leisure time. Economists are even developing accounts for the damages from air pollution

and global warming. When these further accounts are completed, we will have a more comprehensive financial picture of the economy.

But be warned that even the most refined economic accounts still measure only economic activity. They do not attempt to—indeed, cannot—measure the ultimate satisfactions, pleasures, or pains of people in their everyday lives. This point was eloquently put by Robert Kennedy in one of his last speeches:

The gross national product does not allow for the health of our children, the quality of their education, or the joy of their play. It does not include the beauty of our poetry or the strength of our marriages; the intelligence of our public debate or the integrity of our public officials. It measures neither our wit nor our courage; neither our wisdom nor our learning; neither our compassion nor our devotion to our country.

PRICE INDEXES AND INFLATION

This chapter has up to now focused on measuring national output and its components. But people today worry about overall price trends, that is to say, about inflation. What do these terms mean?

Let us begin with a careful definition:

A price index (with symbol P) is a measure of the average level of prices. **Inflation** (with symbol π , or “pi”) denotes a rise in the general level of prices. The **rate of inflation** is defined as the rate of change of the general price level and is measured as follows:

$$\text{Rate of inflation in year } t = \pi_t = 100 \times \frac{P_t - P_{t-1}}{P_{t-1}}$$

Most periods in recent history have been ones of positive inflation. The opposite of inflation is **deflation**, which occurs when the general price level is falling. Deflations have been rare in the last half-century. In the United States, the last time consumer prices actually fell from one year to the next was 1955. Sustained deflations, in which prices fall steadily over a period of several years, are associated with depressions, such as those that occurred in the United States in the 1890s and the 1930s. More recently, Japan experienced a deflation over much of the last two decades as its economy suffered a prolonged recession.

Price Indexes

When newspapers tell us “Inflation is rising,” they are really reporting the movement of a price index. A price index is a weighted average of the price of a basket of goods and services. In constructing price indexes, economists weight individual prices by the economic importance of each good. The most important price indexes are the consumer price index, the GDP price index, and the producer price index.

The Consumer Price Index (CPI). The most widely used measure of the overall price level is the consumer price index, also known as the CPI, calculated by the U.S. Bureau of Labor Statistics (BLS). The CPI is a measure of the average price paid by urban consumers for a market basket of consumer goods and services. Each month, government statisticians record the prices of around 80,000 goods and services for more than 200 major categories. The prices are then arranged into the following eight major groups, listed with some examples:

- Food and beverages (breakfast cereal, milk, and snacks)
- Housing (rent of primary residence, owner's equivalent rent, bedroom furniture)
- Apparel (shirts and sweaters, jewelry)
- Transportation (new vehicles, gasoline, motor vehicle insurance)
- Medical care (prescription drugs, physicians' services, eyeglasses)
- Recreation (televisions, sports equipment, admissions)
- Education and communication (college tuition, computer software)
- Other goods and services (haircuts, funeral expenses)

How are the different prices weighted in constructing price indexes? It would clearly be silly merely to add up the different prices or to weight them by their mass or volume. Rather, a price index is constructed by *weighting each price according to the economic importance of the commodity in question*.

In the case of the traditional CPI, each item is assigned a fixed weight proportional to its relative importance in consumer expenditure budgets; the weight for each item is proportional to the total spending by consumers on that item as determined by a survey of consumer expenditures in the

2005–2006 period. As of 2008, housing-related costs were the single biggest category in the CPI, taking up more than 42 percent of consumer spending budgets. By comparison, the cost of new cars and other motor vehicles accounts for only 7 percent of the CPI's consumer expenditure budgets.



Calculating the CPI

It is worth spending a moment on the exact technique that is used to calculate CPI changes. The formula in the text is correct, but we need to explain how the formula works when there are many goods and services. The change in the overall CPI is the weighted average of the change of the components:

% change in CPI in period t

$$= 100 \times \left\{ \sum_{\text{All items}} [\text{weight of good } i \text{ in } (t - 1)] \times [\% \text{ change in the price of good } i \text{ from } (t - 1) \text{ to } t] \right\}$$

To take a concrete example, the following table shows the actual price-change and relative-importance data:

Expenditure category	Relative importance, December 2007 (%)	Percentage change over the last year
Food and beverages	14.9	4.4
Housing	42.4	3.0
Apparel	3.7	-1.4
Transportation	17.7	8.2
Medical care	6.2	4.6
Recreation	5.6	1.3
Education and communication	6.1	3.0
Other goods and services	3.3	3.2
All items	100.0	4.0

The rate of inflation over the period from March 2007 to March 2008 is seen to be 4.0 percent per year. (Question 9 at the end of this chapter examines this calculation further.)

This example captures the essence of how the traditional CPI measures inflation. The only difference between

this simplified calculation and the actual ones is that the CPI contains many more commodities and regions. Otherwise, the procedure is exactly the same.

GDP Price Index. Another widely used price index is the *GDP price index* (also sometimes referred to as the GDP deflator), which we met earlier in this chapter. The GDP price index is the price of all goods and services produced in the country (consumption, investment, government purchases, and net exports) rather than of a single component (such as consumption). This index also differs from the traditional CPI because it is a chain-weighted index that takes into account the changing shares of different goods (see the discussion of chain weights on page 393). In addition, there are price indexes for components of GDP, such as for investment goods, computers, personal consumption, and so forth, and these are sometimes used to supplement the CPI.

The Producer Price Index (PPI). This index, dating from 1890, is the oldest continuous statistical series published by the BLS. It measures the level of prices at the wholesale or producer stage. It is based on over 8000 commodity prices, including prices of foods, manufactured products, and mining products. The fixed weights used to calculate the PPI are the net sales of each commodity. Because of its great detail, this index is widely used by businesses.



Getting the Prices Right

Measuring prices accurately is one of the central issues of empirical economics. Price indexes affect not only obvious things like the inflation rate. They also are embedded in measures of real output and productivity. And through government policies, they affect monetary policy, taxes, government transfer programs like social security, and many private contracts.

The purpose of the consumer price index is to measure the cost of living. You might be surprised to learn that this is a difficult task. Some problems are intrinsic to price indexes. One issue is the *index-number problem*, which involves how the different prices are weighted or averaged. Recall that the traditional CPI uses a fixed weight for each good. As a result, the cost of living is

overestimated compared to the situation where consumers substitute relatively inexpensive for relatively expensive goods.

The case of energy prices can illustrate the problem. When gasoline prices rise sharply, people tend to reduce their gasoline purchases, buy smaller cars, and travel less. Yet the CPI assumes that they buy the same quantity of gasoline even though gasoline prices may have doubled. The overall rise in the cost of living is thereby exaggerated. Statisticians have devised ways of minimizing such index-number problems by using different weighting approaches, such as adjusting the weights as expenditures change, but government statisticians are just beginning to experiment with these newer approaches for the CPI.

A more important problem arises because of the difficulty of adjusting price indexes to capture the contribution of *new and improved goods and services*. An example will illustrate this problem. In recent years, consumers have benefited from compact fluorescent lightbulbs; these lightbulbs deliver light at approximately one-fourth the cost of the older, incandescent bulbs. Yet none of the price indexes incorporate the quality improvement. Similarly, as CDs and MP3s replaced long-playing records, as cable TV with hundreds of channels replaced the older technology with a few fuzzy channels, as air travel replaced rail or road travel, and in thousands of other improved goods and services, the price indexes did not reflect the improved quality.

Recent studies indicate that if quality change had been properly incorporated into price indexes, the CPI would have risen less rapidly in recent years. This problem is especially acute for medical care. In this sector, reported prices have risen sharply in the last two decades. Yet we have no adequate measure of the quality of medical care, and the CPI completely ignores the introduction of new products, such as pharmaceuticals which replace intrusive and expensive surgery.

A panel of distinguished economists led by Stanford's Michael Boskin examined this issue and estimated that the upward bias in the CPI was slightly more than 1 percent per year. This is a small number with large implications. It indicates that our real-output numbers may have been underestimated by the same amount. If the CPI bias carries through to the GDP deflator, then the growth in output per hour worked in the United States would be understated by around 1 percent per year.

This finding also implies that cost-of-living adjustments (which are used for social security benefits and the tax system) have overcompensated people for changes in the

cost of living. The bias would have substantial effects on overall taxes and benefits over a period of many years. Price indexes are not just abstruse concepts of interest only to a handful of technicians. Proper construction of price and output indexes affects our government budgets, our retirement programs, and even the way we assess our national economic performance.

In response to its own research and to its critics, the BLS has undertaken a major overhaul of the CPI. The most important innovation was the publication starting in 2002 of a "chained consumer price index" that augments the fixed-weight price index with a changing-weight system (like the chain weights used in the GDP accounts discussed on page 393 above) that accounts for consumer substitution. Over the decade since it was published, the chain CPI did indeed rise more slowly than the traditional CPI. It appears that critics were correct that the traditional CPI overstates inflation, although the size of the overstatement is likely to be less than the large number estimated by the Boskin Commission.³

ACCOUNTING ASSESSMENT

This chapter has examined the way economists measure national output and the overall price level. Having reviewed the measurement of national output and analyzed the shortcomings of the GDP, what should we conclude about the adequacy of our measures? Do they capture the major trends? Are they adequate measures of overall social welfare? The answer was aptly stated in a review by Arthur Okun:

It should be no surprise that national prosperity does not guarantee a happy society, any more than personal prosperity ensures a happy family. No growth of GDP can counter the tensions arising from an unpopular and unsuccessful war, a long overdue self-confrontation with conscience on racial injustice, a volcanic eruption of sexual mores, and an unprecedented assertion of independence by the young. Still, prosperity . . . is a precondition for success in achieving many of our aspirations.⁴

³ See this chapter's Further Reading section for a symposium on CPI design.

⁴ *The Political Economy of Prosperity* (Norton, New York, 1970), p. 124.



SUMMARY

- The national income and product accounts contain the major measures of income and product for a country. The gross domestic product (GDP) is the most comprehensive measure of a nation's production of goods and services. It comprises the dollar value of consumption (C), gross private domestic investment (I), government purchases (G), and net exports (X) produced within a nation during a given year. Recall the formula:

$$\text{GDP} = C + I + G + X$$

This will sometimes be simplified by combining private domestic investment and net exports into total gross national investment ($I^T = I + X$):

$$\text{GDP} = C + I^T + G$$

- We can match the upper-loop, flow-of-product measurement of GDP with the lower-loop, flow-of-cost measurement, as shown in Figure 20-1. The flow-of-cost approach uses factor earnings and carefully computes value added to eliminate double counting of intermediate products. And after summing up all (before-tax) wage, interest, rent, depreciation, and profit income, it adds to this total all indirect tax costs of business. GDP does not include transfer items such as social security benefits.
- By use of a price index, we can "deflate" nominal GDP (GDP in current dollars) to arrive at a more accurate measure of real GDP (GDP expressed in dollars of some base year's purchasing power). Use of such a price index corrects for the "rubber yardstick" implied by changing levels of prices.
- Net investment is positive when the nation is producing more capital goods than are currently being used up in the form of depreciation. Since depreciation is hard to estimate accurately, statisticians have more confidence in their measures of gross investment than in those of net investment.
- National income and disposable income are two additional official measurements. Disposable income (DI) is what people actually have left—after all tax payments, corporate saving of undistributed profits, and transfer adjustments have been made—to spend on consumption or to save.
- Using the rules of the national accounts, measured saving must exactly equal measured investment. This is easily seen in a hypothetical economy with nothing but households. In a complete economy, *private*

saving and government surplus equal domestic investment plus net foreign investment. The identity between saving and investment is just that: saving must equal investment no matter whether the economy is in boom or recession, war or peace. It is a consequence of the definitions of national income accounting.

- Gross domestic product and even net domestic product are imperfect measures of genuine economic welfare. In recent years, statisticians have started correcting for nonmarket activities such as unpaid work at home and environmental externalities.
- Inflation occurs when the general level of prices is rising (and deflation occurs when it is falling). We measure the overall price level and rate of inflation using price indexes—weighted averages of the prices of thousands of individual products. The most important price index is the consumer price index (CPI), which traditionally measured the cost of a fixed market basket of consumer goods and services relative to the cost of that bundle during a particular base year. Recent studies indicate that the CPI trend has a major upward bias because of index-number problems and omission of new and improved goods, and the government has undertaken steps to correct some of this bias.
- Recall the useful formulas from this and the prior chapter:
 - For calculating single-period growth of GDP:
$$\begin{aligned} &\text{Growth of real GDP in year } t \\ &= 100 \times \frac{\text{GDP}_t - \text{GDP}_{t-1}}{\text{GDP}_{t-1}} \end{aligned}$$
 - For calculating inflation with a single good:
$$\text{Rate of inflation in year } t = \pi_t = 100 \times \frac{P_t - P_{t-1}}{P_{t-1}}$$
 - Multiyear growth rate:
$$\text{Growth from } (t-n) \text{ to } t:$$

$$g_t^{(n)} = 100 \times \left[\left(\frac{X_t}{X_{t-n}} \right)^{1/n} - 1 \right]$$
 - For calculating the CPI with multiple goods:
$$\begin{aligned} &\% \text{ change in CPI} \\ &= 100 \times \left[\sum_{\text{All items}} (\text{weight}_i) \times (\% \text{ change } p_i) \right] \end{aligned}$$

CONCEPTS FOR REVIEW

national income and product accounts (national accounts)
 real and nominal GDP
 GDP deflator
 $GDP = C + I + G + X$
 net investment =
 gross investment – depreciation
 GDP in two equivalent views:
 product (upper loop)
 earnings (lower loop)

intermediate goods, value added
 $NDP = GDP - \text{depreciation}$
 government transfers
 disposable income (DI)
 investment-saving identity:
 $I = S$
 $I^T = I + X = S^P + S^G = S^T$
 inflation, deflation

price index:
 CPI
 GDP price index
 PPI
 growth-rate formulas

FURTHER READING AND INTERNET WEBSITES

Further Reading

A magnificent compilation of historical data on the United States is Susan Carter et al., *Historical Statistics of the United States: Millennial Edition* (Cambridge, 2006). This is available online from many college websites at hsus.cambridge.org/HSUSWeb/HSUSEntryServlet. A review of the issues involving measuring the consumer price index is contained in “Symposium on the CPI,” *Journal of Economic Perspectives*, Winter 1998.

Robert Kennedy’s remarks are from “Recapturing America’s Moral Vision,” March 18, 1968, in *RFK: Collected Speeches* (Viking Press, New York, 1993).

Websites

The premium site for the U.S. national income and product accounts is maintained by the Bureau of Economic Analysis (BEA) at www.bea.gov. This site also contains

issues of *The Survey of Current Business*, which discusses recent economic trends.

A comprehensive launching pad for government data in many areas is “FRED,” assembled by the Federal Reserve Bank of St. Louis at research.stlouisfed.org/fred2. The best single statistical source for data on the United States is *The Statistical Abstract of the United States*, published annually. It is available online at www.census.gov/compendia/statab/. Many important data sets can be found at www.economagic.com/.

A recent review of alternative approaches to augmented and environmental accounting is contained in a report by the National Academy of Sciences in William Nordhaus and Edward Kokkelenberg, eds., *Nature’s Numbers: Expanding the National Accounts to Include the Environment* (National Academy Press, Washington, D.C., 1999), available at www.nap.edu.

