

6 Aggregate Production

Up to now we have focused on the problem of how to describe and measure the economy. We now turn to the problem of how to explain it. We begin with aggregate supply. What determines how much GDP the firm sector wants to produce? What determines how productively the economy uses labor and physical inputs to production? How does the business cycle affect the use of inputs – especially the level of employment? This chapter sets the stage for Chapter 7, in which we examine economic growth in the long term, and Chapter 8 and 9, in which we examine labor markets.

In the last four chapters we learned what GDP and its components are, how they typically behave, and how they appear to be related to inflation, employment, industrial production and other variables. It is now time to turn from description to explanation? Examples of the kind of a questions that we would like to answer include:

- Why is GDP in the United States vastly higher in 2005 than in 1955 or 1905?
- Why is GDP in the United States higher than in any other country in the world?

Or, more specifically:

- What would happen to GDP if immigration rules were relaxed at 5 percent more workers joined the U.S. labor force?
- How much should we expect employment to fall in a recession or rise in a boom?

To begin to answer questions of this sort, recall from Chapter 2, section 2.7, that GDP can be seen through the eyes of producers (firms) as aggregate supply or through the eyes of purchasers (consumers, investors, government, foreigners) as aggregate demand. The national-income-accounting identities ensure that *ex post* aggregate supply

equals aggregate demand. But aggregate supply and aggregate demand are different things, and *ex ante* they need not be equal. Aggregate demand is governed by the choices of spenders, given their incomes. Aggregate supply is governed by the choices of firms, given the available productive resources. In the end, it is these productive resources that set the upper limit to GDP. Aggregate demand may fall short, and GDP may turn out to be less than it could be. But no matter how much aggregate demand rises, GDP cannot exceed the ability of the workers, factories, and offices to produce. In this and the following three chapters examine aggregate supply.

We know from our discussion of the circular flow of income and production (Chapter 2, section 2.2) that the firm sector purchases factors of production and transforms them into final goods and services. This is the heart of aggregate supply. The decisions that govern the production process are not made by the firm sector as a whole. Rather they are made by each individual firm. As a result, even though we are, in the end, concerned primarily with the aggregate product (GDP), looking at the production problem from the point of view of the firm is a valuable prelude to main business of the chapter. It will help us to build up our intuitions and to cement essential terminology in our minds.

6.1 The Production Decisions of Firms

6.1.1 PRODUCTION POSSIBILITIES

Technology

Consider a particular product, say, a barrel of gasoline. An oil refinery uses raw petroleum, energy, various chemicals, and various types of labor, among other **INPUTS**, to

produce its **OUTPUT** – gasoline. The design of the refinery and its work practices embody a particular **TECHNIQUE** for producing gasoline. A technique is like a recipe for a cake. The inputs are the list of the ingredients used, the output (or possibly outputs) is the cake, and the recipe is the set of instructions for making the cake. Like the recipe, the technique is not the machinery or any of the physical inputs but the *knowledge* of how to make a certain sort of machinery or how to organize production. Technical knowledge guides the construction of machinery, the layout of a plant, the organization of the corporation, and the practices on the shop floor. Yet the technique is not any of these tangible things. It is the knowledge itself.

The process for producing gasoline is represented schematically in Figure 6.1. Notice that the process is incompletely specified. A few of the inputs have been listed. They are divided into different classes: material inputs, energy, services, and labor. Within each class various sorts of inputs are listed, but the recipe for making gasoline is very complicated, and these only hint at the complete description, which may or may not be recorded anywhere, even in the offices of the oil company. And, of course, the actual quantities needed are not listed.

Techniques are not usually unique. The same gasoline might be made with different inputs. During the Second World War, when its crude petroleum supplies were cut off, Germany developed a technology for turning coal into gasoline and other petrochemicals. South Africa used similar technologies to thwart the anti-apartheid embargoes during the twenty years before majority rule was established. The United States investigated technologies for extracting petroleum products from shale found in

western states during the oil crises of the 1970s and early 1980s. **TECHNOLOGY** is *the set of available techniques*.

