PART THREE

Factor Markets: Labor, Land, and Capital

CHAPTER

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	1,056	8.6	
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.*

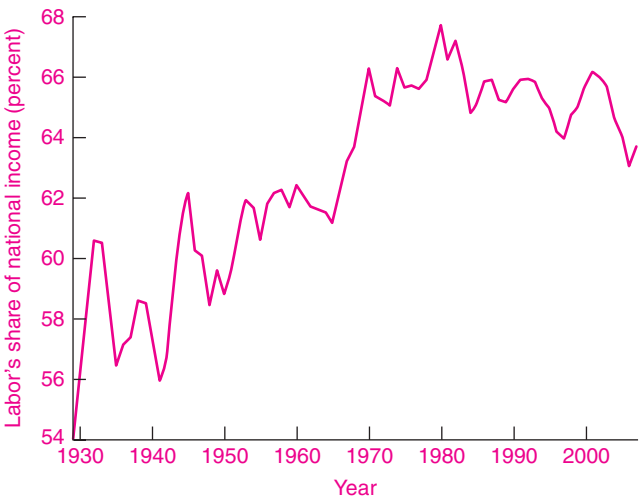

FIGURE 12-1. The Share of Labor in National Income

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.

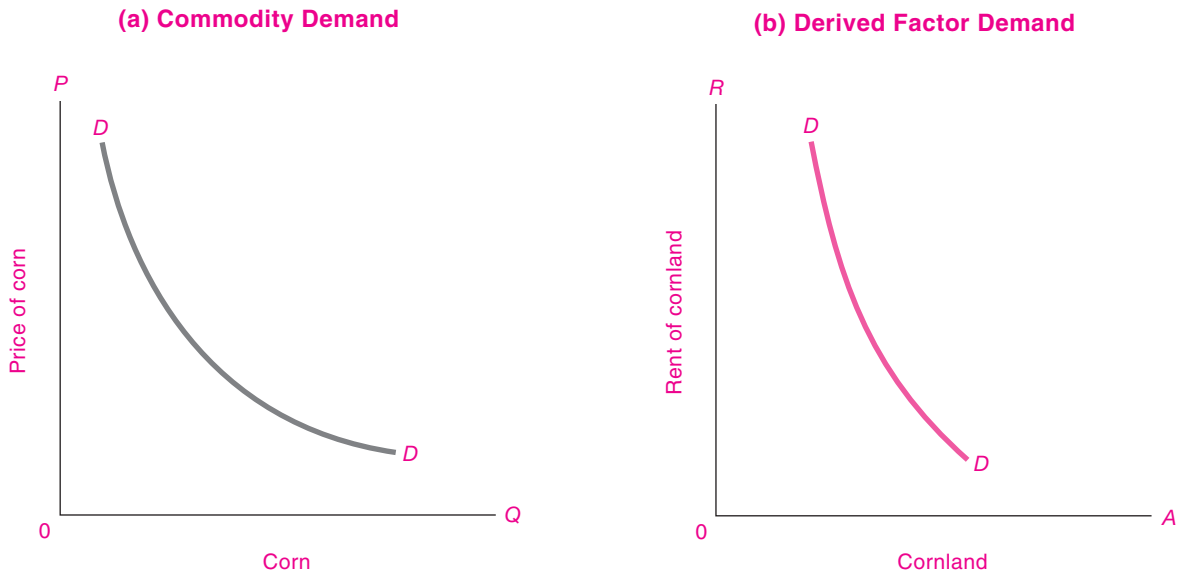

FIGURE 12-2. Demand for Factors Is Derived from Demand for Goods They Produce

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.

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 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.

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 × \$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 × 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:

$$\text{Marginal revenue product of labor } (MRP_L) = MR \times MP_L$$
$$\text{Marginal revenue product of land } (MRP_A) = MR \times MP_A$$

and so forth.

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

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

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:

Marginal product of labor × output price = price of labor = wage rate

Marginal product of land × output price = price of land = rent

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:

$$\frac{\text{Marginal product of labor}}{\text{Price of labor}} = \frac{\text{marginal product of land}}{\text{price of land}} = \cdots = \frac{1}{\text{marginal revenue}}$$

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

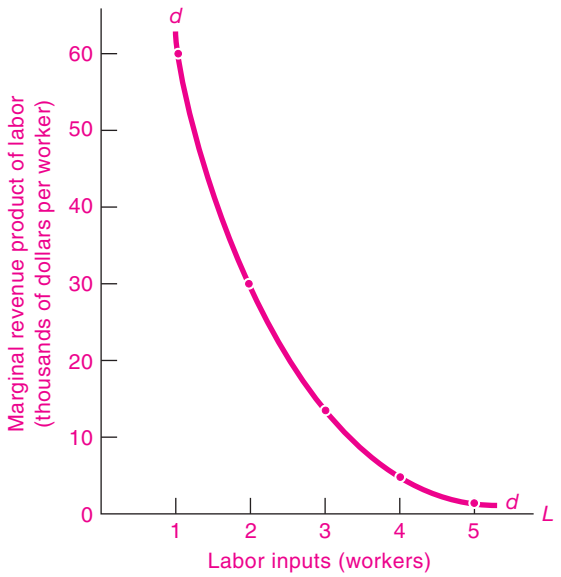

FIGURE 12-3. Demand for Inputs Derived through Marginal Revenue Products

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.