QUESTIONS FOR DISCUSSION

1. Define carefully the following and give an example of each:
 - a. Consumption
 - b. Gross private domestic investment
 - c. Government consumption and investment purchases (in GDP)
 - d. Government transfer payments (not in GDP)
 - e. Exports
2. You sometimes hear, “You can’t add apples and oranges.” Show that we can and do add apples and oranges in the national accounts. Explain how.
3. Examine the data in the appendix to Chapter 19. Locate the figures for nominal and real GDP for 2006 and 2007. Calculate the GDP deflator. What were the rates of growth of nominal GDP and real GDP for 2007? What was the rate of inflation (as measured by

- the GDP deflator) for 2007? Compare the rate of inflation using the GDP deflator with that using the CPI.
4. Robinson Crusoe produces upper-loop product of \$1000. He pays \$750 in wages, \$125 in interest, and \$75 in rent. What must his profit be? If three-fourths of Crusoe's output is consumed and the rest invested, calculate Crusoeland's GDP with both the product and the income approaches and show that they must agree exactly.
5. Here are some brain teasers. Can you see why the following are not counted in U.S. GDP?
- The gourmet meals produced by a fine home chef
 - The purchase of a plot of land
 - The purchase of an original Rembrandt painting
 - The value I get in 2009 from playing a 2005 compact disc
 - Damage to houses and crops from pollution emitted by electric utilities
 - Profits earned by IBM on production in a British factory
6. Consider the country of Agrovia, whose GDP is discussed in "A Numerical Example" on page 396. Construct a set of national accounts like that in Table 20-6 assuming that wheat costs \$5 per bushel, there is no depreciation, wages are three-fourths of national output, indirect business taxes are used to finance 100 percent of government spending, and the balance of income goes as rent income to farmers.
7. Review the discussion of bias in the CPI. Explain why failure to consider the quality improvement of a new good leads to an upward bias in the trend of the CPI.
- Pick a good you are familiar with. Explain how its quality has changed and why it might be difficult for a price index to capture the increase in quality.
8. In recent decades, women have worked more hours in paid jobs and fewer hours in unpaid housework.
- How would this increase in work hours affect GDP?
 - Explain why this increase in measured GDP will overstate the true increase in output. Also explain how a set of augmented national accounts which includes home production would treat this change from nonmarket work to market work.
 - Explain the paradox, "When a person marries his or her gardener, GDP goes down."
9. Examine the price-change numbers shown in the example on page 403.
- Use the formula to calculate the increase in the CPI from March 2007 to March 2008 to two decimal places. Verify that the number shown in the table is correct to a single decimal place.
 - The level of the CPI in March 2007 was 205.10. Calculate the CPI for March 2008.
10. Robert Kennedy's remarks about the shortcomings of measures of national output also contained the following: "The Gross National Product includes air pollution and advertising for cigarettes, and ambulances to clear our highways of carnage. It counts special locks for our doors, and jails for the people who break them. GNP includes the destruction of the redwoods and the death of Lake Superior." List ways that the accounts can be redesigned to incorporate these effects.



Micawber's equation:

Income 20 pounds; expenditure 19 pounds, 19 shillings and sixpence = happiness.

Income 20 pounds; annual expenditure 20 pounds and sixpence = misery.

Charles Dickens

David Copperfield

The major components of national output are consumption and investment. Naturally, nations want high levels of consumption—items such as housing, food, education, and recreation. The purpose of the economy is, after all, to transform inputs like labor and capital into consumption.

But saving and investment—that part of output that is not consumed—also play a central role in a nation’s economic performance. Nations that save and invest large fractions of their incomes tend to have rapid growth of output, income, and wages; this pattern characterized the United States in the nineteenth century, Japan in the twentieth century, and the miracle economies of East Asia in recent decades. By contrast, nations that consume most of their incomes, like many poor countries in Africa and Latin America, have obsolete capital, low educational standards, and backward techniques; they experience low rates of growth of productivity and real wages. High consumption relative to income spells low investment and slow growth; high saving leads to high investment and rapid growth.

The interaction between spending and income plays quite a different role during business-cycle expansions and contractions. When consumption

grows rapidly, this increases total spending or aggregate demand, raising output and employment in the short run. America’s economic boom of the late 1990s was largely fueled by rapid growth in consumer spending, but when American consumers tightened their belts, this contributed to the recession of 2007–2009.

Because consumption and investment are so central to macroeconomics, we devote this chapter to them.

A. CONSUMPTION AND SAVING

This section considers consumption and saving behavior, beginning with individual spending patterns and then looking at aggregate consumption behavior. Recall from Chapter 20 that *consumption* (or, more precisely, personal consumption expenditures) is expenditures by households on final goods and services. *Saving* is that part of personal disposable income that is not consumed.

Consumption is the largest single component of GDP, constituting 70 percent of total spending over the last decade. What are the major elements of consumption? Among the most important categories

Category of consumption	Value of consumption (\$, billion, 2007)	Percent of total
Durable goods	1,083	11.2%
Motor vehicles and parts	440	
Furniture and household equipment	415	
Other	227	
Nondurable goods	2,833	29.2%
Food	1,329	
Clothing and shoes	374	
Energy goods	367	
Other	763	
Services	5,794	59.7%
Housing	1,461	
Household operation	526	
Transportation	357	
Medical care	1,681	
Recreation	403	
Other	1,366	
Total personal consumption expenditures	9,710	100.0%

TABLE 21-1. The Major Components of Consumption

We divide consumption into three categories: durable goods, nondurable goods, and services. The service sector is growing in importance as basic needs for food are met and as health, recreation, and education claim a larger part of family budgets.

Source: U.S. Bureau of Economic Analysis, available at www.bea.gov.

are housing, motor vehicles, food, and medical care. Table 21-1 displays the major elements, broken down into the three main categories of durable goods, nondurable goods, and services. The items themselves are familiar, but their relative importance, particularly the increasing importance of services, is worth a moment's study.

Budgetary Expenditure Patterns

How do the patterns of consumption spending differ across different households in the United States? No two families spend their disposable incomes in exactly the same way. Yet statistics show that there is a predictable regularity in the way people allocate their expenditures among food, clothing, and other major items. The thousands of budgetary investigations of household spending patterns show remarkable agreement on the general, qualitative

patterns of behavior.¹ Figure 21-1 on page 410 tells the story.

Poor families must spend their incomes largely on the necessities of life: food and shelter. As income increases, expenditure on many food items goes up. People eat more and eat better. There are, however, limits to the extra money people will spend on food when their incomes rise. Consequently, the proportion of total spending devoted to food declines as income increases.

Expenditure on clothing, recreation, and automobiles increases more than proportionately to after-tax

¹ The spending patterns shown in Fig. 21-1 are called "Engel's Laws," after the nineteenth-century Prussian statistician Ernst Engel. The average behavior of consumption expenditure does change fairly regularly with income. But averages do not tell the whole story. Within each income class, there is a considerable spread of consumption around the average.

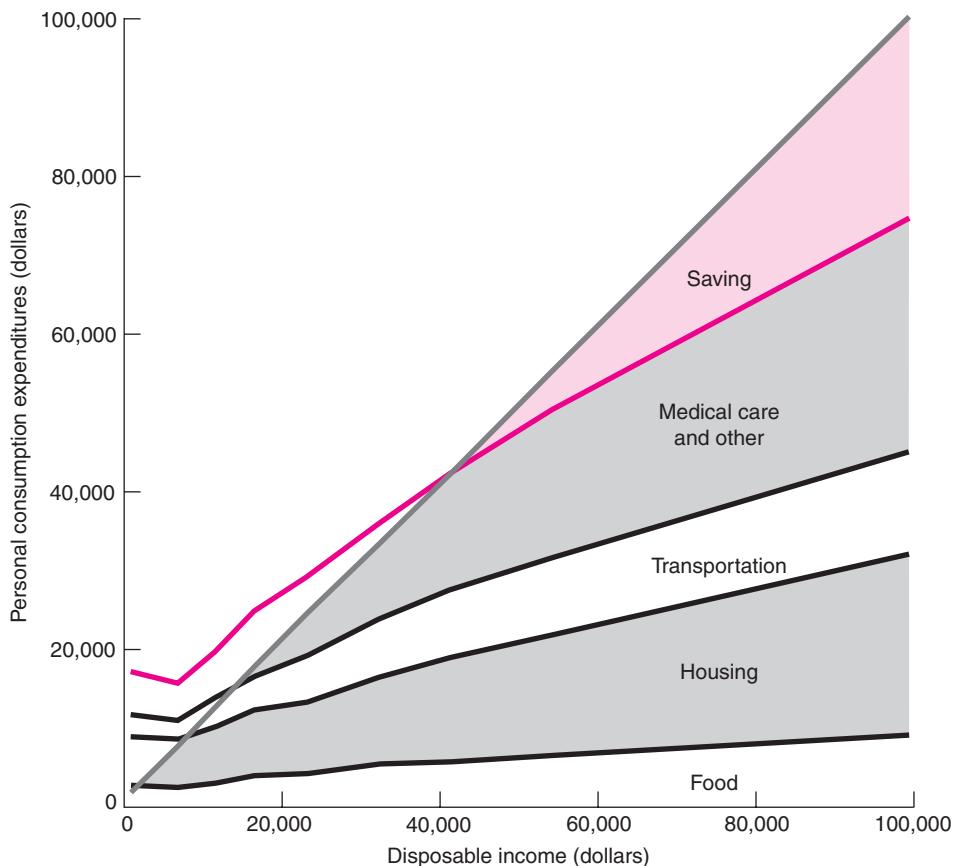


FIGURE 21-1. Family Budget Expenditures Show Regular Patterns

Surveys verify the importance of disposable income as a determinant of consumption expenditures. Notice the drop in food as a percentage of income as incomes rise. Note also that saving is negative at low incomes but rises substantially at high incomes.

Source: U.S. Department of Labor, *Consumer Expenditure Surveys*, 1998, available on the Internet at www.bls.gov/csxstd.htm.

income, until high incomes are reached. Spending on luxury items increases in greater proportion than income. Finally, as we look across families, note that saving rises rapidly as income increases. Saving is the greatest luxury of all.



The Evolution of Consumption in the Twentieth Century

Continual changes in technology, incomes, and social forces have led to dramatic changes in U.S. consumption patterns over time. In 1918, American households on average spent 41 percent of their

incomes on food and drink. By comparison, households now spend only about 14 percent on these items. What lies behind this striking decline? The major factor is that spending on food tends to grow more slowly than incomes. Similarly, spending on apparel has fallen from 18 percent of household income at the beginning of the twentieth century to only 4 percent today.

What are the “luxury goods” that Americans are spending more on? One big item is transportation. In 1918, Americans spent only 1 percent of their incomes on vehicles—but of course Henry Ford didn’t sell his first Model T until 1908. Today, there are 1.2 cars for every licensed driver in the United States. It is not surprising that

11 cents out of every dollar of spending goes for automotive transportation expenses. What about recreation and entertainment? Households now lay out large sums for televisions, cellular phones, and digital video recorders, items that did not exist 75 years ago. Housing services take about the same fraction of expenditures—15 percent of the total. However, those dollars today can buy a much larger house packed with consumer durables that make housework less of a chore.

Over the last decade, the biggest increase in consumption spending has been for health care. Surprisingly, consumers' out-of-pocket expenses for health care take about the same share of the *household* budget as they did in the early part of the twentieth century. The major increase has come as governments pay for an ever-larger fraction of health care.

CONSUMPTION, INCOME, AND SAVING

Income, consumption, and saving are all closely linked. More precisely, **personal saving** is that part of disposable income that is not consumed; saving equals income minus consumption.

The relationship between income, consumption, and saving for the United States in 2007 is shown in Table 21-2. Begin with personal income (composed, as Chapter 20 showed, of wages, fringe benefits, interest, rents, dividends, transfer payments, and so forth). In 2007, 12.8 percent of personal income went to personal taxes. This left \$10,171 billion of **personal**

Item	Amount, 2007 (\$, billion)
Personal income	11,663
Less: Personal taxes	1,493
Equals: Disposable personal income	10,171
Less: Personal outlays (consumption and interest)	10,113
Equals: Personal saving	57.4
Memo: Personal saving as percent of disposable personal income	0.6

TABLE 21-2. Saving Equals Disposable Income Less Consumption

Source: U.S. Bureau of Economic Analysis, available at www.bea.gov.

disposable income. Household outlays for consumption (including interest) amounted to 99.4 percent of disposable income, leaving \$57 billion as personal saving. The last item in the table shows the important **personal saving rate**. This is equal to personal saving as a percent of disposable income—a tiny 0.6 percent in 2007.

Economic studies have shown that income is the primary determinant of consumption and saving. Rich people save more than poor people, both absolutely and as a percent of income. The very poor are unable to save at all. Instead, as long as they can borrow or draw down their wealth, they tend to dissave. That is, they tend to spend more than they earn, reducing their accumulated savings or going deeper into debt.

Table 21-3 contains illustrative data on disposable income, saving, and consumption drawn from budget studies on American households. The first column shows seven different levels of disposable income. Column (2) indicates saving at each level of income, and the third column indicates consumption spending at each level of income.

The *break-even point*—where the representative household neither saves nor dissaves but consumes all its income—comes at \$25,000. Below the

	(1) Disposable income (\$)	(2) Net saving (+) or dissaving (-) (\$)	(3) Consumption (\$)
A	24,000	-200	24,200
B	25,000	0	25,000
C	26,000	200	25,800
D	27,000	400	26,600
E	28,000	600	27,400
F	29,000	800	28,200
G	30,000	1,000	29,000

TABLE 21-3. Consumption and Saving Are Primarily Determined by Income

Consumption and saving rise with disposable income. The break-even point at which people have zero saving is shown here at \$25,000. How much of each extra dollar of income do people devote to extra consumption at this income level? How much to extra saving? (Answer: 80 cents and 20 cents, respectively, when we compare row B and row C.)

break-even point, say, at \$24,000, the household actually consumes more than its income; it dissaves (see the -\$200 item). Above \$25,000 it begins to show positive saving [see the +\$200 and other positive items in column (2)].

Column (3) shows the consumption spending for each income level. Since each dollar of income is divided between the part consumed and the remaining part saved, columns (3) and (2) are not independent; they must always exactly add up to column (1).

To understand the way consumption affects national output, we need to introduce some new tools. We need to understand how each dollar of additional income is divided between additional saving and additional consumption. This relationship is shown by:

- The consumption function, relating consumption and income

- Its twin, the saving function, relating saving and income

The Consumption Function

One of the most important relationships in all macroeconomics is the **consumption function**. The consumption function shows the relationship between the level of consumption expenditures and the level of disposable personal income. This concept, introduced by Keynes, is based on the hypothesis that there is a stable empirical relationship between consumption and income.

We can see the consumption function most vividly in the form of a graph. Figure 21-2 plots the seven levels of income listed in Table 21-3. Disposable income [column (1) of Table 21-3] is placed on the horizontal axis, and consumption [column (3)] is on the vertical axis. Each of the income-consumption

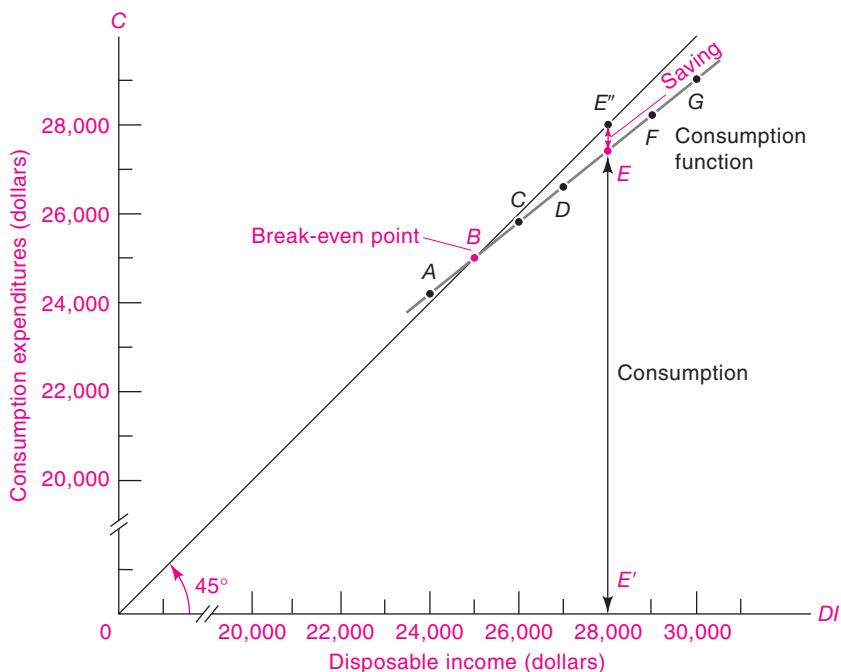


FIGURE 21-2. A Plot of the Consumption Function

The curve through A, B, C, \dots, G is the consumption function. The horizontal axis depicts the level of disposable income (DI). For each level of DI , the consumption function shows the dollar level of consumption (C) for the household. Note that consumption rises with increases in DI . The 45° line helps locate the break-even point and helps our eye measure net saving.

Source: Table 21-3.

combinations is represented by a single point, and the points are then connected by a smooth curve.

The relationship between consumption and income shown in Figure 21-2 is called the **consumption function**.

The “Break-Even” Point. To understand the figure, it is helpful to look at the 45° line drawn northeast from the origin. Because the vertical and horizontal axes have exactly the same scale, the 45° line has a very special property. At any point on the 45° line, the distance up from the horizontal axis (consumption) exactly equals the distance across from the vertical axis (disposable income). You can use your eyes or a ruler to verify this fact.