The Production Function

Which technique a firm chooses depends on the costs of the inputs, the productivity of the technique itself, and the selling price of the output. The techniques that produce gasoline from coal or shale are not usually cost effective. They have been adopted only in special circumstances in which the cost and availability of the crude petroleum input was prohibitive. But crude petroleum itself is not a single product; it comes in a variety of grades with different characteristics; and it takes a somewhat different recipe to convert each grade to gasoline. Which type of crude petroleum to buy and which recipe to use are key business decisions for oil companies, in which price and availability play an important part. California refineries, for instance, are specialized both in the type of crude that they process (North Slope crude from Alaska) and the type of gasoline they produce (a special formulation to meet California air pollution controls).

The producers themselves must have detailed understanding of the available technology, but this is much too complicated in most cases for the economist. Typically, economists work with a simplified picture of the production process. They summarize the wide variety of inputs in a few indicative categories – the **FACTORS OF PRODUCTION**. For example, all of the heterogeneous types of work that go into production are represented by a single category, **LABOR**, while all of the physical inputs are represented by another category, **capital services**. And the set of techniques (the technology) that governs the transformation of capital service and labor into output are represented

mathematically by a **PRODUCTION FUNCTION**. For example, the refining process is represented as

$$\text{gasoline} = f(\text{labor}, \text{capital services}).$$

We can think about production even more abstractly. Instead of gasoline, consider any output – call it y . (The variable name y is chosen to echo the variable name for GDP (Y), because GDP is just the sum of all the final production in the economy. We adopt a convention of writing variable names for the inputs and outputs of firms (microeconomic quantities) as lower-case letters, while writing the variable names for the analogous macroeconomic quantities as upper-case letters.) Similarly, instead of specialized labor and capital, think of general forms for labor and capital – call them l and k . A general production function can be written as

$$(6.1) \quad y = f(l, k).$$

Each point on the production function represents a particular technique: how a particular combination of capital and labor is transformed into a particular quantity of output. Not every technique is represented. If there are two techniques, each using the same quantities of inputs, but producing different quantities of output, the technique with the higher output *dominates* the technique with the lower output. No rational firm would ever choose a dominated technique. The production function represents the set of nondominated techniques.

Measurement Issues

In principle, the production function, $f(\cdot)$, can be quantified. But measurement poses some tricky problems. The variable *gasoline* is measured in gallons, but what are the right units for the variables *labor* or *capital*? For any one kind of labor, it is natural to use the number of workers or the number of worker-hours as the unit. But some kinds of labor are more efficient than others, and some kinds of labor differ on other dimensions. The work of an electrician is qualitatively different from the work of a truck driver.

Similarly, what unit of measurement is appropriate for the wide variety of different physical inputs – electricity, petroleum, pipes, buildings, roadways, paper clips, computers, ovens, and so forth? We faced this same problem (see Chapters 2 and 3) in adding up final goods and services to measure real GDP. The solution was to find a common denominator – money. The same solution is applied to capital services, which are most often measured in real (or constant) dollars.

There is another difficulty in the case of capital services. The output of a production process (for example, the gasoline) is a flow (so many gallons per unit time) and the inputs should also be flows. For labor this is relatively easy: number of workers per year or number of worker-hours per week. Some physical inputs are easily measured as flows. Electricity, for example, can be measured as kilowatt-hours per week. It is trickier for the flow of services from machinery or buildings or other long-lived capital goods. When a rental market exists, it may be easier – the rent on the floor space of an office measures the flow of property services to an accounting firm or the lease price of an airliner measures the flow of aircraft services to an airline. But it is much harder to

assess the exact flow of services from the machinery of an oil refinery or a hydroelectric dam as inputs to the production of gasoline or electricity.