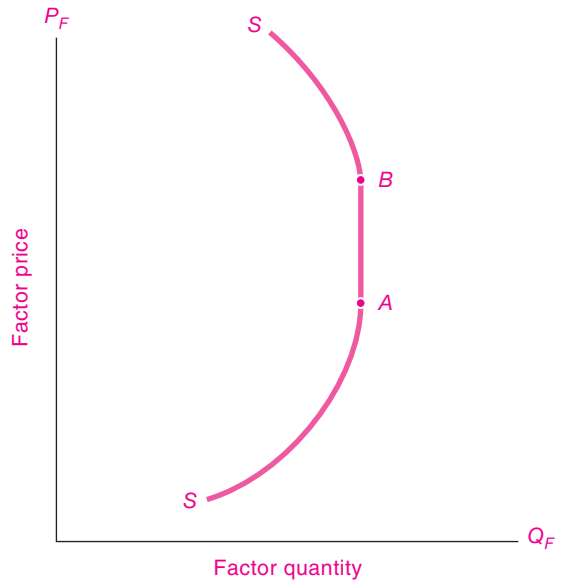

FIGURE 12-4. Supply Curve for Factors of Production

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.

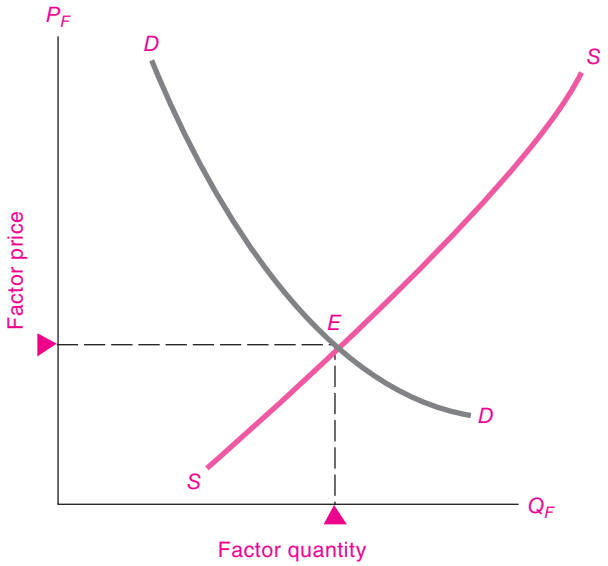

FIGURE 12-5. Factor Supply and Derived Demand Interact to Determine Factor Prices and Income Distribution

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.