The 45° line tells us immediately whether consumption spending is equal to, greater than, or less than the level of disposable income. The point where the consumption schedule intersects the 45° line is the **break-even point**—it is the level of disposable income at which households just break even.

This break-even point is at *B* in Figure 21-2. Here, consumption expenditures exactly equal disposable income; the household is neither a borrower nor a saver. To the right of point *B*, the consumption function lies below the 45° line. The relationship between income and consumption can be seen by examining the thin blue line from *E'* to *E* in Figure 21-2. At an income of \$28,000, the level of consumption is \$27,400 (see Table 21-3). We can

see that consumption is less than income by the fact that the consumption function lies below the 45° line at point *E*.

What a household is not spending, it must be saving. The 45° line enables us to find how much the household is saving. Net saving is measured by the vertical distance from the consumption function up to the 45° line, as shown by the *EE''* saving arrow in green.

The 45° line tells us that to the left of point *B* the household is spending more than its income. The excess of consumption over income is “dissaving” and is measured by the vertical distance between the consumption function and the 45° line.

To review:

At any point on the 45° line, consumption exactly equals income and the household has zero saving. When the consumption function lies above the 45° line, the household is dissaving. When the consumption function lies below the 45° line, the household has positive saving. The amount of dissaving or saving is always measured by the vertical distance between the consumption function and the 45° line.

The Saving Function

The **saving function** shows the relationship between the level of saving and income. This is shown graphically in Figure 21-3. Again we show disposable income on the horizontal axis; but now saving, whether negative or positive in amount, is on the vertical axis.

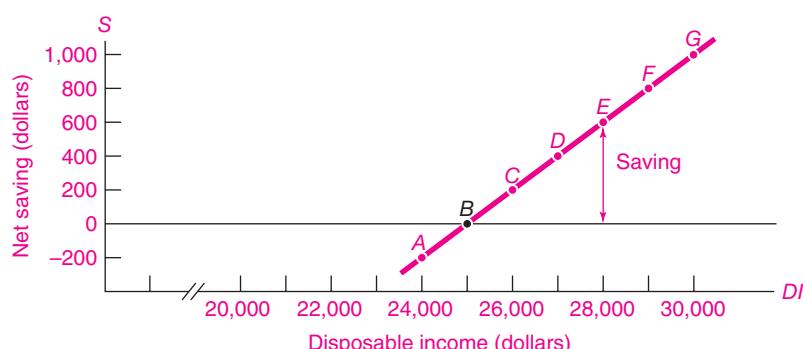


FIGURE 21-3. The Saving Function Is the Mirror Image of the Consumption Function

This saving schedule is derived by subtracting consumption from income. Graphically, the saving function is obtained by subtracting vertically the consumption function from the 45° line in Fig. 21-2. Note that the break-even point *B* is at the same \$25,000 income level as in Fig. 21-2.

This saving function comes directly from Figure 21-2. It is the vertical distance between the 45° line and the consumption function. For example, at point A in Figure 21-2, we see that the household's saving is negative because the consumption function lies above the 45° line. Figure 21-3 shows this dissaving directly—the saving function is below the zero-saving line at point A. Similarly, positive saving occurs to the right of point B because the saving function is above the zero-saving line.

The Marginal Propensity to Consume

Modern macroeconomics attaches much importance to the response of consumption to changes in income. This concept is called the marginal propensity to consume, or *MPC*.

The **marginal propensity to consume** is the extra amount that people consume when they receive an extra dollar of disposable income.

The word "marginal" is used throughout economics to mean extra or additional. For example,

"marginal cost" means the additional cost of producing an extra unit of output. "Propensity to consume" designates the desired level of consumption. The *MPC* is therefore the additional or extra consumption that results from an extra dollar of disposable income.

Table 21-4 rearranges Table 21-3's data in a more convenient form. First, verify its similarity to Table 21-3. Then, look at columns (1) and (2) to see how consumption expenditure goes up with higher levels of income.

Column (3) shows how we compute the marginal propensity to consume. From B to C, income rises by \$1000, going from \$25,000 to \$26,000. How much does consumption rise? Consumption grows from \$25,000 to \$25,800, an increase of \$800. The extra consumption is therefore 0.80 of the extra income. Out of each extra dollar of income, 80 cents goes to consumption and 20 cents goes to saving.

The example shown here is a linear consumption function—one in which the *MPC* is constant. You can verify that the *MPC* is everywhere 0.80. In reality,

	(1) Disposable income (after taxes) (\$)	(2) Consumption expenditure (\$)	(3) Marginal propensity to consume <i>MPC</i>	(4) Net saving (\$) (4) = (1) - (2)	(5) Marginal propensity to save <i>MPS</i>
A	24,000	24,200	$800/1,000 = 0.80$	-200	$200/1,000 = 0.20$
B	25,000	25,000	$800/1,000 = 0.80$	0	$200/1,000 = 0.20$
C	26,000	25,800	$800/1,000 = 0.80$	200	$200/1,000 = 0.20$
D	27,000	26,600	$800/1,000 = 0.80$	400	$200/1,000 = 0.20$
E	28,000	27,400	$800/1,000 = 0.80$	600	$200/1,000 = 0.20$
F	29,000	28,200	$800/1,000 = 0.80$	800	$200/1,000 = 0.20$
G	30,000	29,000	$800/1,000 = 0.80$	1,000	$200/1,000 = 0.20$

TABLE 21-4. The Marginal Propensities to Consume and to Save

Each dollar of disposable income not consumed is saved. Each extra dollar of disposable income goes either into extra consumption or into extra saving. Combining these facts allows us to calculate the marginal propensity to consume (*MPC*) and the marginal propensity to save (*MPS*).

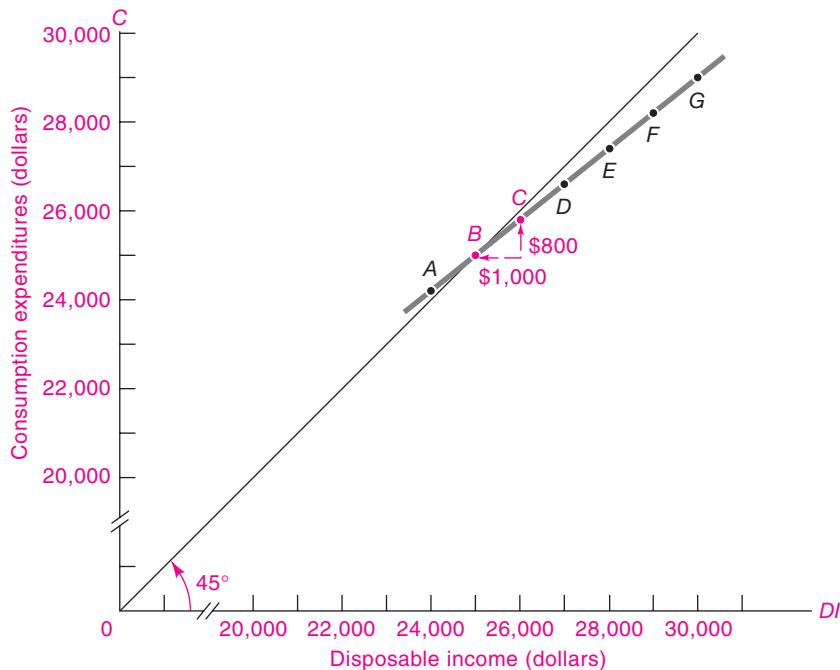


FIGURE 21-4. The Slope of the Consumption Function Is Its MPC

To calculate the marginal propensity to consume (*MPC*), we measure the slope of the consumption function by forming a right triangle and relating height to base. From point *B* to point *C*, the increase in consumption is \$800 while the change in disposable income is \$1000. The slope, equal to the change in *C* divided by the change in *DI*, gives the *MPC*. If the consumption function is everywhere upward-sloping, what does this imply about the *MPC*? If the line is a straight line, with a constant slope, what does this imply about the *MPC*?

consumption functions are unlikely to be exactly linear, but this is a reasonable approximation for our purposes.

Marginal Propensity to Consume as Geometrical Slope. We now know how to calculate the *MPC* from data on income and consumption. Figure 21-4 shows how we can calculate the *MPC* graphically. Near points *B* and *C* a little right triangle is drawn. As income increases by \$1000 from point *B* to point *C*, the amount of consumption rises by \$800. The *MPC* in this range is therefore $\$800/\$1000 = 0.80$. But, as the appendix to Chapter 1 showed, the numerical slope of a line is “the rise over the run.”² We can therefore see that the slope of the consumption

function is the same as the marginal propensity to consume.

The slope of the consumption function, which measures the change in consumption per dollar change in disposable income, is the marginal propensity to consume.

The Marginal Propensity to Save

Along with the marginal propensity to consume goes its mirror image, the marginal propensity to save, or *MPS*. The **marginal propensity to save** is defined as the fraction of an extra dollar of disposable income that goes to extra saving.

Why are *MPC* and *MPS* related like mirror images? Recall that disposable income equals consumption plus saving. This implies that each extra dollar of disposable income must be divided between

² For curved lines, we calculate the slope as the slope of the tangent line at a point.

extra consumption and extra saving. Thus if MPC is 0.80, then MPS must be 0.20. (What would MPS be if MPC were 0.6? Or 0.99?) Comparing columns (3) and (5) of Table 21-4 confirms that at any income level, MPC and MPS must always add up to exactly 1, no more and no less. $MPS + MPC = 1$, always and everywhere.

Brief Review of Definitions

Let's review briefly the main definitions we have learned:

1. The *consumption function* relates the level of consumption to the level of disposable income.
2. The *saving function* relates saving to disposable income. Because what is saved equals what is not consumed, saving and consumption schedules are mirror images.
3. The *marginal propensity to consume (MPC)* is the amount of extra consumption generated by an extra dollar of disposable income. Graphically, it is given by the slope of the consumption function.
4. The *marginal propensity to save (MPS)* is the extra saving generated by an extra dollar of disposable income. Graphically, this is the slope of the saving schedule.
5. Because the part of each dollar of disposable income that is not consumed is necessarily saved, $MPS = 1 - MPC$.

NATIONAL CONSUMPTION BEHAVIOR

Up to now we have examined the budget patterns and consumption behavior of typical families at different incomes. Let's now consider consumption for the entire nation. This transition from household behavior to national trends exemplifies the methodology of macroeconomics: We begin by examining economic activity on the individual level and then add up or aggregate the totality of individuals to study the way the overall economy operates.

Why are we interested in national consumption trends? Consumption behavior is crucial for understanding both short-term business cycles and long-term economic growth. In the short run, consumption is a major component of aggregate spending. When consumption changes sharply, the change

is likely to affect output and employment through its impact on aggregate demand. This mechanism will be described in the chapters on Keynesian macroeconomics.

Additionally, consumption behavior is crucial because what is not consumed—that is, what is saved—is available for investment in new capital goods, and capital serves as a driving force behind long-term economic growth. *Consumption and saving behavior are key to understanding economic growth and business cycles.*

Determinants of Consumption

We begin by analyzing the major forces that affect consumer spending. What factors in a nation's life and livelihood set the pace of its consumption outlays?

Disposable Income. Figure 21-5 shows how closely consumption followed current disposable income over the period 1970–2008. When DI declines in recessions, consumption usually follows the decline. Increases in DI , say, following tax cuts, stimulate consumption growth. The effects of the large cuts in personal taxes in 1981–1983 can be seen in the growth of DI and C .

Permanent Income and the Life-Cycle Model of Consumption. The simplest theory of consumption uses only the current year's income to predict consumption expenditures. Consider the following examples, which suggest why other factors might also be important:

If bad weather destroys a crop, farmers will draw upon their previous savings to finance consumption.

Similarly, law-school students borrow for consumption purposes while in school because they expect that their postgraduate incomes will be much higher than their meager student earnings.

In both these circumstances, people are in effect asking, "Given my current and future income, how much can I consume today without incurring excessive debts?"

Careful studies show that consumers generally choose their consumption levels with an eye to both current income and long-run income prospects. In order to understand how consumption depends

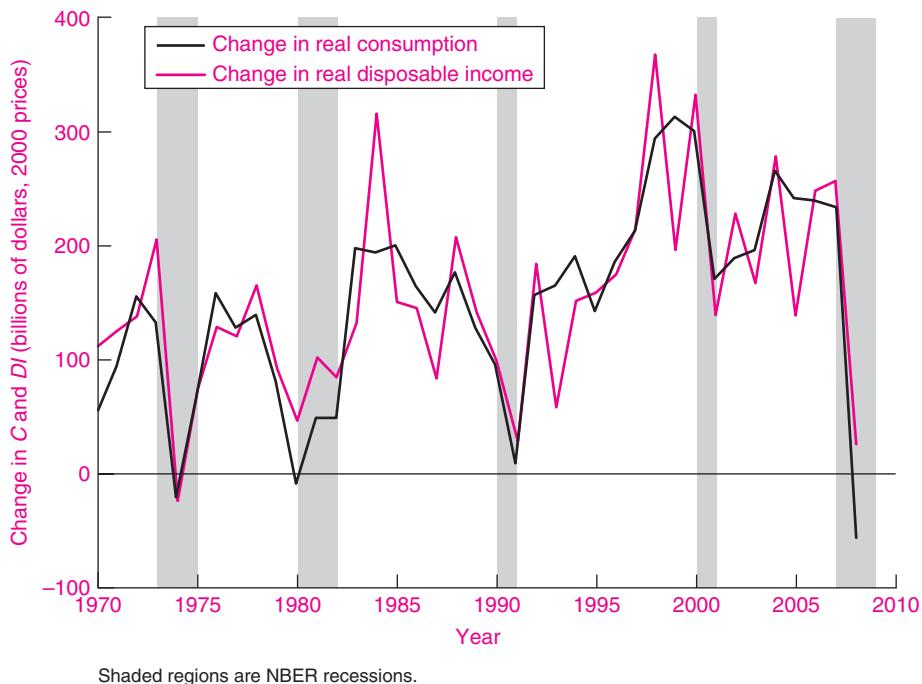


FIGURE 21-5. Changes in Consumption and Disposable Income, 1970–2008

Note how changes in consumption track changes in disposable income. Macroeconomists can forecast consumption accurately based on the historical consumption function. Recessions usually produce declines in consumption as income declines.

Source: U.S. Bureau of Economic Analysis. Real disposable income is calculated using the price index for personal consumption expenditures.

on long-term income trends, economists have developed the permanent-income theory and the life-cycle hypothesis.

Permanent income is the trend level of income—that is, income after removing temporary or transient influences due to windfall gains or losses. According to the permanent-income theory, consumption responds primarily to permanent income. This approach implies that consumers do not respond equally to all income shocks. If a change in income appears permanent (such as being promoted to a secure and high-paying job), people are likely to consume a large fraction of the increase in income. On the other hand, if the income change is clearly transitory (for example, if it arises from a one-time bonus or a good harvest), a significant fraction of the additional income may be saved.

The *life-cycle hypothesis* assumes that people save in order to smooth their consumption over their lifetime. One important objective is to have an adequate

retirement income. Hence, people tend to save while working so as to build up a nest egg for retirement and then spend out of their accumulated savings in their twilight years. One implication of the life-cycle hypothesis is that a program like social security, which provides a generous income supplement for retirement, will reduce saving by middle-aged workers since they no longer need to save as much for retirement.

Wealth and Other Influences. A further important determinant of the amount of consumption is wealth. Consider two consumers, each earning \$50,000 per year. One has \$200,000 in the bank, while the other has no savings at all. The first person may consume part of wealth, while the second has no wealth to draw down. The fact that higher wealth leads to higher consumption is called the *wealth effect*.

Wealth usually changes slowly from year to year. However, when wealth rises or declines sharply, this

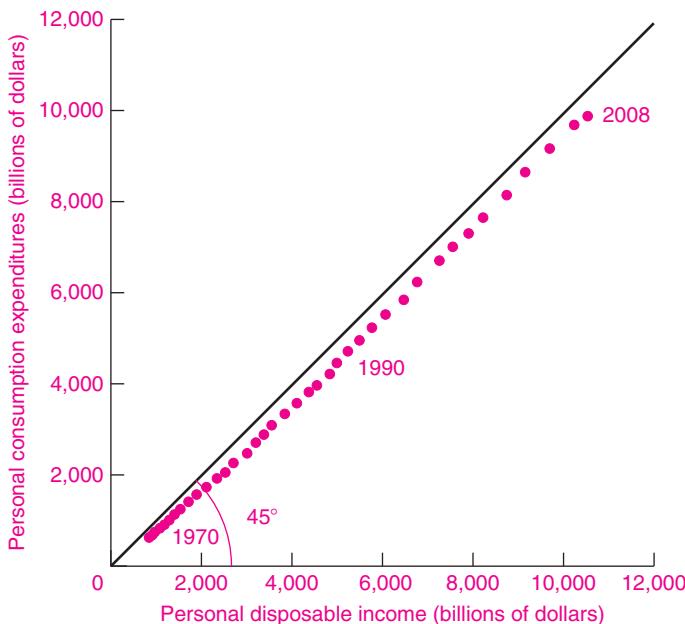


FIGURE 21-6. A Consumption Function for the United States, 1970–2008

The figure shows a scatter plot of personal disposable income and consumption. We have drawn a 45° line that shows where consumption exactly equals DI . Next, draw a consumption-function line through the points. Make sure you understand why the slope of the line you have drawn is the MPC . Can you verify that the MPC slope of the fitted line is close to 0.96?