Economists often take a shortcut to avoid the problem of measuring capital services. Instead of entering the contribution of physical inputs as a *flow* (capital services), they enter it as the *stock* of CAPITAL. Capital is typically measured in dollars without a time dimension. If the flow of capital services is roughly proportional to the stock of capital, then this shortcut provides a good approximation.

Like “investment,” “capital” is one of the tricky words in economics. In ordinary life and often in business, “capital” means an amount of money tied up in a firm or a property. In macroeconomics, however, “capital” refers not primarily to money but instead to *the physical means of production*. Even though it costs money to buy the inputs to gasoline production, one cannot refine gasoline using money in the process. It is easy to get confused on this point because we generally measure the amount of capital in monetary terms. In a simplified description of the production process, we refer to the total real monetary cost of the physical inputs. Still, we must not forget that it is the non-labor inputs (not the money in which we measure them) that do the work.

Basic Properties of the Production Function

The production function (equation (6.1)) cannot be graphed in two dimensions. Instead, in Figure 6.2 we use two related graphs. Look first at panel (A). It shows the relationship of output (y) to a varying labor input (l) on the assumption that the capital input (k) is held to a constant value k_0 . That is, it shows the relationship between output and labor *ceteris paribus* (a Latin phrase frequently used by economists meaning “other

things equal,” often abbreviated *cet. par.*). The constancy of capital in panel (A) is indicated by the bar over the capital variable (\bar{k}_0). Panel (B) of Figure 6.2 shows the same production function from a different point of view. The amount of output (y) is shown *ceteris paribus* (l held constant at \bar{l}_0) as an increasing function of capital (k). (For convenience, we refer to a graph of the production function with variable labor and constant capital as the *labor production function* and to one with variable capital and constant labor as the *capital production function*.)

Why is the production function drawn in just this way? It is mainly a reflection of some commonsense ideas about production that can be summed up in three basic properties:

Property 1: *If the inputs are zero, output is zero – that is, the production function goes through the origin (“there’s no such thing as a free lunch”).*

Property 2: *The production function is increasing in each of the inputs – that is, it slopes up. For a given amount of physical inputs more can be squeezed out of them using more labor. For a given amount of labor, more can be produced using additional capital.*

Property 3: *The production function displays **DIMINISHING RETURNS TO FACTORS OF PRODUCTION** – that is, each increase in any factor of production, holding other factors constant, raises output but at a slower rate than previous increases. The production function is drawn as bowed (or concave) toward the horizontal axes to reflect diminishing returns.*

A farm provides a classic illustration of these properties. Consider a farmer who has a fixed amount of land, tractors, fertilizer, seed, and other physical inputs. If neither

he nor his family nor any hired help work, there is no corn crop (the production function goes through zero). The more work they do, the bigger the crop will be. At first the returns are great. Each additional hour of work means more crop planted, fertilized, and tended, at least up to the point that the entire amount of seed and land have been used (the production function slopes up). But even after that, additional work in keeping down weeds, or more carefully monitoring the use of water, fertilizer, and pesticides produces some additional corn. But as more and more jobs are tended to, the jobs become less and less important to increasing the final harvest (the production function shows diminishing returns).

The story is exactly the same for capital. Consider the farm again, but this time imagine that every member of the family works a full day. If they use no land, tractors, seeds, fertilizer, and so forth, they get nothing. The more they plow and plant (the more tractors, plows, and seed they use), the more corn they grow at least until the entire farm is planted. Once the farm is planted, additional use of water, fertilizer, pesticides, tilling machines, and so forth add to the final crop. But as more and more needs are met, the remaining needs become less and less important.

Look again at Figure 6.2. Notice that at the input values (l_0, k_0) the curves in both panels take the value y_0 . Both curves represent the same production function, equation (6.1), from different points of view.