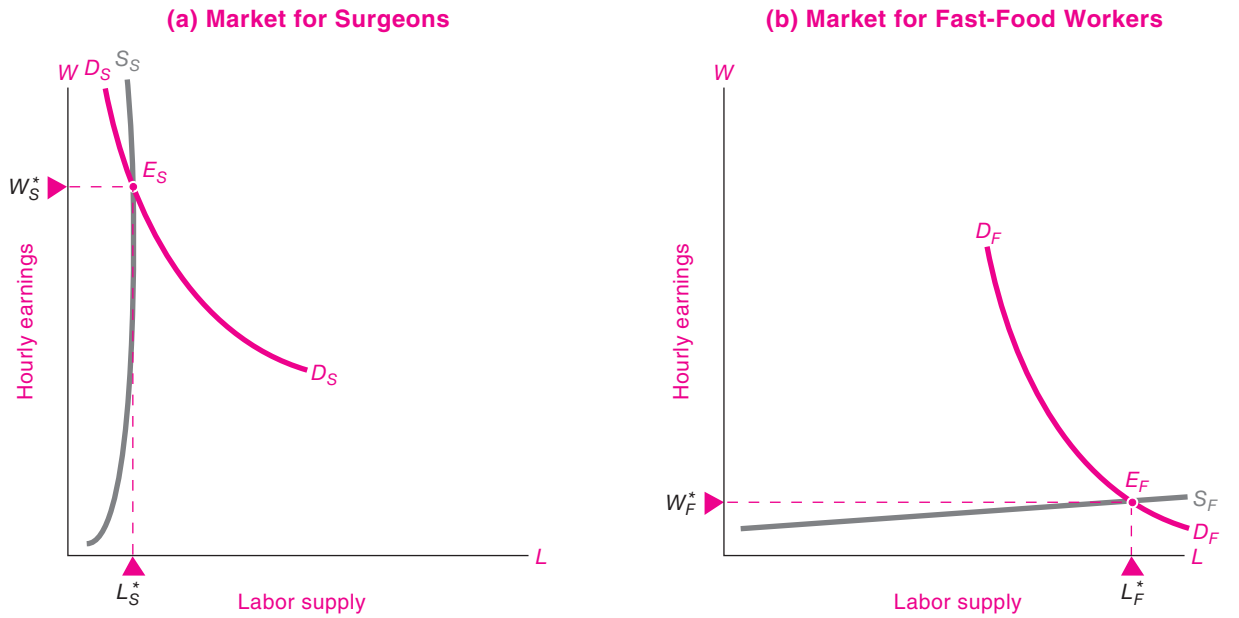

FIGURE 12-6. The Markets for Surgeons and Fast-Food Workers

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

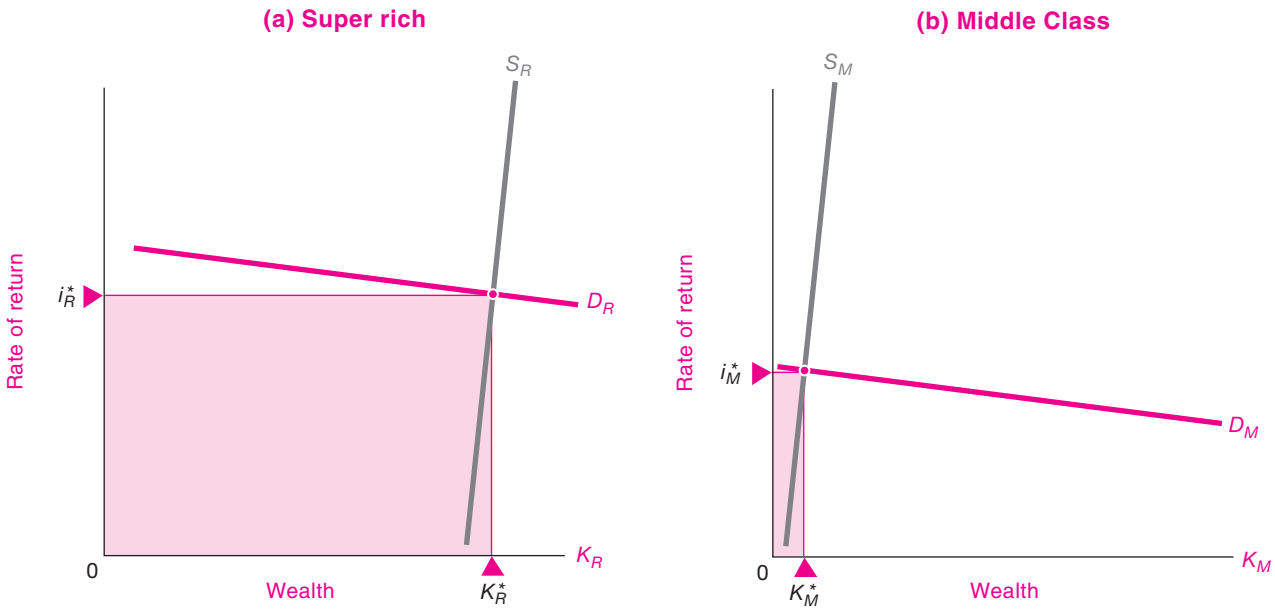

FIGURE 12-7. Differences in Total Returns to Wealth

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 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 $W = 5$ and $L = 1$ million, total wages = 5 million); this is shown by the dark rectangle, *0SEN*.

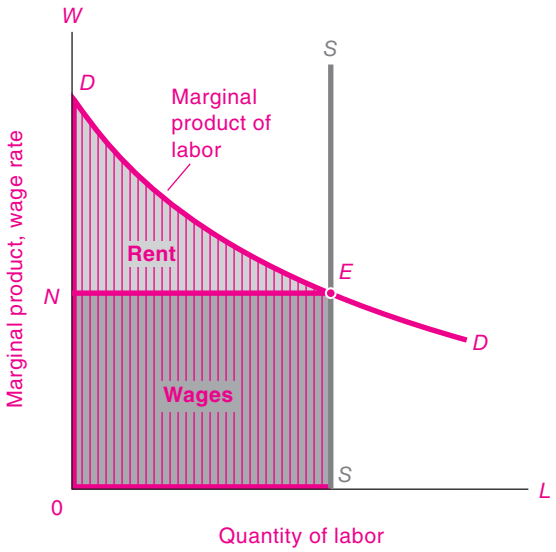

FIGURE 12-8. Marginal Product Principles Determine Factor Distribution of Income

Each vertical slice represents the marginal product of that unit of labor. Total national output *0DES* 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 *0N* times the quantity of labor *0S*). Land rents get the residual upper triangle *NDE*.

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 *MP*s 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:

$$\frac{MP_L}{P_L} = \frac{MP_A}{P_A} = \cdots = \frac{1}{\text{marginal revenue}}$$

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			
		____	5	____
1	30			
		____	5	____
2	50			
		____	5	____
3	60			
		____	5	____
4	65			
		____	5	____
5	68			
		____	5	____
6	68			

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.

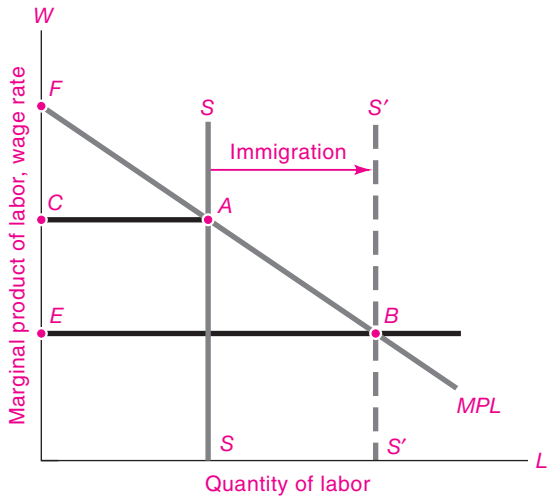

FIGURE 12-9.

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?

CHAPTER

13

The Labor Market

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.

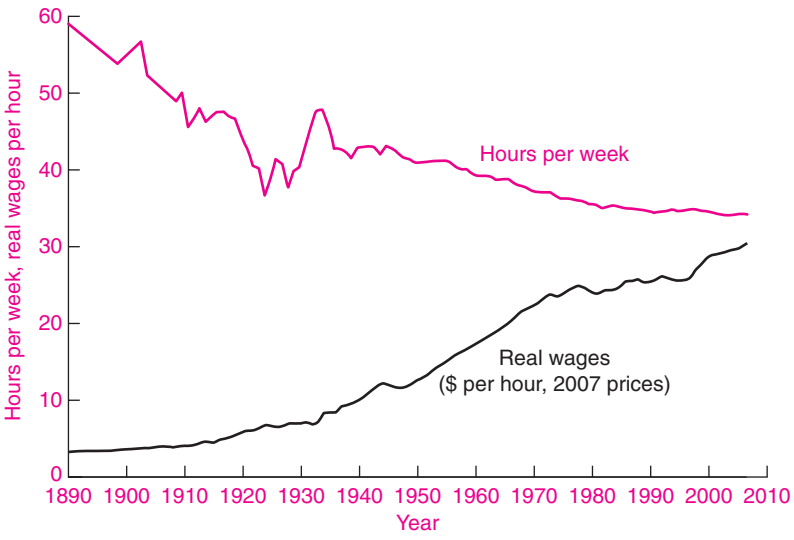

FIGURE 13-1. Wages Have Improved as Hours of Work Have Declined

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

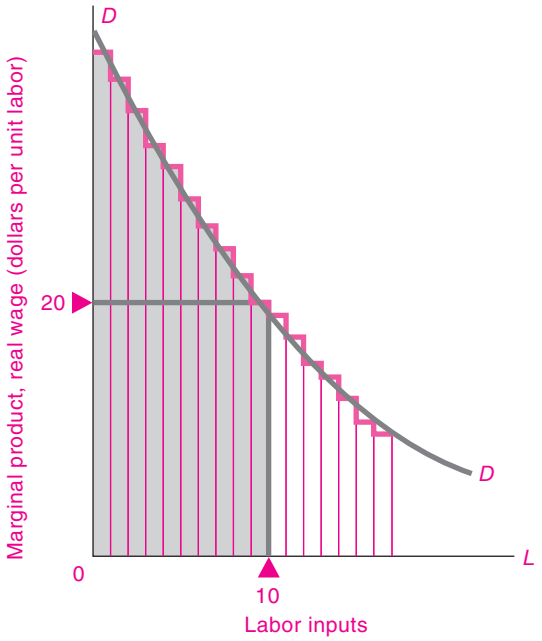

FIGURE 13-2. Demand for Labor Reflects Marginal Productivity

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 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