Source: U.S. Bureau of Economic Analysis.

may lead to major changes in consumption spending. One important historical case was the stock market crash in 1929, when fortunes collapsed and paper-rich capitalists became paupers overnight. Economic historians believe that the sharp decline in wealth after the 1929 stock market crash reduced consumption spending and contributed to the depth of the Great Depression.

Over the last decade, the rise and decline of housing prices had a marked effect on consumption. From 2000 to 2006, the total value of household real estate rose over \$7000 billion (about \$70,000 per household). Many households refinanced their homes, took out home equity loans, or dipped into their savings. This is one of the reasons for the decline in the saving rate in recent years, as we will see shortly.

However, what went up then went down. By early 2009, the average price of residential houses had declined almost 30 percent from the peak in 2006. The wealth effect from declining housing values was a drag on consumer spending during this period.

The National Consumption Function

Having reviewed the theory of consumption behavior, we conclude that the determinants are complex,

including disposable income, wealth, and expectations of future income. We can plot the simplest consumption function in Figure 21-6. The scatter diagram shows data for the period 1970–2008, with each point representing the level of consumption and disposable income for a given year.

In addition, you might draw a line in Figure 21-6 through the scatter points and label it "Fitted consumption function." This fitted consumption function shows how closely consumption has followed disposable income over the period shown. In fact, economic historians have found that a close relationship between disposable income and consumption holds back to the nineteenth century.



The Declining Personal Saving Rate

Although consumption behavior tends to be stable over time, the personal saving rate dropped sharply in the United States over the last three decades. The personal saving rate as measured in the national accounts averaged around 8 percent of personal disposable income over most of the twentieth century. Starting about 1980, however, it began to decline and is now close to zero. (See Figure 21-7.)

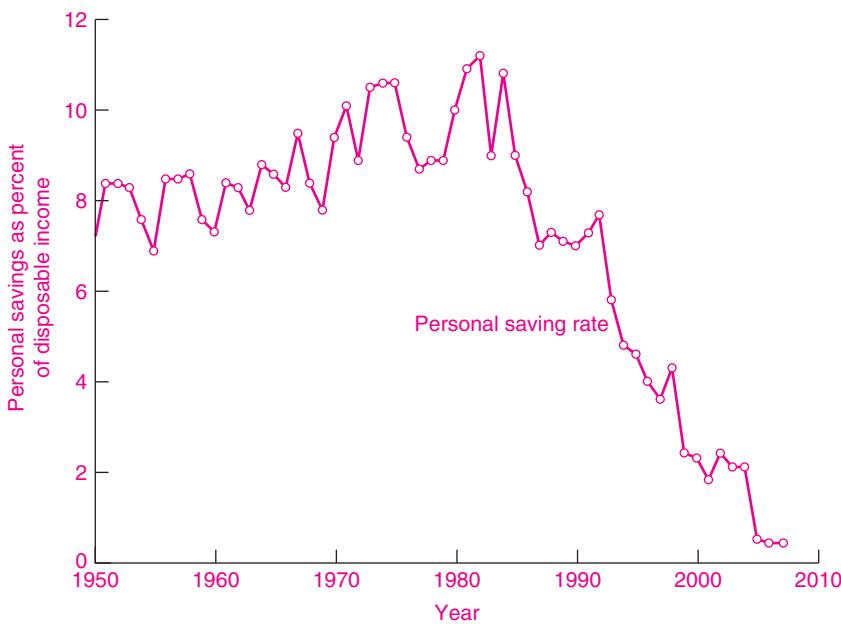


FIGURE 21-7. Personal Saving Rate Has Declined

After rising slowly over the postwar period, the personal saving rate took a sharp nosedive after 1980.

Source: U.S. Bureau of Economic Analysis.

This drop alarmed many economists because, over the long run, the growth in a nation's capital stock is largely determined by its national saving rate. National saving is composed of private and government saving. A high-saving nation has a rapidly growing capital stock and enjoys a rapid growth in its potential output. When a nation's saving rate is low, its equipment and factories become obsolete and its infrastructure begins to rot away. (This discussion abstracts away from borrowing abroad, but that cannot be a substantial fraction of income forever.)

What were the reasons for the sharp decline in the personal saving rate? This is a controversial question today, but economists point to the following potential causes:

- **Social security.** Some economists argued that the social security system has removed some of the need for private saving. In earlier times, as the life-cycle model of consumption suggests, a household would save during working years to build up a nest egg for retirement. When the government collects social security taxes and pays out social security benefits, people have less need to save for retirement. Other income-support systems have a similar effect, reducing the need to save for a rainy day. Disaster insurance for farmers, unemployment insurance for workers, and medical care for the poor and elderly all reduce the precautionary motive for people to save.

- **Financial markets.** For most of economic history, financial markets had numerous imperfections. People found it hard to borrow funds for worthwhile purposes, whether to buy a house, finance an education, or start a business. As financial markets developed, often with the help of the government, new loan instruments allowed people to borrow more easily. One example is the proliferation of credit cards, which encourage people to borrow (even though the interest rates are quite high). A generation ago, it would be difficult to borrow more than \$1000 unless a person had substantial assets. Today, credit-card solicitations arrive daily in the mail. It is not unusual to receive multiple promotions offering credit lines of \$5,000 or more in a single week!

Perhaps the biggest and ultimately most troublesome source of finance was the "subprime" mortgages that proliferated in the early 2000s. These were loans at as much as 100 percent of the value of a house, sometimes to people with no documented income. When housing prices declined, literally hundreds of billions of dollars of these loans were in default, and investors worldwide took huge losses.

- **The rapid growth in paper wealth.** Part of the decline in personal saving in the 1990–2007 period was caused by the rapid increase in personal wealth. First, the

stock market boomed, and then housing prices took off. Economists calculate that the wealth effect alone might have contributed to a decline in the personal saving rate of 3 percentage points by the late 2000s.

Alternative Measures of Saving

You might at this point ask, “If people are saving so little, why are there so many rich people?” This question raises an important point about measuring personal saving. Saving looks different to the household than to the nation as a whole. This is so because saving as measured in the national income and product accounts is not the same as that measured by accountants or in balance sheets. The *national-accounts measure of saving* is the difference between disposable income (excluding capital gains) and consumption. The *balance-sheet measure of saving* calculates the change in real net worth (that is, assets less liabilities, corrected for inflation) from one year to the next; this measure includes real capital gains.

If we examine the balance-sheet savings rate for the decade from 1997 to 2007—the viewpoint from the dining room table, so to speak—the savings rate was relatively high. Average household net worth over this period in 2007 prices rose from \$157,000 to \$191,000. The change in net worth was 17 percent of disposable income. So the balance-sheet saving rate was 17 percent, while the national-account saving rate shown in Figure 21-7 was 2 percent.

Does this alternative view mean that we can breathe a sigh of relief? Probably not. The reason is that the high saving over the last decade was largely an increase in “paper wealth.” A rise in stock prices or the prices of existing assets like housing does not necessarily reflect the productivity or “real wealth” of the economy. Although people feel richer when asset prices rise in a speculative bubble, the economy cannot produce more cars, computers, food, or housing. Indeed, if everyone wanted to sell their houses, they would find that prices would fall and they could not convert their paper wealth into consumption.

Hence, economists are justified in worrying about the decline in the national-accounts saving rate. While consumers may *feel* richer because of a booming stock or housing market, an economy is *actually* richer only when its productive tangible and intangible assets increase.

B. INVESTMENT

The second major component of private spending, after consumption, is investment. Investment plays two roles in macroeconomics. First, because it is a large and volatile component of spending, investment often leads to changes in aggregate demand and affects the business cycle. In addition, investment leads to capital accumulation. Adding to the stock of buildings and equipment increases the nation’s potential output and promotes economic growth in the long run.

Thus investment plays a dual role, affecting short-run output through its impact on aggregate demand and influencing long-run output growth through the impact of capital formation on potential output and aggregate supply.



The Meaning of “Investment” in Economics

Remember that macroeconomists use the term “investment” or “real investment” to mean additions to the stock of productive assets or capital goods like computers or trucks. When Amazon.com builds a new warehouse or when the Smiths build a new house, these activities represent investment.

Many people speak of “investing” when buying a piece of land, an old security, or any title to property. In economics, these purchases are really financial transactions or “financial investments,” because what one person is buying, someone else is selling, and the net effect is zero. There is investment only when real capital is produced.

DETERMINANTS OF INVESTMENT

In this discussion, we focus on *gross private domestic investment*, or I . This is the domestic component of national investment. Recall, however, that I is but one component of total social investment, which also includes foreign investment, government investment, and intangible investments in human capital and improved knowledge.

The major types of gross private domestic investment are the building of residential structures; investment in business fixed equipment, software, and structures; and additions to inventory. In this discussion, we focus on business investment, but the principles apply to investments by other sectors as well.

Why do businesses invest? Ultimately, businesses buy capital goods when they expect that this action will earn them a profit—that is, will bring them revenues greater than the costs of the investment. This simple statement contains the three elements essential to understanding investment: revenues, costs, and expectations.

Revenues

An investment will bring the firm additional revenue if it helps the firm sell more product. This suggests that the overall level of output (or GDP) will be an important determinant of investment. When factories are lying idle, firms have relatively little need for new factories, so investment is low. More generally, investment depends upon the revenues that will be generated by the state of overall economic activity. Most studies find that investment is very sensitive to the business cycle.

Costs

A second important determinant of the level of investment is the costs of investing. Because investment goods last many years, reckoning the costs of investment is somewhat more complicated than doing so for other commodities like coal or wheat. For durable goods, the cost of capital includes not only the price of the capital good but also the interest rate that borrowers pay to finance the capital as well as the taxes that firms pay on their incomes.

To understand this point, note that investors often raise the funds for buying capital goods by borrowing (say, through a mortgage or in the bond market). What is the cost of borrowing? It is the *interest rate* on borrowed funds. Recall that the interest rate is the price paid for borrowing money for a period of time; for example, you might have to pay 8 percent to borrow \$1000 for a year. In the case of a family buying a house, the interest rate is the mortgage interest rate.

Additionally, taxes can have a major effect on investment. One important tax is the federal corporation income tax. This tax takes up to 35 cents of the last dollar of corporate profits, thereby discouraging investment in the corporate sector. Sometimes, the government gives tax breaks to particular activities or sectors. For example, the government encourages home ownership by allowing homeowners to deduct real-estate taxes and mortgage interest from their taxable income.

Expectations

Additionally, profit expectations and business confidence are central to investment decisions. Investment is a gamble on the future. This means that business investments require a weighing of certain present costs with uncertain future profits. If businesses are concerned that political conditions in Russia are unstable, they will be reluctant to invest there. Conversely, if businesses believe that Internet commerce is the key to riches, they will invest heavily in that sector.

However, economists also realize that emotions weigh in the balance, that some investments are moved as much by intuition as by spreadsheets. This point was emphasized by J. M. Keynes as one of the reasons for the instability of a market economy:

Even apart from the instability due to speculation, there is the instability due to the characteristic of human nature that a large proportion of our positive activities depend on spontaneous optimism rather than mathematical expectations, whether moral or hedonistic or economic. Most, probably, of our decisions to do something positive, the full consequences of which will be drawn out over many days to come, can only be taken as the result of *animal spirits*—a spontaneous urge to action rather than inaction, and not as the outcome of a weighted average of quantitative benefits multiplied by quantitative probabilities.

Thus, investment decisions hang by a thread on expectations and forecasts. But accurate forecasting is difficult. Businesses spend much energy analyzing investments and trying to narrow the uncertainties about their investments.

We can sum up our review of the forces lying behind investment decisions as follows:

Businesses invest to earn profits. Because capital goods last many years, investment decisions depend on (1) the level of output produced by the new investments, (2) the interest rates and taxes that influence the costs of the investment, and (3) business expectations about the state of the economy.

THE INVESTMENT DEMAND CURVE

In analyzing the determinants of investment, we focus particularly on the relationship between interest rates and investment. This linkage is crucial because interest rates (influenced by central banks) are the

major instrument by which governments influence investment. To show the relationship between interest rates and investment, economists use a schedule called the **investment demand curve**.

Consider a simplified economy where firms can invest in different projects: A, B, C, and so forth, up to H. These investments are so durable (like power plants or buildings) that we can ignore the need for replacement. Further, they yield a constant stream of net income each year, and there is no inflation. Table 21-5 shows the financial data on each of the investment projects.

Consider project A. This project costs \$1 million. It has a very high return—\$1500 per year of revenues per \$1000 invested (this is a rate of return of 150 percent per year). Columns (4) and (5) show the cost of investment. For simplicity, assume that the investment is financed purely by borrowing at the market interest rate, here taken alternately as 10 percent per year in column (4) and 5 percent in column (5).

Thus at a 10 percent annual interest rate, the cost of borrowing \$1000 is \$100 a year, as is shown in all entries of column (4); at a 5 percent interest rate, the borrowing cost is \$50 per \$1000 borrowed per year.

Finally, the last two columns show the *annual net profit* from each investment. For lucrative project A, the net annual profit is \$1400 a year per \$1000 invested at a 10 percent interest rate. Project H loses money.

To review our findings: In choosing among investment projects, firms compare the annual revenues from an investment with the annual cost of capital, which depends upon the interest rate. The difference between annual revenue and annual cost is the annual net profit. When annual net profit is positive, the investment makes money, while a negative net profit denotes that the investment loses money.

Look again at Table 21-5 and examine the last column, showing annual net profit at a 5 percent interest rate. Note that at this interest rate, investment

Project	(1) Total investment in project (\$, million)	(2) Annual revenues per \$1,000 invested	(3)		(5) Cost per \$1,000 Borrowed at Annual Interest Rate of:	(6) Annual Net Profit per \$1,000 Borrowed at Annual Interest Rate of:		
			(4)			(6) = (3) - (4)		
			10% (\$)	5% (\$)		10% (\$)	5% (\$)	
A	1	1,500	100	50		1,400	1,450	
B	4	220	100	50		120	170	
C	10	160	100	50		60	110	
D	10	130	100	50		30	80	
E	5	110	100	50		10	60	
F	15	90	100	50		-10	40	
G	10	60	100	50		-40	10	
H	20	40	100	50		-60	-10	

TABLE 21-5. The Profitability of Investment Depends on the Interest Rate

The economy has eight investment projects, ranked in order of return. Column (2) shows the investment in each project. Column (3) calculates the perpetual return each year per \$1000 invested. Columns (4) and (5) then show the cost of the project, assuming all funds are borrowed, at interest rates of 10 and 5 percent; this is shown per \$1000 borrowed.

The last two columns calculate the annual net profit per \$1000 invested in the project. If net profit is positive, profit-maximizing firms will undertake the investment; if negative, the investment project will be rejected.

Note how the cutoff between profitable and unprofitable investments moves as the interest rate rises. (Where would the cutoff be if the interest rate rose to 15 percent per year?)

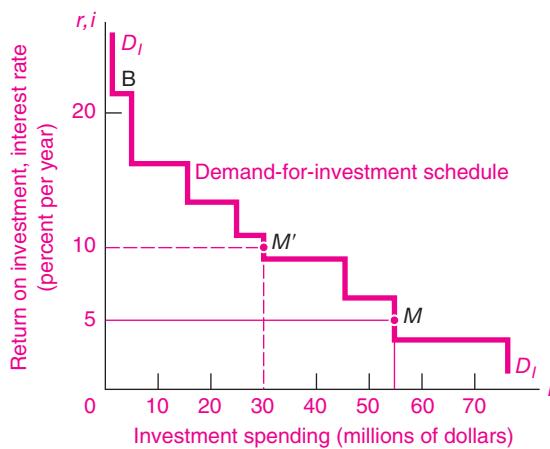


FIGURE 21-8. Investment Depends upon Interest Rate

The downward-stepping demand-for-investment schedule plots the amount that businesses would invest at each interest rate, as calculated from the data in Table 21-5. Each step represents a lump of investment: project A has such a high rate of return that it is off the figure; the highest visible step is project B, shown at the upper left. At each interest rate, all investments that have positive net profit will be undertaken.

projects A through G would be profitable. We would thus expect profit-maximizing firms to invest in all seven projects, which [from column (2)] total up to \$55 million in investment. Thus at a 5 percent interest rate, investment demand would be \$55 million.