What would happen to the two curves if one of the factors of production changed? For example, in Figure 6.3, l_0 increases to l_1 . Since capital remains fixed at k_0 , this is a movement along the production function in panel (A) showing that output rises to y_1 . Notice that in panel (B), the point (k_0, y_1) is not a point on the original production

function. This is because that function was drawn on the condition that labor was fixed at l_0 , but now there is a larger labor input. In general, the same amount of capital with more labor produces more output, no matter how much capital there is, so the curve must shift up. In fact, at k_0 it must pass through y_1 , because this is the amount of output that the production function (equation (6.1)) assigns to (l_1, k_0) as is clear from panel (A). *A movement along one curve corresponds to a shift in the other curve.* Notice that the new curve in panel (B) is steeper, but essentially the same shape as the old curve – in particular, it still must go through the origin. The student should work out the consequences for the diagram of a change in the capital input holding the labor input constant (see Problems 6.1 and 6.2 at the end of the chapter).

Returns to Scale

The thought experiment behind diminishing returns to a factor of production holds one input constant while increasing the other. What happens if both inputs are allowed to vary? Production functions are classified by how output responds to equal proportional increases in inputs. There are three possibilities: *when all inputs increase by the same proportion*, a production shows

- **CONSTANT RETURNS TO SCALE** if output rises by the same proportion as the inputs;
- **Increasing returns to scale** if output rises by a greater proportion than the inputs; and
- **Decreasing returns to scale** if output rises by a smaller proportion than the inputs.

Figure 6.4 illustrates *constant returns to scale*. The production function passes through the initial input-output combinations – (l_0, y_0) and (k_0, y_0) . The **average product of labor** (abbreviated **apl**) is defined as *the number of units of output per unit of labor* (that is, y/l). Notice that the slope of a ray from the origin through (l_0, y_0) is y_0/l_0 . In other words, the slope of ray from the origin to a point of the labor production function is the average product of labor. Similarly, the **average product of capital** (abbreviated **apk**) is defined as *the number of units of output per unit of capital* (that is, y/k). And it is indicated on the capital production function as the slope of a ray from the origin to the production point.

When a production function displays constant returns to scale, the average products remain constant as the inputs are increased by the same proportion. In Figure 6.4, both labor and capital are increased by 20 percent ($l_1 = 1.2l_0$ and $k_1 = 1.2k_0$). Given the increase in labor, the capital production function must shift upwards; and given the increase in capital, the labor production function must also shift upwards. But how far? A 20 percent increase in output would leave the point of intersection of each curve vertically above l_1 and k_1 on the same rays from the origin, so that $y_1 = 1.2 y_0$.

Figure 6.5 illustrates *increasing returns to scale*. Again, each input is increased by 20 percent, but since output increases by more than 20 percent, each curve must shift upward so that the production points, (l_1, y_1) and (k_1, y_1) , which lie above the original rays from the origin. Rays from the origin to the new production points are steeper: increasing returns to scale raises the average product of each input. The student should analyze decreasing returns to scale using an analogous diagram (see Problem 6.3 at the end of this chapter).

It is important to remember that thought experiment behind returns to a factor of production allows only one factor at a time to vary, while that behind returns to scale allows all factors to vary proportionally. The production functions in Figures 6.4 and 6.5 both display diminishing returns to labor and capital, while the first displays constant returns, and the second increasing returns, to scale.

6.1.2 OPTIMAL PRODUCTION

The production function displays the technological possibilities open to a firm. How does a firm choose which technique to use? Firms, of course, wish to make profits, so the question reduces to how does the firm choose the level of output and the mix of labor and capital that generates the greatest profit?

Profit Maximization

Profit is defined as:

$$\text{Profits} = \text{Revenues} - \text{Costs}$$

What happens if a firm produces more and more units of output? More output means more goods to sell, so revenues rise. But of course, more output requires more inputs, so costs rise as well. The question must be, which rises faster?

To analyze the firm's profit-maximization problem more carefully, consider a firm producing with a fixed stock of capital. What happens if the firm produces one more unit of output? The extra cost of the labor (or of any additional unit of an input) is

known as the *marginal cost*. To economists, the word “marginal” means *having to do with a small additional increment*. So,

marginal cost (MC) is *the small increase in the cost of the factors of production attributable to a one unit increase in output*.