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/ichccpwsuppt02.txt* and estimates by the authors. Note these estimates use market exchange rates and not purchasing-power-parity exchange rates.

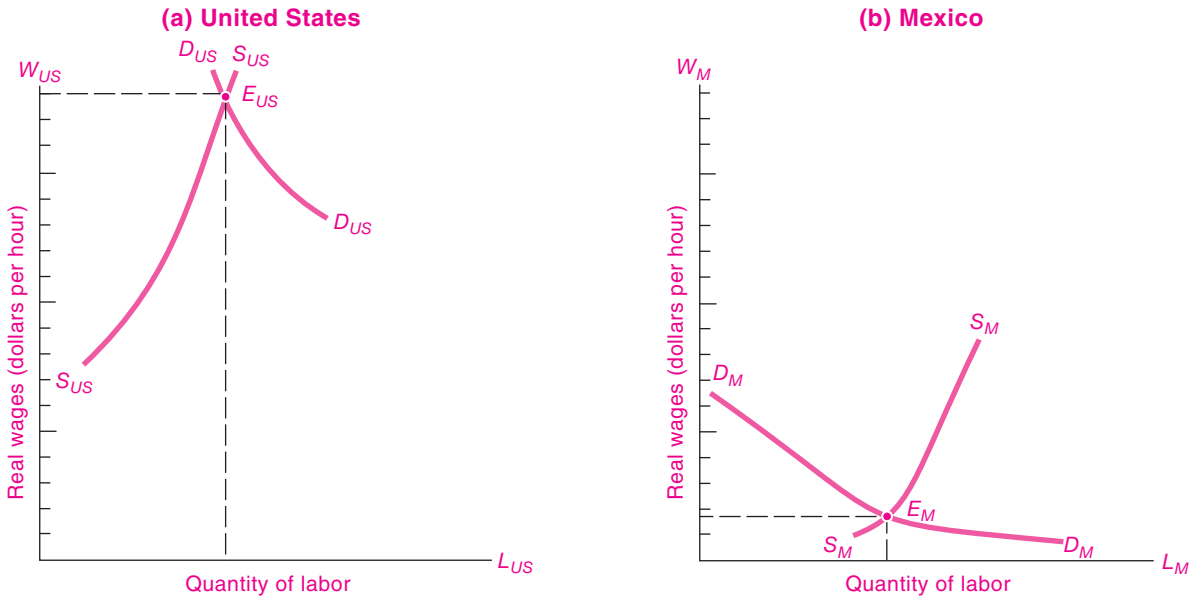

FIGURE 13-3. Favorable Resources, Skills, Management, Capital, and Technology Explain High U.S. Wages

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?

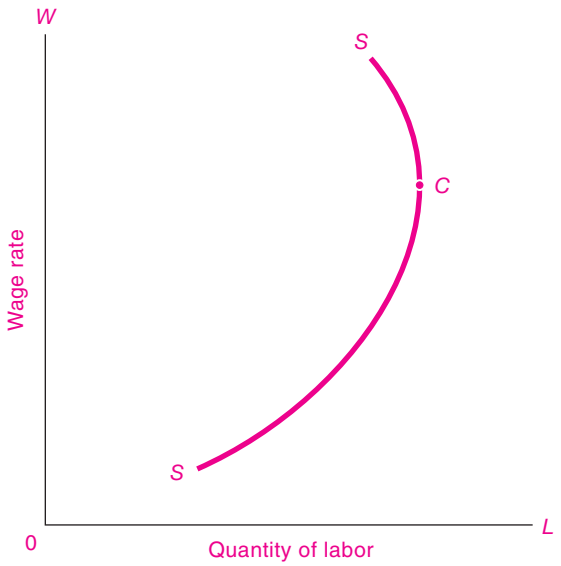

FIGURE 13-4. As Wages Rise, Workers May Work Fewer Hours

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

Labor-Supply Patterns

Group of workers	Labor-Force Participation Rate (% of population) 1960	2007	Response of labor supply to increase in real wages
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

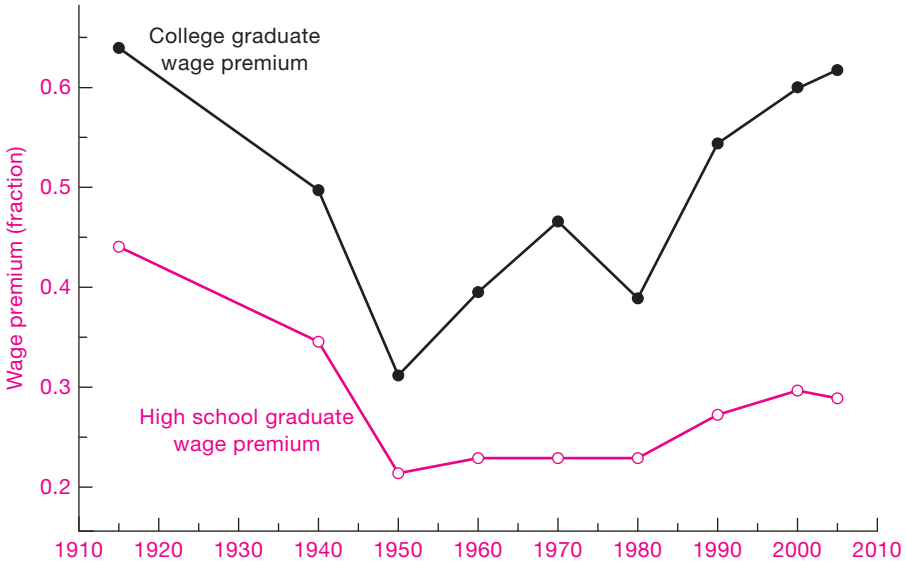

FIGURE 13-5. Relative Income Gains Have Been Dramatic for College Graduates

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 nonunion 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