However, suppose that the interest rate rises to 10 percent. Then the cost of financing these investments would double. We see from column (6) that investment projects F and G become unprofitable at an interest rate of 10 percent; investment demand would fall to \$30 million.

We show the results of this analysis in Figure 21-8. This figure shows the *demand-for-investment schedule*, which is here a downward-sloping step function of the interest rate. This schedule shows the amount of investment that would be undertaken at each interest rate; it is obtained by adding up all the investments that would be profitable at each level of the interest rate.

Hence, if the market interest rate is 5 percent, the desired level of investment will occur at point M, which shows investment of \$55 million. At this interest rate, projects A through G are undertaken. If interest rates were to rise to 10 percent, projects F

and G would be squeezed out; in this situation, investment demand would lie at point M' with total investment of \$30 million.³

Shifts in the Investment Demand Curve

We have seen how interest rates affect the level of investment. Investment is affected by other forces as well. For example, an increase in the GDP will shift the investment demand curve out, as shown in Figure 21-9(a) on the next page.

An increase in business taxation would depress investment. Say that the government taxes away half the net yield in column (3) of Table 21-5, with interest costs in columns (4) and (5) not being deductible. The net profits in columns (6) and (7) would therefore decline. [Verify that at a 10 percent interest rate, a 50 percent tax on column (3) would raise the cutoff to between projects B and C, and the demand for investment would decline to \$5 million.] The case of a tax increase on investment income is shown in Figure 21-9(b).

We can also see how expectations enter the picture from a historical example. In the late 1990s, investors became infatuated with the Internet and the “new economy.” They poured money into now-defunct companies on the basis of wild projections. Some seasoned investors even succumbed to the “animal spirits,” as, for example, when Time Warner paid \$180 billion for the online company AOL. Figure 21-9(c) illustrates how a bout of business optimism would shift out the investment demand schedule in the 1990s. When the technology-stock bubble burst in 2000, the demand for investment in software and equipment fell sharply as well, and the curve in Figure 21-9(c) shifted sharply back to the left. These are but two examples of how expectations can have powerful effects on investment.

After learning about the factors affecting investment, you will not be surprised to discover that investment is the most volatile component of spending. Investment behaves unpredictably because it depends on such uncertain factors as the success or failure of new and untried products, changes in tax rates and interest rates, political attitudes and

³ We will later see that when prices are changing, it is appropriate to use a real interest rate, which represents the nominal or money interest rate corrected for inflation.

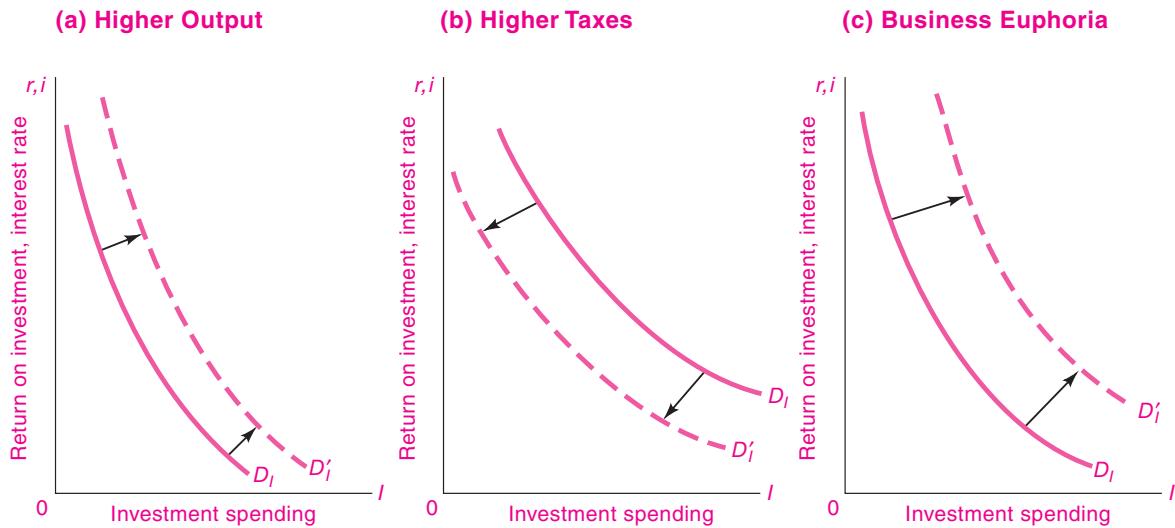


FIGURE 21-9. Shifts in Investment Demand Function

In the demand-for-investment (D_I) schedule, the arrows show the impact of (a) a higher level of GDP, (b) higher taxes on capital income, and (c) a burst of business euphoria.

approaches to stabilizing the economy, and similar changeable events of economic life. *In virtually every business cycle, investment fluctuations have been the driving force behind boom or bust.*

ON TO THE THEORY OF AGGREGATE DEMAND

We have now completed our introduction to the basic concepts of macroeconomics. We have examined the determinants of consumption and investment and seen how they can fluctuate from year to year, sometimes quite sharply.

At this point, macroeconomics branches into two major subjects—business cycles and economic

growth. We begin our survey in the next chapter with business cycles, which concern the short-term fluctuations in output, employment, and prices. Modern business-cycle theories rely primarily on the Keynesian approach. This analysis shows the impact of financial shocks and changes in investment, government spending and taxation, and foreign trade. These shocks are amplified through induced consumption effects and determine aggregate demand. We will learn that the wise application of government fiscal and monetary policies can reduce the severity of recessions and inflation, but also that poor policies can amplify shocks. The theories of consumption and investment surveyed in this chapter will be the major players in our business-cycle drama.



SUMMARY

A. Consumption and Saving

1. Disposable income is an important determinant of consumption and saving. The consumption function is the schedule relating total consumption to total

disposable income. Because each dollar of disposable income is either saved or consumed, the saving function is the other side or mirror image of the consumption function.

2. Recall the major features of consumption and saving functions:
 - a. The consumption (or saving) function relates the level of consumption (or saving) to the level of disposable income.
 - b. The marginal propensity to consume (*MPC*) is the amount of extra consumption generated by an extra dollar of disposable income.
 - c. The marginal propensity to save (*MPS*) is the extra saving generated by an extra dollar of disposable income.
 - d. Graphically, the *MPC* and the *MPS* are the slopes of the consumption and saving schedules, respectively.
 - e. $MPS \equiv 1 - MPC$.
3. Adding together individual consumption functions gives us the national consumption function. In simplest form, it shows total consumption expenditures as a function of disposable income. Other variables, such as permanent income or long-term income trends as well as wealth, also have a significant impact on consumption patterns.
4. The personal saving rate has declined sharply in the last three decades. To explain this decline, economists point to social security and government health programs, changes in financial markets, and wealth effects. Declining saving hurts the economy because

personal saving is a major component of national saving and investment. While people feel richer because of the booming stock market, the nation's true wealth increases only when its productive tangible and intangible assets increase.

B. Investment

5. The second major component of spending is gross private domestic investment in housing, plant, software, and equipment. Firms invest to earn profits. The major economic forces that determine investment are therefore the revenues produced by investment (primarily influenced by the state of the business cycle), the cost of investment (determined by interest rates and tax policy), and the state of expectations about the future. Because it depends on highly unpredictable future events, investment is the most volatile component of aggregate spending.
6. An important relationship is the investment demand schedule, which connects the level of investment spending to the interest rate. Because the profitability of investment varies inversely with the interest rate, which affects the cost of capital, we can derive a downward-sloping investment demand curve. As the interest rate declines, more investment projects become profitable.

CONCEPTS FOR REVIEW

Consumption and Saving

disposable income, consumption, saving
consumption and saving functions
personal saving rates
marginal propensity to consume (*MPC*)
marginal propensity to save (*MPS*)

$MPC + MPS \equiv 1$
break-even point
 45° line
determinants of consumption:
current disposable income
permanent income
wealth
life-cycle effect

Investment

determinants of investment:
revenues
costs
expectations
role of interest rates in *I*
investment demand function
animal spirits

FURTHER READING AND INTERNET WEBSITES

Further Reading

Economists have studied consumer expenditure patterns in order to improve predictions and aid economic policy. One of the most influential studies is Milton Friedman, *The Theory of the Consumption Function* (University of Chicago Press, 1957). A historical overview by an economic historian

is Stanley Lebergott, *Pursuing Happiness: American Consumers in the Twentieth Century* (Princeton University Press, Princeton, N.J., 1993).

Firms devote much management time to deciding about investment strategies. A good survey can be found in Richard A. Brealey, Stewart C. Myers, and Franklin Allen,

Principles of Corporate Finance (McGraw-Hill, New York, 2009).

Websites

Data on total personal consumption expenditures for the United States are provided at the website of the Bureau of Economic Analysis, www.bea.gov.

Data on family budgets are contained in Bureau of Labor Statistics, "Consumer Expenditures," available at www.bls.gov.

Data and analysis of investment for the U.S. economy are provided by the Bureau of Economic Analysis at www.bea.gov.

Milton Friedman and Franco Modigliani made major contributions to our understanding of the consumption function. Visit the Nobel website at nobelprize.org/nobel_prizes/economics to read about the importance of their contributions to macroeconomics.

QUESTIONS FOR DISCUSSION

1. Summarize the budget patterns for food, clothing, luxuries and saving.
2. In working with the consumption function and the investment demand schedule, we need to distinguish between shifts of and movements along these schedules.
 - a. Define carefully for both curves changes that would lead to shifts of and those that would produce movements along the schedules.
 - b. For the following, explain verbally and show in a diagram whether they are shifts of or movements along the consumption function: increase in disposable income, decrease in wealth, fall in stock prices.
 - c. For the following, explain in words and show in a diagram whether they are shifts of or movements along the investment demand curve: expectation of a decline in output next year, rise of interest rates, increase in taxes on profits.
3. Exactly how were the *MPC* and *MPS* in Table 21-4 computed? Illustrate by calculating *MPC* and *MPS* between points *A* and *B*. Explain why it must always be true that $MPC + MPS \equiv 1$.
4. I consume all my income at every level of income. Draw my consumption and saving functions. What are my *MPC* and *MPS*?
5. Estimate your income, consumption, and saving for last year. If you dissaved (consumed more than your income), how did you finance your dissaving? Estimate the composition of your consumption in terms of each of the major categories listed in Table 21-1.
6. "Along the consumption function, income changes more than consumption." What does this imply for the *MPC* and *MPS*?
7. "Changes in disposable income lead to movements along the consumption function; changes in wealth or other factors lead to a shift of the consumption function." Explain this statement with an illustration of each case.
8. What would be the effects of the following on the investment demand function illustrated in Table 21-5 and Figure 21-8?
 - a. A doubling of the annual revenues per \$1000 invested shown in column (3)
 - b. A rise in interest rates to 15 percent per year
 - c. The addition of a ninth project with data in the first three columns of (J, 10, 70)
 - d. A 50 percent tax on net profits shown in columns (6) and (7)
9. Using the augmented investment demand schedule from question 8(c) and assuming that the interest rate is 10 percent, calculate the level of investment for cases **a** through **d** in question 8.
10. **Advanced problem:** According to the life-cycle model, people consume each year an amount that depends upon their *lifetime* income rather than upon their current income. Assume that you expect to receive future income (in constant dollars) according to the schedule in Table 21-6.
 - a. Assume that there is no interest paid on savings. You have no initial savings. Further assume that you want to "smooth" your consumption (enjoying equal consumption each year) because of diminishing extra satisfaction from extra consumption. Derive your best consumption trajectory for the 5 years, and write the figures in column (3). Then calculate your saving and enter the amounts in column (4); put your end-of-period wealth, or cumulative saving, for each year into column (5). What is your average saving rate in the first 4 years?
 - b. Next, assume that a government social security program taxes you \$2000 in each of your working

(1) Year	(2) Income (\$)	(3) Consumption (\$)	(4) Saving (\$)	(5) Cumulative saving (end of year) (\$)
1	30,000	_____	_____	_____
2	30,000	_____	_____	_____
3	25,000	_____	_____	_____
4	15,000	_____	_____	_____
5*	0	_____	_____	0

*Retired.

TABLE 21-6.

years and provides you with an \$8000 pension in year 5. If you still desire to smooth consumption, calculate your revised saving plan. How has the social security program affected your

consumption? What is the effect on your average saving rate in the first 4 years? Can you see why some economists claim that social security can lower saving?

22

Business Cycles and Aggregate Demand



The fault, dear Brutus, is not in our stars—but in ourselves.

William Shakespeare
Julius Caesar

The American economy has been subject to business cycles since the early days of the Republic. Sometimes, business conditions are healthy, with rapidly growing employment, factories working overtime, and robust profits. The “fabulous 1990s” was such a period for the American economy. The economy grew rapidly; employment and capacity utilization were exceptionally high, and unemployment was low. Yet, unlike the case in earlier long expansions, inflation remained low throughout the 1990s.

Such periods of prosperity often come to an unhappy end. In the nineteenth and early twentieth centuries, and again in 2007–2009, financial crises turned into waves of contagious pessimism, businesses failed, credit conditions tightened, and a downturn in the banking and financial sectors rippled through the rest of the economy. During business downturns, jobs are hard to find, factories are idle, and profits are low. These downturns are usually short and mild, as was the case in the recession that began in March 2001 and ended in November 2001. From time to time the contraction may persist for a decade and cause widespread economic hardships, as during the 1930s in the Great Depression of the 1930s or in Japan in the 1990s.

These short-term fluctuations in economic activity, known as *business cycles*, are the central topic of this chapter. Understanding business cycles

has proved to be one of the most enduring issues in all of macroeconomics. What causes business fluctuations? How can government policies reduce their virulence? Economists were largely unable to answer these questions until the 1930s, when the revolutionary macroeconomic theories of John Maynard Keynes highlighted the importance of the forces of aggregate demand in determining business cycles. Keynesian economics emphasizes that *changes in aggregate demand can have powerful impacts on the overall levels of output, employment, and prices in the short run.*

This chapter describes the basic features of the business cycle and presents the simplest theories of output determination. The structure of this chapter is as follows:

- We begin with a description of the key elements of the business cycle.
- We then summarize the basics of aggregate demand and show how the modern business cycle fits into that framework.
- Next, we develop the multiplier model—the simplest Keynesian example of a model of aggregate demand.
- We close with an application of the multiplier model to the question of the impact of fiscal policy on output.

A. WHAT ARE BUSINESS CYCLES?

Economic history shows that no economy grows in a smooth and even pattern. A country may enjoy several years of economic expansion and prosperity, with rapid increases in stock prices (as in the 1990s) or housing prices (as in the early 2000s). Then, the irrational exuberance may flip over to irrational pessimism as, during the 2007–2009 period, lenders stop issuing mortgages or car loans on favorable terms, banks slow their lending to businesses, and spending declines. Consequently, national output falls, unemployment rises, and profits and real incomes decline.

Eventually the bottom is reached and recovery begins. The recovery may be incomplete, or it may be so strong as to lead to a new boom. Prosperity may mean a long, sustained period of brisk demand, plentiful jobs, and rising living standards. Or it may be marked by a quick, inflationary flare-up in prices and speculation, followed by another slump.

Upward and downward movements in output, inflation, interest rates, and employment form the business cycle that characterizes all market economies.

FEATURES OF THE BUSINESS CYCLE

What exactly do we mean by “business cycles”?

Business cycles are economywide fluctuations in total national output, income, and employment, usually lasting for a period of 2 to 10 years, marked by widespread expansion or contraction in most sectors of the economy.

Economists typically divide business cycles into two main phases: *recession* and *expansion*. Peaks and troughs mark the turning points of the cycle. Figure 22-1 shows the successive phases of the business cycle. The downturn of a business cycle is called a recession. A **recession** is a recurring period of decline in total output, income, and employment, usually lasting from 6 to 12 months and marked by contractions in many sectors of the economy. A recession that is large in both scale and duration is called a **depression**.