The firm sells the marginal output:

marginal revenue (MR) is *the small increase in revenue attributable to the sale of a one unit increase in output*.

It is worth increasing the labor input to the production process, so long as the marginal revenue exceeds the marginal cost ($MR > MC$), because whenever this is true, the additional input means additional profits to the firm.

Imagine that we start off with a very low level of labor (say, on the farm in the earlier example). Each additional unit of output produced and sold adds revenue as well as labor costs. Typically, the first unit produced adds more revenue than costs. But recall the principle of diminishing returns to a factor. Each additional unit of output requires increasingly greater additions of labor – that is, *marginal costs tend to increase as the level of production rises*.

When a market is large relative to the producer (as a wheat market is relative to a wheat farm), the decision of an individual firm to produce or not produce will have no discernible effect on the price at which it sells its output. Such a firm is called a **price**

taker. A market composed of such firms displays **PERFECT COMPETITION**. Since each additional unit of output earns the market price, the marginal revenue faced by a perfect competitor is constant.

When firms are large relative to their market, their output decisions affect market prices. (In order to sell more output, the firm must typically lower its prices.) Such firms are called **price setters**. A market composed of such firms displays **imperfect competition**. (Monopoly, duopoly, and oligopoly are forms of imperfect competition studied by microeconomists.) Typically, under imperfect competition each additional unit of output will sell for less than the previous unit, so the marginal revenue earned by the firm will fall.

Under either perfect or imperfect competition, as output is increased, the gap between marginal revenue and marginal cost narrows, until eventually they are exactly equal. At this point, a profit-maximizing firm should stop expanding output and stop using additional inputs, because to go further would put marginal costs above marginal revenues and begin to cut into profits. The rule then is:

- *a firm maximizes its profits when it increases its use of any factor of production ceteris paribus up to the point that marginal revenue equals marginal cost: $MR = MC$.*

The same maximization rule applies to all forms of competition. To extend the analysis, we have to be more specific about the market organization. Since perfect competition is easier to analyze than imperfect competition and, in most cases, it provides

a good approximation for macroeconomics, we focus on the profit-maximization problem for a perfectly competitive firm.

First, consider marginal revenue. A small increase in output can be indicated as Δy . Since the firm is a perfect competitor, it takes the market price of its output (p) as given beyond its control. Marginal revenue is, then, $MR = p\Delta y$.

Next, consider marginal cost. If capital is held constant, then the small additional amount of labor needed to produce the extra unit of output (Δy) can be indicated as Δl . If a firm takes the wage rate as given beyond its control at, say, w , then the marginal cost is $MC = w\Delta l$. The rule for profit maximization can now be reformulated:

- *a firm maximizes its profits when it increases its use of labor ceteris paribus up to the point that $MR = p\Delta y = w\Delta l = MC$.*

Marginal Products and Factor Prices

Look at the middle two terms in the maximization rule. Divide both by p and Δl to get

$$(6.2) \quad \Delta y / \Delta l = w / p.$$

This equation has a nice interpretation. The expression on the left-hand side, $\Delta y / \Delta l$, is the **MARGINAL PRODUCT OF LABOR (MPL)** defined as *the extra output produced ceteris paribus as the result of an additional unit of labor*. The expression on the right-hand side is the **REAL WAGE RATE (W/P)**. In this case, it is the value of the wage rate expressed in terms of the output it will buy. (If worker in a golf ball factory is paid \$12 per hour and

golf balls sell for 50¢ each, then $w/p = 24$ balls. (We are, of course, familiar with converting other nominal quantities into real equivalents (see Chapters 2 and 4). When we do this with a price index such as the CPI or the GDP deflator, we implicitly express a number of dollars as the number of units of a bundle of commodities that those dollars will buy. In the firm's profit maximization problem, it is not a bundle, but the single commodity produced by the labor that matters.) Another way of expressing the rule for profit maximization is then:

- *a firm maximizes its profits when it increases its use of labor ceteris paribus up to the point that the marginal product of labor equals the real wage rate; that is, up to the point that*

$$(6.3) \quad mpl = \Delta y / \Delta l = w/p.$$

We can illustrate the rule with data for the hypothetical golf-ball manufacturer (Table 6.1). The table shows that each additional worker adds to total production, but at a decreasing rate. The last two columns show the marginal products and marginal revenues for different numbers of workers. Suppose that the firm started with 99 workers. An additional worker (the 100th) has a marginal product of 28 balls per hour, which is greater than the real wage of 24 balls per hour. Hiring that worker adds to profits. Similarly, hiring the 101st worker adds to profits. Notice, however, that for the 101st worker, the marginal product (24 balls per hour) equals the real wage. Hiring the 102nd worker is not a good idea: the marginal product is 16 balls per hour, which is less than the real wage rate, so the firm's profits would fall. The firm's optimizing decision is

to hire 101 workers to produce 5050 balls per hour. The third, fourth, and fifth columns confirm this choice. Revenues rise with each additional worker – in fact, they rise even with the 102nd worker. But up to 101 workers, revenues rise faster than costs; after 101 workers, revenues rise less quickly than costs. Profits reach their peak at 101 workers.

We can relate the maximization rule to the graphical presentation of the production function. In Figure 6.6 a small increase in labor is indicated by Δl and the resulting increase in output by Δy . Look at the line that passes through points A and B on the production function. Its slope is the rise (Δy) over the run (Δl) or $\Delta y/\Delta l$, which is the marginal product of labor.

When we think, as economists often do, in terms of calculus, we imagine Δl getting smaller and smaller until it is infinitesimally small. It is easy to visualize the line through A and B getting steeper as Δl gets smaller, with less and less of the production function lying above it until, in the limit it just touches the production function at point A. In the calculus we indicate discrete changes by Δ and infinitesimal changes by d . Figure 6.7 shows the result of letting Δl become infinitesimally small. The line is now tangent to the production function at point A and the marginal product of labor is now expressed as the first derivative of the production function at point A: dy/dl .

This is a useful result: if we know the mathematical form of the production function, we can compute the marginal product of labor by taking the first derivative. The principle of diminishing marginal returns to labor can also be stated mathematically as the property that the marginal product of labor becomes smaller (dy/dl declines) as the labor input becomes larger (l increases). The rule for profit maximization can be restated:

- *a firm maximizes its profits when it increases its use of labor ceteris paribus up to the point that the marginal product of labor equals the real wage; that is, up to the point that*

$$(6.4) \quad mpl = dy/dl = df(l, \bar{k})/dl = w/p.$$

Notice that the mpl at the profit-maximizing point of production is always less than the average product of labor ($mpl < apl$), another implication of diminishing returns : adding one more unit of labor adds output, but by less than the average of all preceding units.

With appropriate modifications, the same principles apply to capital, holding the labor input constant. We must think of capital services as something for which there is a market price, even if the firm owns its own capital. If a construction company hires a crane for a project, the **rental rate** measures the price of the capital service. If, instead, the company owns the crane itself, it should still regard the price that it would earn if rented on the market (its **IMPLICIT RENTAL RATE**) as its price, since the services that the crane provides are same in either case. We designate the (implicit) rental rate of capital by the Greek letter ν (“nu”). The rule for a profit-maximizing firm is then:

- *a firm maximizes its profits when it increases its use of capital services ceteris paribus up to the point that THE MARGINAL PRODUCT OF CAPITAL (MPK) equals the real (implicit) rental rate. Mathematically, the profit-maximizing point occurs when*

$$(6.5) \quad mpk = \Delta y / \Delta k = v / p$$

or, for infinitesimal changes,

$$(6.6) \quad mpk = dy / dk = df(\bar{l}, k) / dk = v / p .$$

Analogous to the case of labor, the marginal product of capital is always less than the average product of capital ($mpk < apk$) due to diminishing returns to capital.