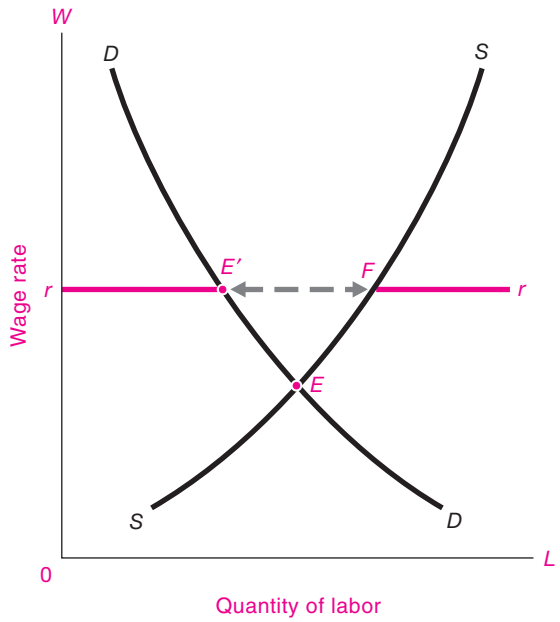

FIGURE 13-6. Unions Set High Standard Wage and Limit Employment

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_pS_p, while the demand for such labor is depicted as D_pD_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.

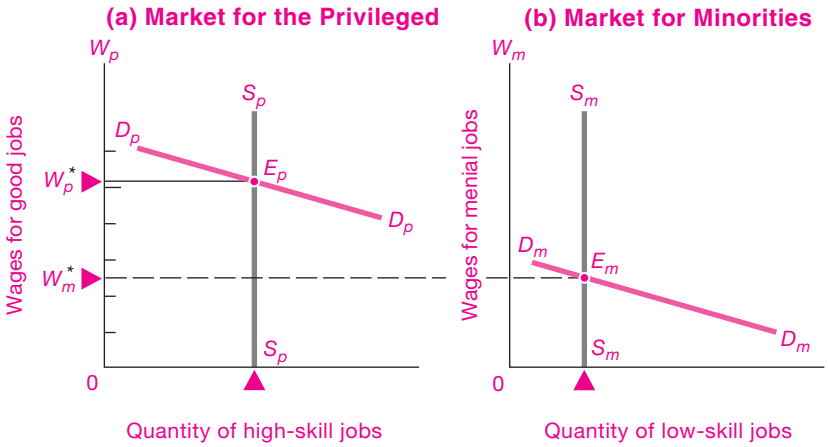

FIGURE 13-7. Discrimination by Exclusion Lowers the Wage Rates of Excluded Minorities

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 noncompeting groups

Labor Market Issues

collective bargaining
unions as monopolies
classical unemployment
discrimination
earnings differentials: quality differences vs. discrimination
statistical discrimination
antidiscrimination policies

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/.*

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.

CHAPTER

14

Land, Natural Resources, and the Environment

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 **inappropriable** when some of the costs and benefits associated with its use do not accrue to its owner. In other words, inappropriable 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 inappropriable 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

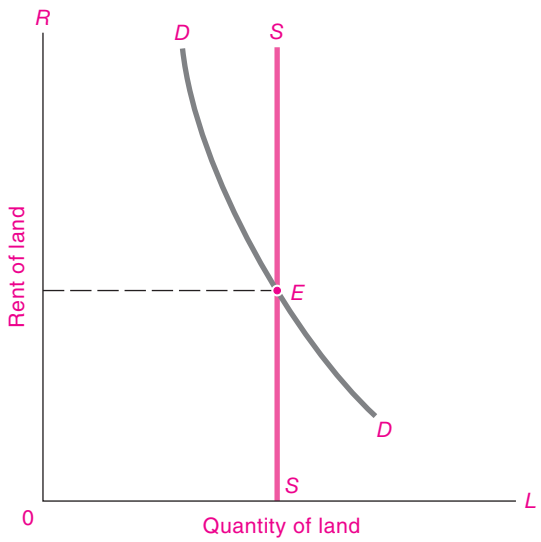

FIGURE 14-1. Fixed Land Must Work for Whatever It Can Earn

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.

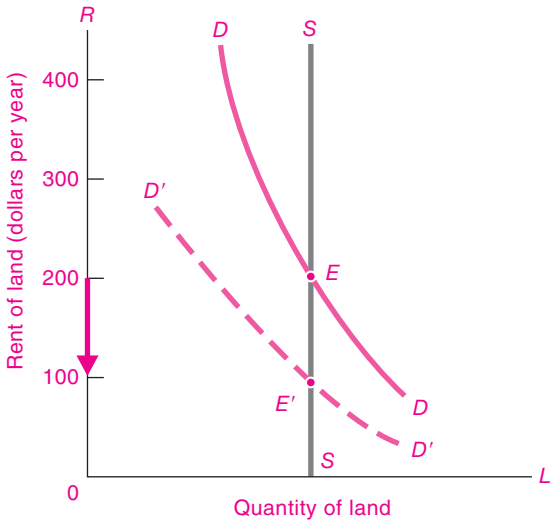

FIGURE 14-2. Tax on Fixed Land Is Shifted Back to Landowners, with Government Skimming Off Pure Economic Rent

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.