The semiofficial judge of the timing of contractions and expansions is the National Bureau of Economic Research (NBER), a private research organization. The NBER defines a recession as “a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment,

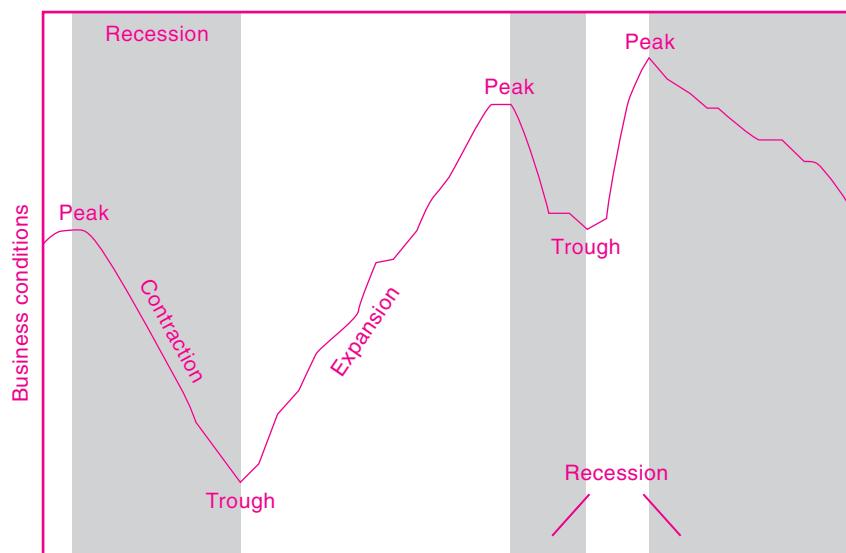


FIGURE 22-1. A Business Cycle, like the Year, Has Its Seasons

Business cycles are the irregular expansions and contractions in economic activity. (These are the actual monthly data on industrial production for a recent business-cycle period.)

industrial production, and wholesale-retail sales." (See "Websites" at the end of this chapter for further information on dating of recessions.)

An alternative definition sometimes used is that a recession occurs when real GDP has declined for two consecutive calendar quarters. (Question 12 at the end of the chapter reviews the difference between the two definitions.)

Although we call these short-term fluctuations "cycles," the actual pattern is irregular. No two business cycles are quite the same. No exact formula, such as might apply to the revolutions of the planets or the swings of a pendulum, can be used to predict the duration and timing of business cycles. Rather, business cycles more closely resemble the irregular fluctuations of the weather. Figure 22-2 shows the American business cycles throughout recent history. Here you can see that business cycles are like mountain ranges, with some valleys that are deep and

broad, as in the Great Depression, and others that are shallow and narrow, as in the recession of 1991.

While individual business cycles are not identical, they often share a family similarity. If a reliable economic forecaster announces that a recession is about to arrive, what are the typical phenomena that you should expect? The following are a few of the *customary characteristics* of a recession:

- Investment usually falls sharply in recessions. Housing has generally been the first to decline, either because of a financial crisis or because the Federal Reserve has raised interest rates to slow inflation. Consumer purchases often decline sharply as well. As businesses slow production lines, real GDP falls.
- Employment usually falls sharply in the early stages of a recession. It sometimes recovers slowly in what are often called "jobless recoveries."

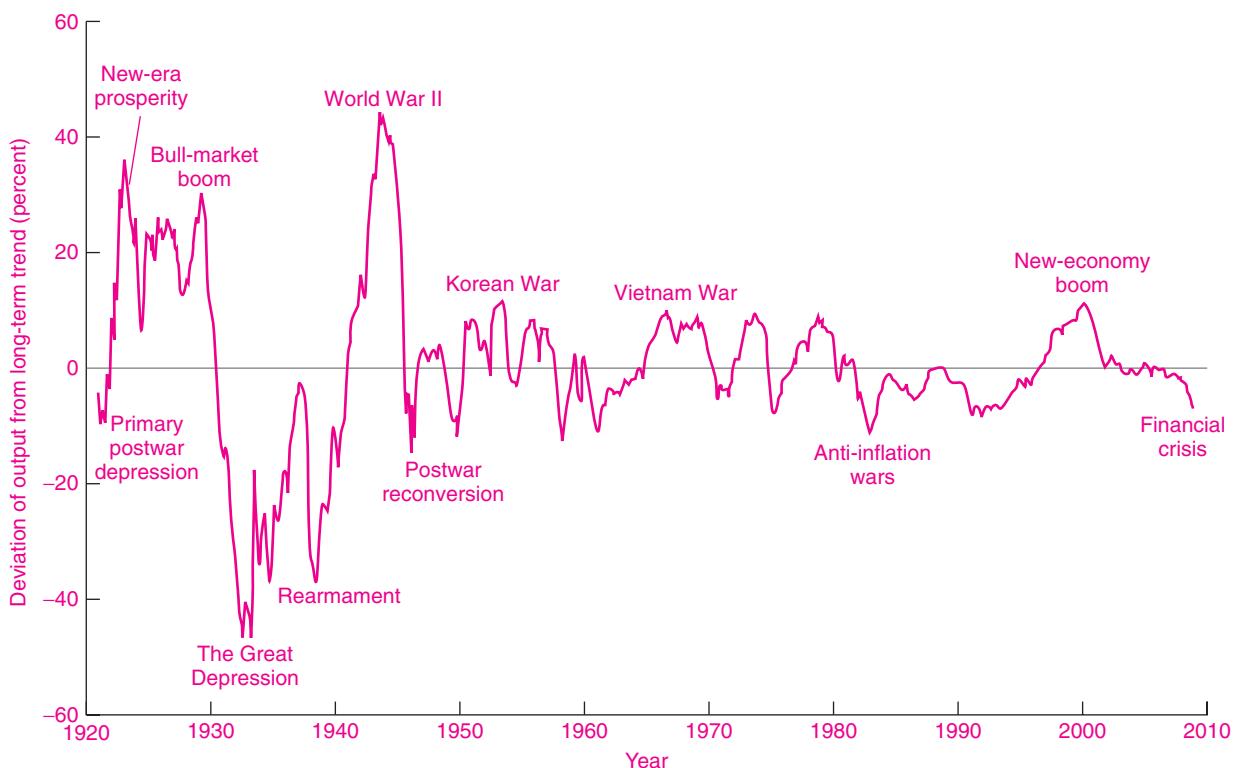


FIGURE 22-2. Business Activity since 1919

Industrial production has fluctuated irregularly around its long-run trend. Can you detect a more stable economy in recent years?

Source: Federal Reserve Board, detrended by authors.

- As output falls, inflation slows and the demand for crude materials declines, and materials' prices tumble. Wages and the prices of services are unlikely to face a similar decline, but they tend to rise less rapidly in economic downturns.
- Business profits fall sharply in recessions. In anticipation of this, common-stock prices usually fall as investors sniff the scent of a business downturn.
- Generally, as business conditions deteriorate and employment falls, the Federal Reserve begins to lower short-term interest rates to stimulate investment, and other interest rates decline as well.

BUSINESS-CYCLE THEORIES

Exogenous vs. Internal Cycles. Over the years, macroeconomists have engaged in vigorous debates about the reasons for business fluctuations. Some think they are caused by monetary fluctuations, others by productivity shocks, and still others by changes in exogenous spending.

There is certainly no end to possible explanations, but it is useful to classify the different theories into two categories: exogenous and internal. The *exogenous* theories find the sources of the business cycle in the fluctuations of factors outside the economic system—in wars, revolutions, and elections; in oil prices, gold discoveries, and population migrations; in discoveries of new lands and resources; in scientific breakthroughs and technological innovations; even in sunspots, climate change, and the weather.

An example of an exogenous cycle was the outbreak of World War II. When Germany and Japan launched wars on Europe and the United States, this led to a rapid military buildup, large increases in spending, and an increase in aggregate demand that propelled the United States out of the Great Depression. Here we saw an exogenous event—a major war—that led to a huge increase in military spending and to the biggest economic expansion of the twentieth century. (We will examine this episode later in this chapter.)

By contrast, the *internal* theories look for mechanisms within the economic system itself. In this approach, every expansion breeds recession and contraction, and every contraction breeds revival and expansion. Many business cycles in U.S. economic history were internal cycles that originated in the financial sector. It is for this reason that we devote much of our attention to monetary and financial economics.

Financial Crises and Business Cycles

One common feature of capitalism around the world is the speculative booms and busts that occurred frequently in the nineteenth century, produced the upheaval of the Great Depression, and reappeared in the United States several times over the last two decades. Below are some important examples.

Panics of Early Capitalism. The nineteenth century witnessed frenzies of investment speculation—notably in canals, land, and railroads. Inevitably, “animal spirits” would take over. Railroads would be overbuilt, land prices would rise too high, and people would take on too much debt. Bankruptcy would lead to bank failures, a run on the banks, and a banking crisis. Output and prices would fall sharply in the panic. Eventually, after the worst excesses were wrung out, the economy would begin to expand again.

Hyperinflation. Sometimes, an overheated economy leads to high inflation, or even hyperinflation. Hyperinflation occurs when prices rise at 100 percent or more *per month*. The most famous hyperinflation in history occurred in Germany in 1923. The government was unable to meet its financial obligations through taxing and borrowing, so it turned to the monetary printing press. By the end of 1923, currency was printed with more and more digits, and the largest banknote in circulation was for 25 billion marks! Central banks today are vigilant in their defense against even the most moderate inflation.

The New-Economy Bubble. The classic pattern of speculative boom was seen again in the late 1990s. The phenomenal pattern of growth and innovation in the “new-economy” sectors—including software, the Internet, and the newly invented dot.com companies—produced a speculative boom in new-economy stocks. Companies sold online dating services, gave away free electronic birthday cards, and issued stock for Flooz.com, which sold a worthless digital currency. College students dropped out of school to become instant millionaires (or so they dreamed). All of this spurred real investment in computers, software, and telecommunications. Investment in information-processing equipment rose by 70 percent from 1995 to 2000, representing one-fifth of the entire rise in real GDP during this period.

Eventually, investors became skeptical about the fundamental value of many of these firms. Losses piled up on top of losses. The urge to buy the stocks before prices rose higher was replaced by the panicky desire to sell before they collapsed. The stock price of a typical new-economy company fell from \$100 per share to pennies by 2003. Many such companies went bankrupt. College dropouts went back to school wiser but seldom richer.

The changed expectations about the new economy and the resulting stock market decline contributed to the recession and slow growth in the 2000–2002 period. Investment in information-processing equipment fell by 10 percent, and investment in computers fell by twice as much. The impressive innovations of the new economy have become a staple feature of modern technology, but, with a few exceptions, investors have little or no profits to show for their efforts.

The Housing Bubble. Less than a decade later, another financial crisis erupted, and this was again the result of rapid innovation. But in this case, the innovation was the process of financial “securitization.” This occurs when a financial instrument, such as a simple home mortgage, is sliced and diced, repackaged, and then sold on securities markets. While securitization itself was not a new phenomenon, the scope of packaging and repackaging grew sharply. Rating agencies failed to provide accurate ratings of the riskiness of these new securities, and many people bought them thinking they were as good as gold. The worst examples were “subprime mortgages,” mortgages provided to people for the entire value of a house on the basis of little or no documentation of their income and job status. By early 2007, the total value of these new securities was over \$1 trillion.

All went well as long as housing prices were rising, as they did starting in 1995. But then in 2006 the housing bubble burst—echoing the end of the speculative dot.com stock-market bubble from a decade earlier. Many of the new securities lost value. It turned out they were not top-grade AAA securities but junk bonds. As banks and other financial institutions suffered large losses, they began to tighten credit, reduce loans, and cut back sharply on new mortgages. Risk premiums rose sharply.

The Federal Reserve took steps to ease monetary conditions—lowering interest rates and extending

credit—but it was flying against powerful headwinds. As the value of stocks fell more sharply than at any time in a century, many financial institutions were on the verge of bankruptcy. Many of the large investment banking firms disappeared. The Federal Reserve and U.S. Treasury loaned massive amounts of federal money and bailed out several financial firms. Yet, even with the strong countercyclical activities, the economy went into a deep recession at the end of 2007.

You begin to see the theme running through all these events. The next few chapters survey our economic theories to explain them.

B. AGGREGATE DEMAND AND BUSINESS CYCLES

We have now begun to understand the short-term changes in output, employment, and prices that characterize business fluctuations in market economies. Most explanations of business cycles rely upon the theory of aggregate demand. This section explains *AD* theory in greater detail.

THE THEORY OF AGGREGATE DEMAND

What are the major components of aggregate demand? How do they interact with aggregate supply to determine output and prices? Exactly how do short-run fluctuations in *AD* affect GDP? We first examine aggregate demand in more detail in order to get a better understanding of the forces driving the economy. Then, in the following sections, we derive the simplest model of aggregate demand: the multiplier model.

Aggregate demand (or *AD*) is the total or aggregate quantity of output that is willingly bought at a given level of prices, other things held constant. *AD* is the desired spending in all product sectors: consumption, private domestic investment, government purchases of goods and services, and net exports. It has four components:

1. *Consumption.* As we saw in the last chapter, consumption (*C*) is primarily determined by

disposable income, which is personal income less taxes. Other factors affecting consumption are longer-term trends in income, household wealth, and the aggregate price level. Aggregate demand analysis focuses on the determinants of *real* consumption (that is, nominal or dollar consumption divided by the price index for consumption).

2. *Investment.* Investment (I) spending includes purchases of buildings, software, and equipment and accumulation of inventories. Our analysis in Chapter 21 showed that the major determinants of investment are the level of output, the cost of capital (as determined by tax policies along with interest rates and other financial conditions), and expectations about the future. The major channel by which economic policy can affect investment is monetary policy.
3. *Government purchases.* A third component of aggregate demand is government purchases of goods and services (G). This includes the purchases of goods like tanks and school books, as well as the services of judges and public-school teachers. Unlike private consumption and investment, this component of aggregate demand is determined directly by the government's spending decisions; when the Pentagon buys a new fighter aircraft, this output directly adds to the GDP.
4. *Net exports.* A final component of aggregate demand is net exports (X), which equal the value of exports minus the value of imports. Imports are determined by domestic income and output, by the ratio of domestic to foreign prices, and by the foreign exchange rate of the dollar. Exports (which are imports of other countries) are the mirror image of imports, and they are determined by foreign incomes and outputs, by relative prices, and by foreign exchange rates. Net exports, then, will be determined by domestic and foreign outputs, relative prices, and exchange rates.

Figure 22-3 shows the AD curve and its four major components. At price level P , we can read the levels of consumption, investment, government purchases, and net exports, which sum to GDP, or Q . The sum of the four spending streams at that price level is aggregate spending, or aggregate demand, at that price level.

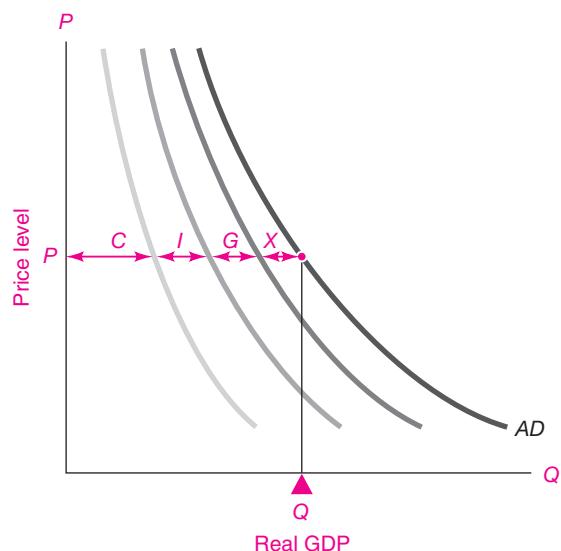


FIGURE 22-3. Components of Aggregate Demand

Aggregate demand (AD) consists of four components—consumption (C), domestic private investment (I), government spending on goods and services (G), and net exports (X).

Aggregate demand shifts when there are changes in macroeconomic policies (such as monetary-policy changes or changes in government expenditures or tax rates) or when exogenous events change spending (as would be the case with changes in foreign output, affecting X , or in business confidence, affecting I).

THE DOWNWARD-SLOPING AGGREGATE DEMAND CURVE

One important point you should notice is that the aggregate demand curve in Figure 22-3 slopes downward. This means that, holding other things constant, the level of real spending declines as the overall price level in the economy rises.