Choosing Input Levels

How should a firm use the rules for profit maximization to choose the best mix of capital and labor? First, it must compute the real wage and the real (implicit) rental rate of capital services. Then it must find the combination of labor and capital that simultaneously sets the marginal product of labor equal to the real wage rate and the marginal product of capital equal to the real (implicit) rental rate. If we have a mathematical representation of the production function, this is easy to do.

Diagrammatically, it is a little tricky, as it would involve a lot of trial and error. Figure 6.8, however, shows the end result in which the firm has chosen the best mix of labor and capital (l_0, k_0) given the wage rate (w_0), (implicit) rental rate (v_0), and product price (p_0) that it faces.

As the prices of factors of production change, the optimal mix of capital and labor also changes. Imagine that the wage rate falls because of an increase in the number of people willing to work, either from a natural boom in population twenty years before or increased immigration or increased participation in the labor force of teenagers or

women. In that case labor would be cheaper to the firm, and the firm would use more of it. This is easy to see on Figure 6.9. Given the wage rate w_0 , the firm initially hires labor of l_0 . The real wage (w_0/p_0) is equal to the marginal product of labor (mpl_0) shown by the slope of the tangent line at point A on the production function. When the wage falls to w_1 , the real wage also falls to w_1/p_0 . The lower real wage must be set equal to a lower marginal product of labor. Because of diminishing returns to labor, the lower marginal product of labor must correspond to a higher total use of labor at a point to the right of A on the production function. This is shown by the flatter tangent line at point B where $w_1/p_0 = mpl_1$ and where the firm hires labor of l_1 .

6.2 Aggregate Supply

6.2.1 THE AGGREGATE PRODUCTION FUNCTION

The individual firm is not the main concern of macroeconomics. We want to explain aggregate real GDP, not particular products such as gasoline, corn, or golf balls. Yet, the microeconomic prelude of the last section provided valuable insights. If there is a relationship between individual inputs and output, there should also be a relationship between aggregate inputs (aggregate capital and aggregate labor) and aggregate output (real GDP). And we may be able to exploit important analogies between the more easily understood behavior of firms and the behavior of aggregate production.

We write the macroeconomic or **AGGREGATE PRODUCTION FUNCTION** by analogy to the microeconomic production function as

$$(6.7) \quad Y = F(L, K)$$

Upper-case letters are used to show aggregates and lower-case letters are used to show micro-level quantities. The micro-quantities y , l , and k were placeholders for the particular output, labor and capital of any firm we wanted to analyze. They were general algebraic variables that could be filled in for special cases. The output variable y , for example, could represent cars or bananas or blue-jeans. Once we aggregate to the level of the whole economy, however, this is no longer true. Y is not a placeholder that can be filled in different ways; it is real GDP. Similarly, L and K are the labor and the capital inputs, not for particular firms, but for the whole economy.

Most of the properties of the microeconomic production functions were derived from a commonsense understanding of production processes and the behavior of firms. Many of these obviously carry over to the aggregate production function. The aggregate production function must pass through zero, because no GDP can be produced without inputs. And, it should slope up, because more inputs generate more output. But are there aggregate diminishing returns? Would the properties of the aggregate production function, derived from adding up the micro production functions for the whole economy, result in a curve with other predictable properties and shape?

It turns out that these questions are extremely difficult. They belong to an area of economics known as *aggregation theory*, and they are well beyond the scope of this book. Instead of trying to figure out what the properties of the aggregate production function are from first principles, our strategy will be to start with a conjecture that the economy can be described by a particular production function, one that shares important properties with microeconomic production functions. We will then test our conjecture

empirically. If it seems to describe the data well, we shall be satisfied that it provides a useful approximation.