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 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 batallions, 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 marginal costs and benefits on the vertical axis and the abatement 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

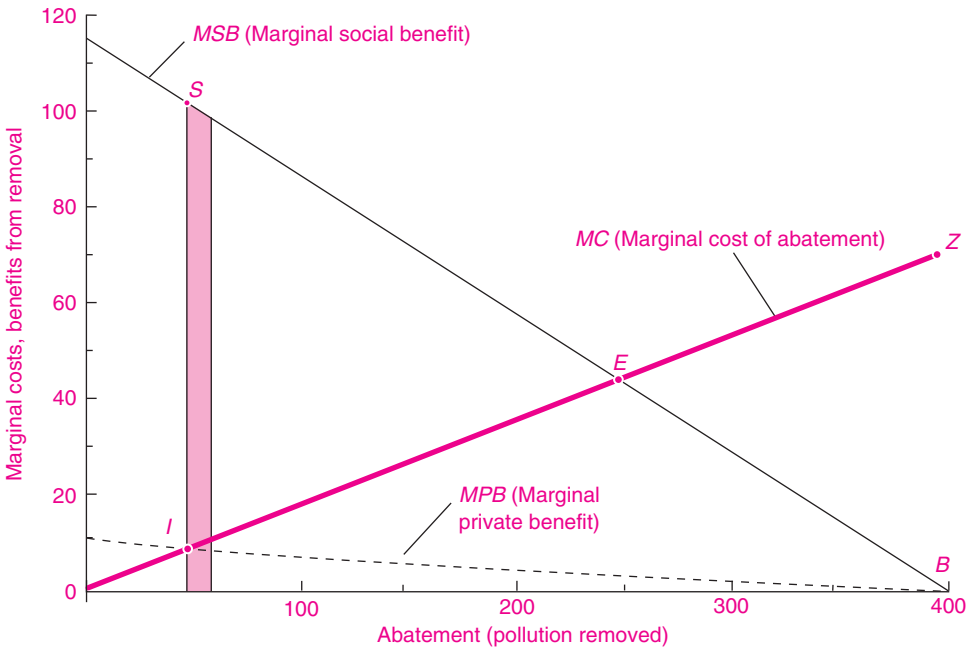

FIGURE 14-3. Inefficiency from Externalities

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

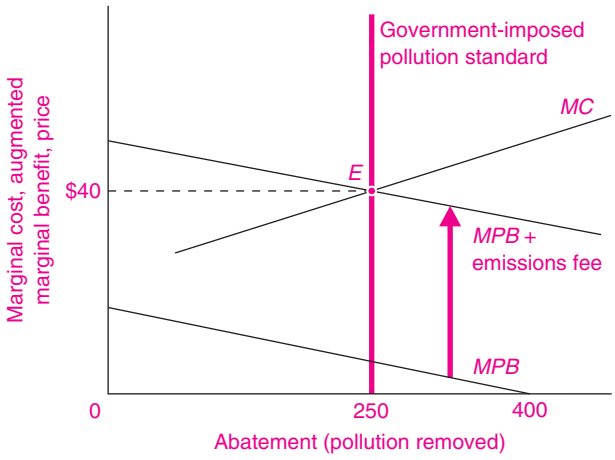

FIGURE 14-4. Pollution Standards and Emissions Fees

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

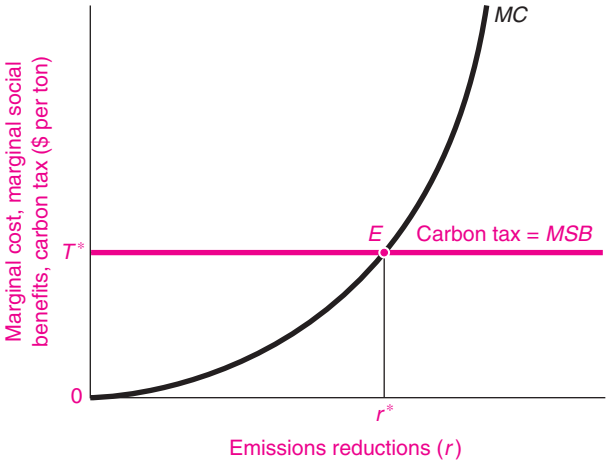

FIGURE 14-5. Carbon Taxes Can Slow Harmful Climate Change

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 CO_2 reduction, be achieved? Since CO_2 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 nonbinding commitments to limit the emissions of greenhouse gases like CO_2.

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 CO_2 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.
   c. Consider the quotation in **b.** Why is this correct for the market as a whole but incorrect for the individual farmer? Explain the fallacy of composition that is at work here.
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?
   a. Steel mills emitting sulfur oxides into the Birmingham air
   b. Smoking by people in restaurants
   c. Smoking by students without roommates in their own rooms
   d. Driving by persons under the influence of alcohol
   e. 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:
   a. Economists believe that the market solves all environmental problems.
   b. Economists always recommend market solutions to environmental problems.
   c. Economists always use market prices to evaluate environmental issues.
   d. 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?

CHAPTER

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 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 ($= 1 \div 5/100$) 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, $N_2, \ldots, N_t, \ldots$) 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.