What is the reason for the downward slope? The basic reason is that there are some elements of income or wealth that do not rise when the price level rises. For example, some items of personal income might be set in nominal dollar terms—some government transfer payments, the minimum wage, and company pensions are examples. When the price level goes up, therefore, real disposable income falls, leading to a decline in real consumption expenditures.

In addition, some elements of wealth may be fixed in nominal terms. Examples here would be holdings of money and bonds, which usually contain promises to pay a certain number of dollars in a given period. If the price level rises, therefore, the real value of wealth declines, and this would again lead to lower levels of real consumption.

We illustrate the impact of a higher price level graphically in Figure 22-4(a) on page 436. Say that the economy is in equilibrium at point *B*, with a price level of 100 and a real GDP of \$3000 billion. Next assume that prices rise by 50 percent, so the price index *P* rises from 100 to 150. Suppose that at that higher price level, real spending declines because of lower real disposable income. Total real spending declines to \$2000 billion, shown at point *C*. We see here how higher prices have reduced real spending.

To summarize:

The *AD* curve slopes downward. This downward slope implies that real spending declines as the price level rises, other things held constant. Real spending declines with a higher price level primarily because of the effect of higher prices on real incomes and real wealth.

Shifts in Aggregate Demand

We have seen that total spending in the economy tends to decline as the price level rises, holding other things constant. But those other things do in fact tend to change, thereby producing changes in aggregate demand. What are the key determinants of changes in aggregate demand?

We can separate the determinants of *AD* into two categories, as shown in Table 22-1. One set includes

Variable	Impact on aggregate demand
Policy Variables	
Monetary policy	Monetary expansion may lower interest rates and loosen credit conditions, inducing higher levels of investment and consumption of durable goods. In an open economy, monetary policy also affects the exchange rate and net exports.
Fiscal policy	Increases in government purchases of goods and services directly increase spending; tax reductions or increases in transfers raise disposable income and induce higher consumption. Tax incentives like an investment tax credit can induce higher spending in a particular sector.
Exogenous Variables	
Foreign output	Output growth abroad leads to an increase in net exports.
Asset values	Rise in stock market increases household wealth and thereby increases consumption; also, higher stock prices lower the cost of capital and thereby increase business investment.
Advances in technology	Technological advances can open up new opportunities for business investment. Important examples have been the railroad, the automobile, and computers.
Other	Defeat of a socialist government stimulates foreign investment; peace breaks out, with an increase in world oil production, and lowers oil prices; good weather leads to lower food prices.

TABLE 22-1. Many Factors Can Increase Aggregate Demand and Shift out the AD Curve

The aggregate demand curve relates total spending to the price level. But numerous other influences affect aggregate demand—some policy variables, others exogenous factors. The table lists changes that would tend to increase aggregate demand and shift out the *AD* curve.

the macroeconomic *policy variables*, which are under government control. These are monetary policy (steps by which the central bank can affect interest rates and other financial conditions) and fiscal policy (taxes and government expenditures). Table 22-1 illustrates how these government policies can affect different components of aggregate demand.

The second set includes *exogenous variables*, or variables that are determined outside the AS-AD framework. As Table 22-1 shows, some of these variables (such as wars or revolutions) are outside the scope of macroeconomic analysis proper, some (such as foreign economic activity) are outside the control of domestic policy, and others (such as the stock market) have significant independent movement.

What are the effects of changes in the variables lying behind the *AD* curve? Consider the economic effects of a sharp increase in military spending, such as took place in World War II. The additional costs of the war included pay for the troops, purchases of ammunition and equipment, and costs of transportation. The effect of these purchases was an increase in *G*. Unless some other component of spending decreased to offset the increase in *G*, the total *AD* curve would shift out and to the right as *G* increased. Similarly, a radical new innovation that increased the profitability of new investment, or an increase in consumer wealth because of higher housing prices, would lead to an increase in aggregate demand and an outward shift of the *AD* curve.

Figure 22-4(b) on page 436 shows how the changes in the variables listed in Table 22-1 would affect the *AD* curve. To test your understanding, construct a similar table showing forces that would tend to decrease aggregate demand (see question 2 at the end of the chapter).



Two Reminders

We pause for two important reminders.

1. We first emphasize the difference between macroeconomic and microeconomic demand curves. Recall from our study of supply and demand that the microeconomic demand curve has the price of an individual commodity on the vertical axis and production of that commodity on the horizontal axis, with all other prices and total consumer incomes held constant.

In the aggregate demand curve, the general price level is on the vertical axis, while total output and incomes vary along the horizontal axis. By contrast, total incomes and output are held constant for the microeconomic demand curve.

Finally, the negative slope of the microeconomic demand curve occurs because consumers substitute other goods for the good in question when its price rises. If the price of meat rises, the quantity demanded falls because consumers tend to substitute bread and potatoes for meat, using more of the relatively inexpensive commodities and less of the relatively expensive one.

The aggregate demand curve is downward-sloping for completely different reasons: Total spending falls when the overall price level rises because consumer real incomes and real wealth fall, reducing consumption, and interest rates rise, reducing investment spending.

2. Remember also the important distinction between the *movement along a curve* and the *shift of a curve*. Figure 22-4(a) shows a case of movement along the aggregate demand curve. This might occur when higher oil prices reduce real disposable income. Figure 22-4(b) shows a shift of the aggregate demand curve. This might occur because of a sharp increase in war spending. Always keep this distinction in mind as you analyze a particular policy or shock.

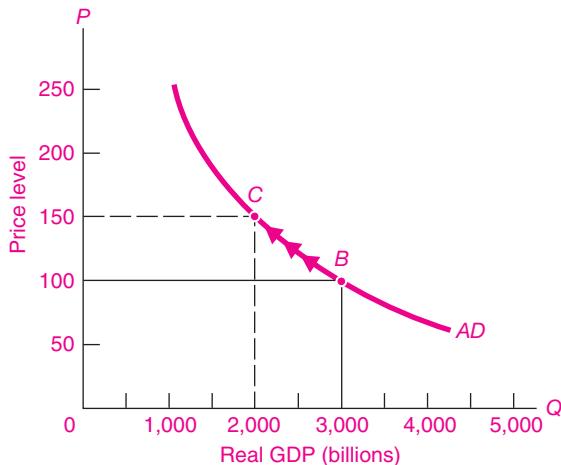
Business Cycles and Aggregate Demand

One important source of business fluctuations is shocks to aggregate demand. A typical case is illustrated in Figure 22-5 on page 436, which shows how a decline in aggregate demand lowers output. Say that the economy begins in short-run equilibrium at point *B*. Then, perhaps because of a financial panic or a tax increase, the aggregate demand curve shifts leftward to *AD'*. If there is no change in aggregate supply, the economy will reach a new equilibrium at point *C*. Note that output declines from *Q* to *Q'*. In addition, prices are now lower than they were at the previous equilibrium, and the rate of inflation falls.

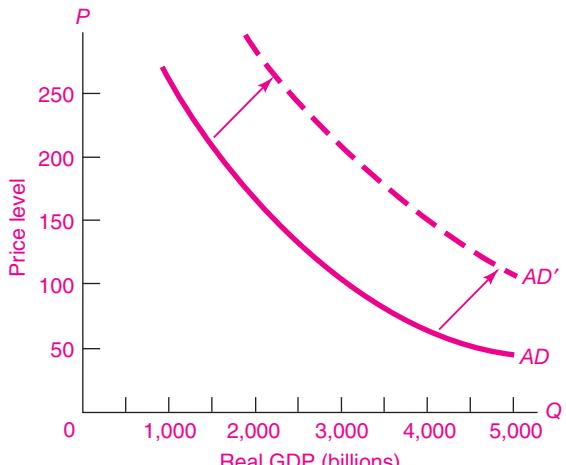
The case of an economic expansion is just the opposite. Suppose that a war leads to a sharp increase in government spending. As a result, the *AD* curve would shift to the right, output and employment would increase, and prices and inflation would rise.

Business-cycle fluctuations in output, employment, and prices are often caused by shifts in aggregate demand. These occur as consumers, businesses,

(a) Movements along the Aggregate Demand Curve

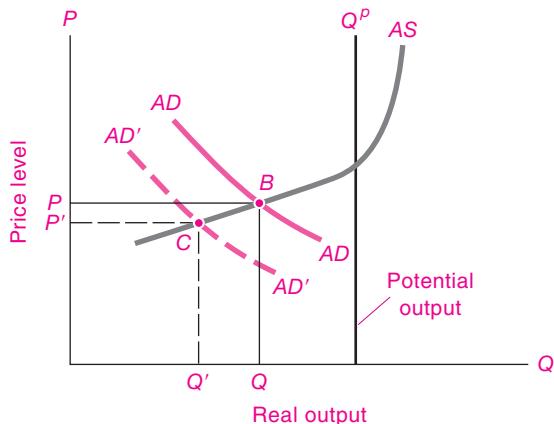


(b) Shifts of Aggregate Demand

**FIGURE 22-4. Movement along vs. Shifts of the Aggregate Demand Curve**

In (a), a higher price level with given nominal money incomes lowers real disposable income; this leads to higher interest rates and declining spending on interest-sensitive investment and consumption. This illustrates a *movement along* the AD curve from B to C when other things are held constant.

In (b), other things are no longer constant. Changes in variables underlying AD —such as the money supply, tax policy, or military spending—lead to changes in total spending at a given price level. This leads to a *shift of* the AD curve.

**FIGURE 22-5. A Decline in Aggregate Demand Leads to an Economic Downturn**

A downward shift in the AD curve along a relatively flat and unchanging AS curve leads to lower levels of output. Note that as a result of the leftward shift in the AD curve, actual output declines relative to potential output and makes a recession worse.

or governments change total spending relative to the economy's productive capacity. When these shifts in aggregate demand lead to sharp business downturns, the economy suffers recessions or even depressions. A sharp upturn in economic activity can lead to inflation.

Is the Business Cycle Avoidable?

The history of business cycles in the United States shows a remarkable trend toward greater stability in the last quarter-century (look back at Figure 22-2). The period through 1940 witnessed numerous crises and depressions—prolonged, cumulative slumps like those of the 1870s, 1890s, and 1930s. Since 1945, business cycles have become less frequent and milder, and many Americans have never witnessed a real Depression.

What were the sources of the Great Moderation? Some believe that capitalism is inherently more stable now than it was in earlier times. Some of that stability comes from a larger and more predictable government sector. Equally important is a better

understanding of macroeconomics that now permits the government to conduct its monetary and fiscal policies so as to prevent shocks from turning into recessions and to keep recessions from snowballing into depressions.

During tranquil periods, people often declare that the business cycle has been vanquished. Is this a realistic possibility? While business cycles have moderated in America over the last quarter-century, they have actually become more prevalent in other economies. So take heed of the following prophetic words of the great macroeconomist Arthur Okun, which are particularly appropriate as the world economy heads into recession in 2007–2009:

Recessions are now generally considered to be fundamentally preventable, like airplane crashes and unlike hurricanes. But we have not banished air crashes from the land, and it is not clear that we have the wisdom or the ability to eliminate recessions. The danger has not disappeared. The forces that produce recurrent recessions are still in the wings, merely waiting for their cue.

C. THE MULTIPLIER MODEL

The basic macroeconomic theory of business cycles holds that shifts in aggregate demand produce the frequent and unpredictable fluctuations in output, prices, and employment known as business cycles. Economists try to understand the *mechanism* by which changes in spending get translated into changes in output and employment. The simplest approach to understanding business cycles is known as the *Keynesian multiplier model*.

When economists attempt to understand why major increases in military spending led to rapid increases in GDP, or why the tax cuts of the 1960s or 1980s ushered in long periods of business-cycle expansions, or why the investment boom of the late 1990s produced America's longest expansion, they often turn to the Keynesian multiplier model for the simplest explanation.

What exactly is the **multiplier model**? It is a macroeconomic theory used to explain how output is determined in the short run. The name “multiplier” comes from the finding that each dollar change in exogenous expenditures (such as investment) leads to more than a dollar change (or a multiplied

change) in GDP. The key assumptions underlying the multiplier model are that wages and prices are fixed and that there are unemployed resources in the economy. In addition, in this introductory chapter, we are ignoring the role of monetary policy and assuming that financial markets do not react to changes in the economy. Additionally, we are for now assuming that there is no international trade and finance. These further elaborations will be introduced in later chapters.

OUTPUT DETERMINED BY TOTAL EXPENDITURES

Our initial discussion of the multiplier model analyzes how investment and consumption spending interact with incomes to determine national output. This is called the *total expenditure approach* to determining national output.

Recall Chapter 21's picture of the national consumption function. We have drawn a reminder graph in Figure 22-6, where the consumption function is

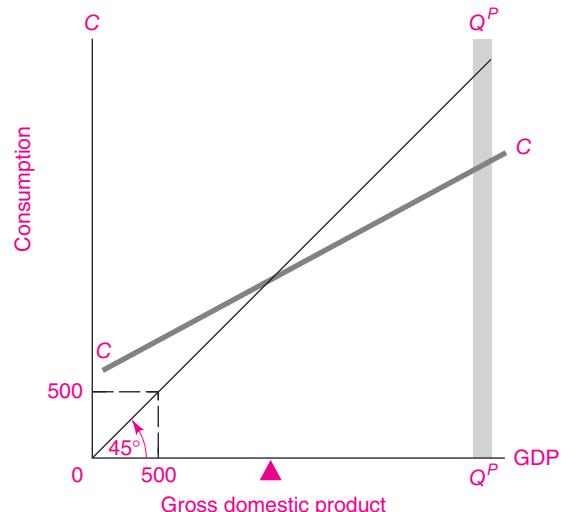


FIGURE 22-6. National Income Determines the Level of Consumption

Recall the consumption function, CC , that was described in Chapter 21. This shows the level of consumption expenditures corresponding to every level of income (where income equals GDP in this simple example). The two points marked “500” emphasize the important property of the 45° line. Any point on the 45° line depicts a vertical distance exactly equal to the horizontal distance. The blue band marked $Q_p Q_p$ shows the level of potential GDP.

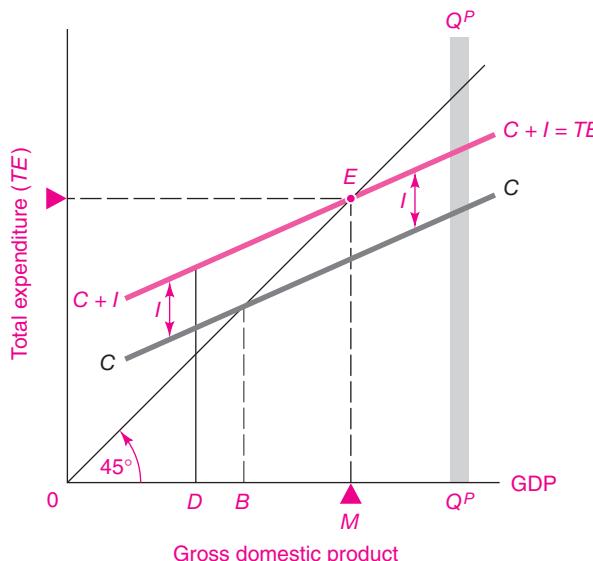


FIGURE 22-7. The Equilibrium Level of National Output Is Determined When Total Expenditure (TE) Equals Output

The blue CC line represents the consumption function (shown in Figure 22-6). The I arrows indicate constant investment. Adding I to CC gives the TE curve of total desired investment plus consumption spending. Along the 45° line, expenditures exactly equal GDP. Equilibrium GDP comes at point E , which is the intersection of the TE line and the 45° line. This is the only level of GDP at which the desired spending on $C + I$ exactly equals output.

the CC line. Recall that the consumption function shows the desired consumption corresponding to each level of income. We have omitted taxes, transfers, and other items, so that personal income equals national income, and national income equals GDP.

We now develop in Figure 22-7 an important new graph showing the total expenditure-output relationship. This graph is sometimes called the “Keynesian cross,” because it shows how output equals expenditure when the expenditure curve crosses the 45° line. (If you are not sure about the significance of the 45° line, look back at Chapter 21’s explanation.)

We begin by drawing the consumption function, CC . We then add total investment to consumption. Normally, investment depends on interest rates, tax policy, and business confidence. To simplify things, we treat investment as an *exogenous* variable, one whose level is determined outside the model. Say that

investment opportunities are such that investment would be exactly \$200 billion per year regardless of the level of GDP. The investment schedule is stacked on top of the consumption schedule in Figure 22-7. Note that the $C + I$ curve is higher than the C curve by exactly the constant amount of I . This parallel feature indicates that investment is constant.