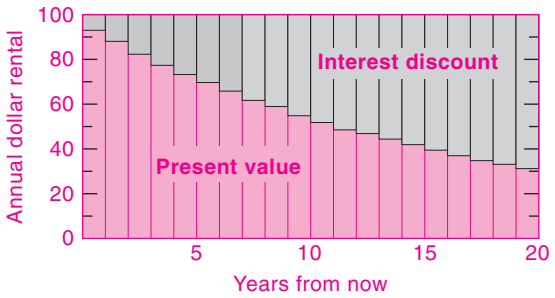

FIGURE 15-1. Present Value of an Asset

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

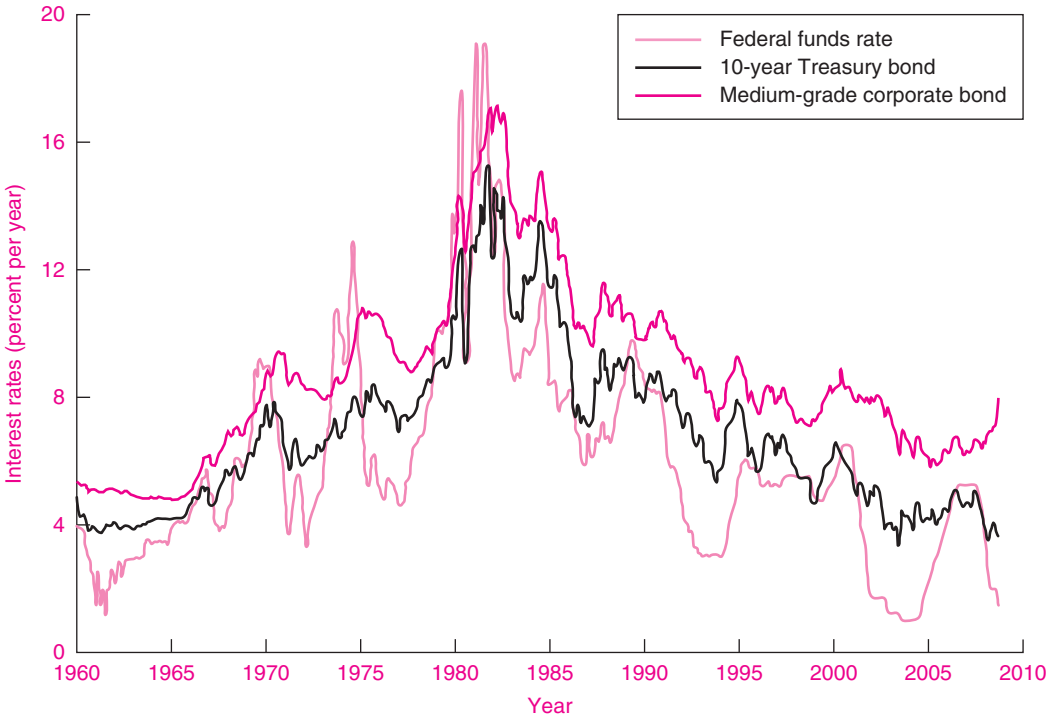

FIGURE 15-2. Most Interest Rates Move Together

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 4¼ percent 10-year inflation-protected bond. Between January 2000 and June 2003, the CPI increased by 12 percent. Therefore, the same

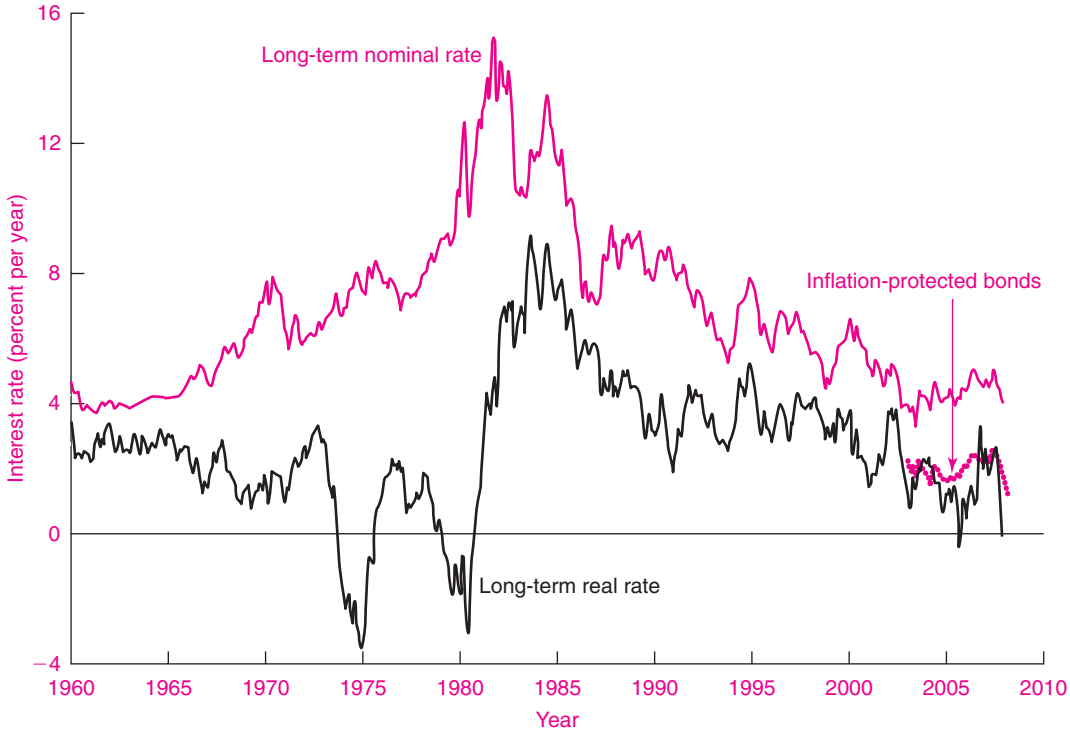

FIGURE 15-3. Nominal vs. Real Interest Rates

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 4¼ percent of $1120, instead of 4¼ 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 × $(1.3)^{10}$], 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

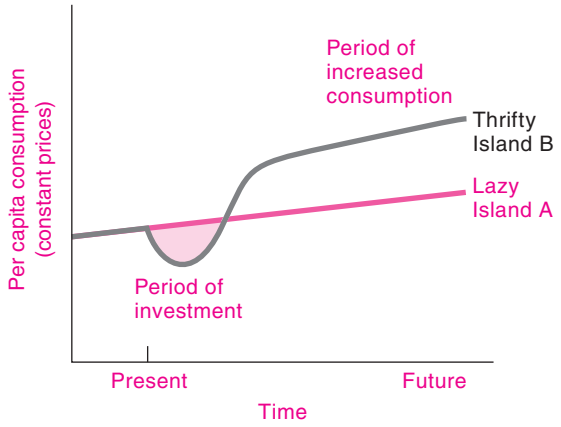

FIGURE 15-4. Investments Today Yield Consumption Tomorrow

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 idiosyncracies 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.

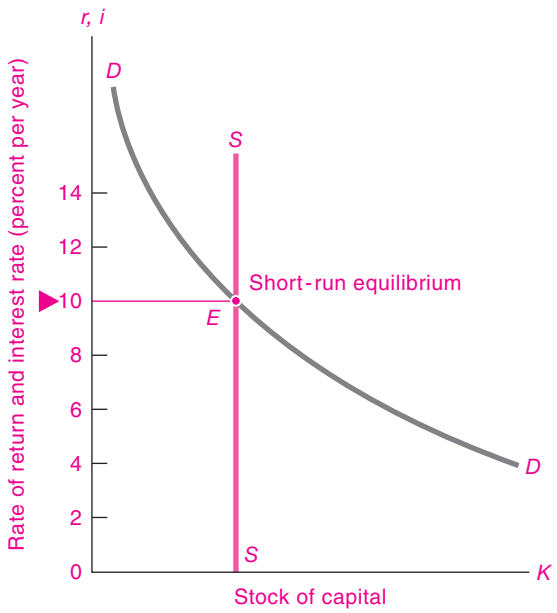

FIGURE 15-5. Short-Run Determination of Interest and Returns

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*.

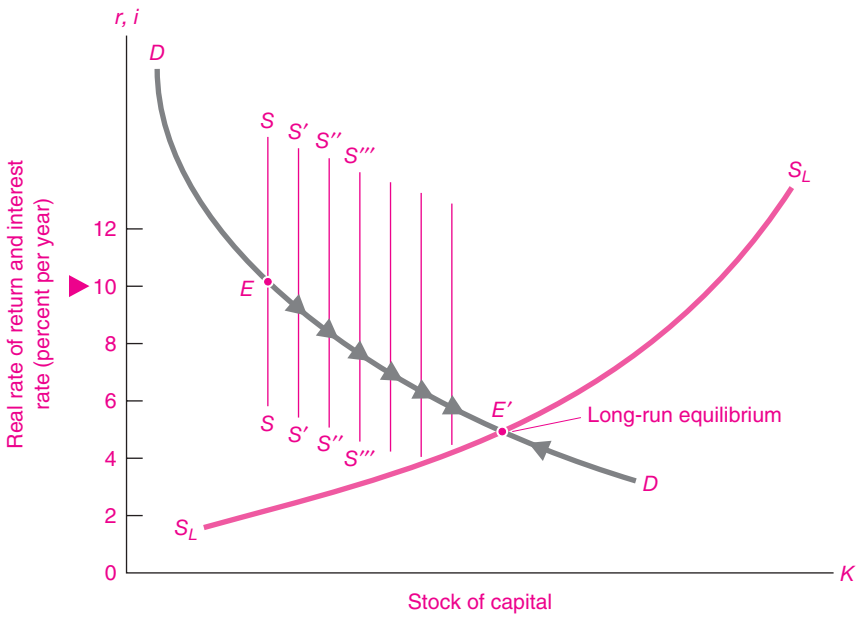

FIGURE 15-6. Long-Run Equilibration of the Supply and Demand for Capital

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 innovational 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.

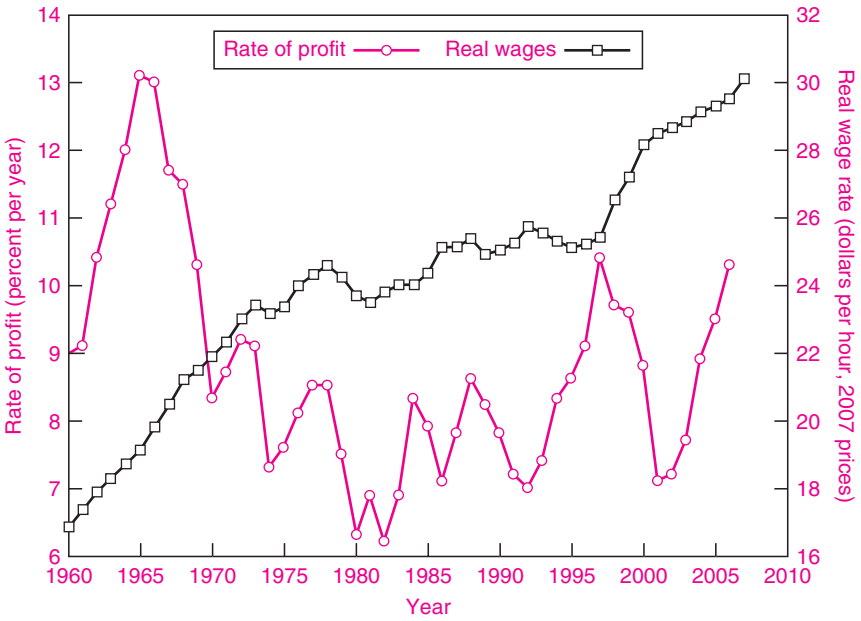

FIGURE 15-7. Trends in Wages and Rate of Profit in the United States

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 $(N_1, N_2, \ldots, N_t)$, 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. $I_0 = 0, I_1 = 12, I_2 = 12, I_3 = 12, \ldots; i = 5.$
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.
   b. What is the value of a perpetuity yielding $16 per year at interest rates of 1, 5, 10, and 20 percent per year?
   c. Compare the answers to **a** and **b.** Which asset is more sensitive to interest-rate changes? Quantify the difference.
5. 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)
6. 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.
7. 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?
8. Recall the algebraic formula for a convergent geometric progression:

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

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. A perpetuity yielding $100 per year
    b. 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.