This $C + I$ curve represents total expenditures (TE), which equals desired investment (which is at fixed level I) plus consumption. This is drawn in Figure 22-7 as the green $C + I$ or TE curve.

Finally, we draw in a 45° line along which expenditure on the vertical axis exactly equals output on the horizontal axis. At any point on the 45° line, total desired expenditure (measured vertically) exactly equals the total level of output (measured horizontally).

We can now calculate the equilibrium level of output in Figure 22-7. Where planned expenditure, represented by the TE curve, equals total output, the economy is in equilibrium.

The total expenditure curve (TE) shows the level of expenditure desired or planned by consumers and businesses corresponding to each level of output. The economy is in equilibrium at the point where the $TE = C + I$ curve crosses the 45° line—at point E in Figure 22-7. Point E is the macroeconomic equilibrium because at that point, the level of desired expenditure on consumption and investment exactly equals the level of total output.

Reminder on the Meaning of Equilibrium

We often look for a macroeconomic “equilibrium” when analyzing business cycles or economic growth. What exactly does this term mean? An **equilibrium** is a situation where the different forces at work are in balance. For example, if you see a ball rolling down a hill, the ball is not in equilibrium because the forces at work are pulling the ball down. This is therefore a **disequilibrium**. When the ball comes to rest in a valley at the bottom of the hill, the forces operating on the ball are in balance. This is therefore an equilibrium.

Similarly, in macroeconomics, an equilibrium level of output is one where the different forces of spending and output are in balance; in equilibrium, the level of output tends to persist until there are changes in the forces affecting the economy.

Applying the equilibrium concept to Figure 22-7, we see that point *E* is an equilibrium. At point *E*, and only at point *E*, does *desired spending on C + I equal actual output*. At any other level of production, desired spending would differ from production. At any level other than *E*, businesses would find themselves producing too little or too much and would want to change the level of production back toward the equilibrium level.

The Adjustment Mechanism

It is not enough to say that point *E* is an equilibrium. We need to understand *why* a certain output is an equilibrium and what would happen if output deviated from that equilibrium. Let's consider three cases: planned spending above output, planned spending below output, and planned spending equal to output.

In the first case, suppose that spending is above output. This is represented by point *D* in Figure 22-7. At this level of output, the *C + I* spending line is above the 45° line, so planned *C + I* spending would be greater than output. This means that consumers would be buying more goods than businesses had anticipated. Auto dealers would find their lots emptying, and the backlog for computers would be getting longer and longer.

In such a disequilibrium situation, auto dealers and computer stores would respond by increasing their orders. Automakers would recall workers from layoff and gear up their production lines, while computer makers would add additional shifts. As a result of this increased production, output would increase. *Therefore, a discrepancy between total planned expenditure and total output leads to an adjustment of output.*

You should also work through what happens in the second case, where output is below equilibrium.

Finally, take the third case, where planned expenditure exactly equals output. At equilibrium, firms will find that their sales are equal to their forecasts. Inventories will be at their planned levels. There will not be any unexpected orders. Firms cannot improve profits by changing output because planned consumption needs have been met. So production, employment, income, and spending will remain the same. In this case GDP stays at point *E*, and we can rightly call it an *equilibrium*.

The equilibrium level of GDP occurs at point *E*, where planned spending equals planned production.

At any other output, the total desired spending on consumption and investment differs from the planned production. Any deviation of plans from actual levels will cause businesses to change their production and employment levels, thereby returning the system to the equilibrium GDP.

A Numerical Analysis

An example may help show why the equilibrium level of output occurs where planned spending and planned output are equal.

Table 22-2 shows a simple example of consumption, saving, and output. The break-even level of income, where consumption equals income, is \$3000 billion (\$3 trillion). Each \$300 billion change of income is assumed to lead to a \$100 billion change in saving and a \$200 billion change in consumption. In other words, the *MPC* is assumed to be constant and equal to $\frac{2}{3}$.

We assume that investment is exogenous and always sustainable at \$200 billion, as shown in column (4) of Table 22-2.

Columns (5) and (6) are the crucial ones. Column (5) shows the total GDP. It is simply column (1) copied again into column (5). The figures in column (6) represent total planned expenditures at each level of GDP; that is, it equals the planned consumption spending plus planned investment. It is the *C + I* schedule from Figure 22-7 in numbers.

When businesses as a whole are producing too much output (higher than the sum of what consumers and businesses want to purchase), inventories of unsold goods will be piling up.

Reading from the top row of Table 22-2, we see that if firms are initially producing \$4200 billion of GDP, planned or desired spending [shown in column (6)] is only \$4000 billion. In this situation, excess inventories will be accumulating. Firms will respond by reducing their production levels, and GDP will fall. In the opposite case, represented in the bottom row of Table 22-2, total spending is \$3000 billion but output is only \$2700 billion. Inventories are being depleted and firms will expand operations, raising output.

We see, then, that when businesses as a whole are temporarily producing more than they can profitably sell, they will reduce production and GDP will fall. When they are selling more than their current production, they will increase their output, and GDP will rise.

GDP Determination Where Output Equals Planned Spending (billions of dollars)							
(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Levels of GDP and DI	Planned consumption	Planned saving $(3) = (1) - (2)$	Planned investment	Level of GDP $(5) = (1)$	Total planned consumption and investment, TE $(6) = (2) + (4)$	Resulting tendency of output	
4,200	3,800	400	200	4,200	> 4,000	↓	Contraction
3,900	3,600	300	200	3,900	> 3,800	↓	Contraction
3,600	3,400	200	200	3,600	= 3,600		Equilibrium
3,300	3,200	100	200	3,300	< 3,400	↑	Expansion
3,000	3,000	0	200	3,000	< 3,200	↑	Expansion
2,700	2,800	-100	200	2,700	< 3,000	↑	Expansion

TABLE 22-2. Equilibrium Output Can Be Found Arithmetically at the Level Where Planned Spending Equals GDP

The darker green row depicts the equilibrium GDP level, where the \$3600 that is being produced is just matched by the \$3600 that households plan to consume and that firms plan to invest. In upper rows, firms will be forced into unintended inventory investment and will respond by cutting back production until equilibrium GDP is reached. Interpret the lower rows' tendency toward expansion of GDP toward equilibrium.

Only when the level of actual output in column (5) exactly equals planned expenditure (TE) in column (6) will the economy be in equilibrium. In equilibrium, and only in equilibrium, business sales will be exactly sufficient to justify the current level of aggregate output. In equilibrium, GDP will neither expand nor contract.

THE MULTIPLIER

Where is the multiplier in all this? To answer this question, we need to examine how a change in exogenous investment spending affects GDP. It is logical that an increase in investment will raise the level of output and employment. But by how much? The multiplier model shows that an increase in investment will increase GDP by an amplified or multiplied amount—by an amount greater than itself.

The **multiplier** is the impact of a 1-dollar change in exogenous expenditures on total output. In the simple $C + I$ model, the multiplier is the ratio of the change in total output to the change in investment.

Note that the definition of the multiplier speaks of the change in output per unit change in *exogenous*

expenditures. This indicates that we are taking certain components of spending as given outside the model. In the case in hand, the exogenous component is investment. Later, we will see that the same approach can be used to determine the effect of changes in government expenditures, exports, and other items on total output.

For example, suppose investment increases by \$100 billion. If this causes an increase in output of \$300 billion, the multiplier is 3. If, instead, the resulting increase in output is \$400 billion, the multiplier is 4.

Woodsheds and Carpenters. Why is it that the multiplier is greater than 1? Let's suppose that I hire unemployed workers to build a \$1000 woodshed. My carpenters and lumber producers will get an extra \$1000 of income. But that is not the end of the story. If they all have a marginal propensity to consume of $\frac{2}{3}$, they will now spend \$666.67 on new consumption goods. The producers of these goods will now have extra incomes of \$666.67. If their MPC is also $\frac{2}{3}$, they in turn will spend \$444.44, or $\frac{2}{3}$ of \$666.67 (or $\frac{2}{3}$ of $\frac{2}{3}$ of \$1000). The process will go on, with each new round of spending being $\frac{2}{3}$ of the previous round.

Thus an endless chain of *secondary consumption spending* is set in motion by my *primary investment* of \$1000. But, although an endless chain, it is an ever-diminishing one. Eventually it adds up to a finite amount.

Using straightforward arithmetic, we can find the total increase in spending in the following manner:

$$\begin{array}{r}
 \$1000.00 \\
 + \\
 666.67 \\
 + \\
 444.44 \\
 + \\
 296.30 \\
 + \\
 197.53 \\
 + \\
 \vdots \\
 \hline
 \$3000.00
 \end{array}
 =
 \left\{
 \begin{array}{l}
 1 \times \$1000 \\
 + \\
 \frac{2}{3} \times \$1000 \\
 + \\
 (\frac{2}{3})^2 \times \$1000 \\
 + \\
 (\frac{2}{3})^3 \times \$1000 \\
 + \\
 (\frac{2}{3})^4 \times \$1000 \\
 + \\
 \vdots \\
 \hline
 \frac{1}{1 - \frac{2}{3}} \times \$1000, \text{ or } 3 \times \$1000
 \end{array}
 \right.$$

This shows that, with a *MPC* of $\frac{2}{3}$, the multiplier is 3; it consists of the 1 of primary investment plus 2 extra of secondary consumption responding.

The same arithmetic would give a multiplier of 4 for a *MPC* of $\frac{3}{4}$, because $1 + \frac{3}{4} + (\frac{3}{4})^2 + (\frac{3}{4})^3 + \dots$ eventually adds up to 4. For a *MPC* of $\frac{1}{2}$, the multiplier would be 2.¹

The size of the multiplier thus depends upon how large the *MPC* is. It can also be expressed in terms of the twin concept, the *MPS*. For a *MPS* of $\frac{1}{4}$, the *MPC* is $\frac{3}{4}$ and the multiplier is 4. For a *MPS* of $\frac{1}{3}$, the multiplier is 3. If the *MPS* were $1/x$, the multiplier would be x .

By this time it should be clear that the simple multiplier is always the inverse, or reciprocal, of the marginal propensity to save. It is thus equal to $1/(1 - \text{MPC})$. Our simple multiplier formula is

$$\begin{aligned}
 \text{Change in output} &= \frac{1}{\text{MPS}} \times \text{change in investment} \\
 &= \frac{1}{1 - \text{MPC}} \times \text{change in investment}
 \end{aligned}$$

¹ The formula for an infinite geometric progression is

$$1 + r + r^2 + r^3 + \dots + r^n + \dots = \frac{1}{1 - r}$$

as long as *MPC* (r) is less than 1 in absolute value.

The Multiplier Model Compared with the AS-AD Model

As you study the multiplier model, you might begin to wonder how this model fits in with the *AS-AD* model of Chapter 19. These are not, in fact, different approaches. Rather, the multiplier model is a special case of the aggregate demand-and-supply model. It explains how *AD* is affected by consumption and investment spending under certain precise assumptions.

One of the key assumptions in the multiplier analysis is that prices and wages are fixed in the short run. This is an oversimplification, for many prices adjust quickly in the real world. But this assumption captures the point that if some wages and prices are sticky—which is most definitely the case—then some of the adjustment to *AD* shifts will come through output adjustments. We will return to this important point in later chapters.

We can show the relationship between the multiplier analysis and the *AS-AD* approach in Figure 22-8. Part (b) displays an *AS* curve that becomes completely vertical when output equals potential output. However, when there are unemployed resources—to the left of potential output in the graph—output will be determined primarily by the strength of aggregate demand. As investment increases, this increases *AD*, and equilibrium output rises.

The same economy can be described by the multiplier diagram in the top panel of Figure 22-8. The multiplier equilibrium gives the same level of output as the *AS-AD* equilibrium—both lead to a real GDP of Q . They simply stress different features of output determination.

This discussion again points to a crucial feature of the multiplier model. While it is a useful model for describing recessions or even depressions, it cannot apply to periods of full employment. Once factories are operating at full capacity and all workers are employed, the economy simply cannot produce more output.

D. FISCAL POLICY IN THE MULTIPLIER MODEL

For centuries, economists have understood the *allocational* role of fiscal policy (government tax and spending programs). It has long been known that

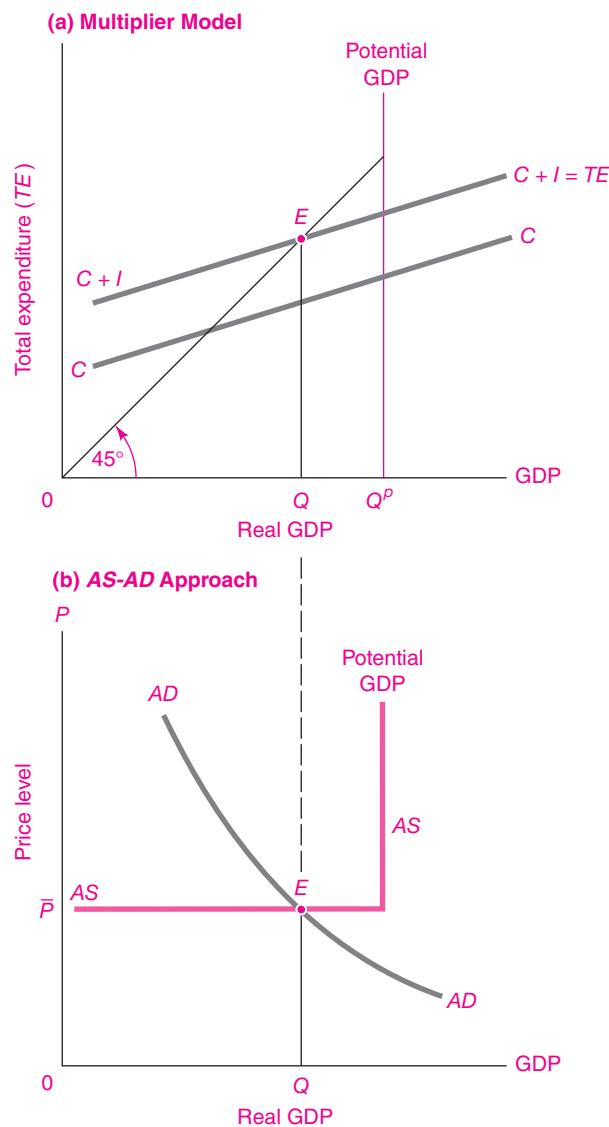


FIGURE 22-8. How the Multiplier Model Relates to the AS-AD Approach

The multiplier model is a way of understanding the workings of the AS-AD equilibrium.

(a) The top panel shows the output-expenditure equilibrium in the multiplier model. At point E , the spending line just cuts the 45° line, leading to equilibrium output of Q .

(b) The equilibrium can also be seen in the bottom panel, where the AD curve cuts the AS curve at point E . In this simplest business-cycle model wages and prices are assumed to be fixed, so the AS curve is horizontal until full employment is reached. Both approaches lead to exactly the same equilibrium output, Q .

fiscal programs are instrumental in deciding how the nation's output should be divided between collective and private consumption and how the burden of payment for collective goods should be divided among the population.

Only with the development of modern macroeconomic theory has a surprising fact been uncovered: Government fiscal powers also have a major *macroeconomic* impact upon the short-run movements of output, employment, and prices. The knowledge that fiscal policy has powerful effects upon economic activity led to the *Keynesian approach to macroeconomic policy*, which is the active use of government action to moderate business cycles. This approach was described by the Nobel Prize-winning macroeconomist James Tobin as follows:

Keynesian policies are, first, the explicit dedication of macroeconomic policy instruments to real economic goals, in particular full employment and real growth of national income. Second, Keynesian demand management is activist. Third, Keynesians have wished to put both fiscal and monetary policies in consistent and coordinated harness in the pursuit of macroeconomic objectives.

In this section we use the multiplier model to show how government purchases affect output.

HOW GOVERNMENT FISCAL POLICIES AFFECT OUTPUT

To understand the role of government in economic activity, we need to look at government purchases and taxation, along with the effects of those activities on private-sector spending. We now modify our earlier analysis by adding G to $C + I$ to get a new total expenditure curve $TE = C + I + G$. This new schedule can describe the macroeconomic equilibrium when government, with its spending and taxing, is in the picture.

It will simplify our task in the beginning if we analyze the effects of government purchases with total taxes collected held constant (taxes that do not change with income or other economic variables are called *lump-sum taxes*). But even with a fixed dollar value of taxes, we can no longer ignore the distinction between disposable income and gross domestic product. Under simplified conditions (including no foreign trade, transfers, or depreciation), we know