6.2.2 THE COBB-DOUGLAS PRODUCTION FUNCTION

Many mathematical functions have the properties we expect from a production function. A particularly useful one is the Cobb-Douglas production function. It was popularized in economics by Paul Douglas (1892-1976), a labor economist who later became a prominent U.S. Senator, and Charles Cobb, a mathematician and Douglas's colleague at Amherst College in the 1920s. We shall first describe the Cobb-Douglas production function and some of its properties and then explain why it is a good one for elementary macroeconomic analysis.

The **COBB-DOUGLAS PRODUCTION FUNCTION** can be written:

$$(6.8) \quad Y = F(L, K) = AL^\alpha K^{1-\alpha}.$$

A is a positive constant that can be interpreted as an index of the state of technology. It is often referred to as a measure of **TOTAL FACTOR (OR MULTIFACTOR) PRODUCTIVITY (TFP)**: for the same labor and capital inputs, higher A implies higher GDP. The exponent α is a constant greater than zero and less than unity ($0 < \alpha < 1$). As we shall see shortly, it can be interpreted as labor's share of GDP, and its value will depend on the actual value of the labor share in the economy.

On a graph, the Cobb-Douglas production function has the same shape as the micro production function in Figure 6.2. Box 6.1 provides a numerical example of some

of the useful properties of the Cobb-Douglas production function. **[Insert Box 6.1 near here.]** These include all three properties mentioned in section 6.1.1:

- **Property 1.** It is obvious from a glance at equation (6.8) that if both K and L are zero, Y is also zero – there is no “free lunch.”
- **Property 2.** Similarly, it is obvious that if either K or L increases, Y increases.
- **Property 3.** To see that the Cobb-Douglas production function displays diminishing returns to factors of production, we must look at the marginal products. Each marginal product can be written in at least two equivalent forms (see the *Guide*, section G.16.1, for the derivation):

$$(6.9) \quad mpL = \alpha \left(\frac{\bar{K}}{L} \right)^{1-\alpha} = \frac{\alpha Y}{L};$$

$$(6.10) \quad mpK = (1-\alpha) \left(\frac{\bar{L}}{K} \right)^{1-\alpha} = \frac{(1-\alpha)Y}{K}.$$

(Note that we write the marginal products with upper-case L and K as a reminder that we are working with aggregates.) Looking at the middle term in equation (6.9), it is obvious that the marginal product of labor gets smaller as L becomes larger. Equation (6.10) shows that the same is true for the marginal product of capital as K becomes larger.

In addition, the Cobb-Douglas production function displays three other important properties:

- **Property 4.** An increase in one factor of production raises the marginal product of the other factor of production. For example, raising \bar{K} , in the middle term of

- (6.9) raises mpL . Similarly, raising \bar{L} in (6.10) raises mpK . (Graphically, when a Cobb-Douglas production function shifts up – while always going through the origin – the slope of every point must be higher than the slope a point vertically below it on the original curve.)
- **Property 5.** An increase in the level of total factor productivity (A) raises both marginal products. An increase in A raises Y , which, looking at the right-hand terms of (6.9) and (6.10), must increase mpL and mpK .
 - **Property 6.** The Cobb-Douglas production function displays constant returns to scale. (See Box 6.1 for a numerical example and the *Guide*, section G.16.1 for a proof.)

6.2.3 DOES THE COBB-DOUGLAS PRODUCTION FUNCTION PROVIDE A GOOD MODEL OF AGGREGATE SUPPLY?

What is the evidence that the Cobb-Douglas production function describes the economy of the United States? Any empirically acceptable production function must clearly fulfill properties 1 and 2. While properties 3-6 appeal to our intuitions gained from thinking about the production functions of individual firms, it is not certain that they carry over to the aggregate economy. To provide some evidence that the Cobb-Douglas production function is a good description, we shall put it to a test. We derive an important joint implication of the Cobb-Douglas production function and the assumption that firms follow the maximization rules of perfect competition (see section 6.1.2). The test proceeds in two steps: first, we show that our assumptions imply that the labor and capital shares in GDP are constant; then, we show, to a good approximation, that those shares are constant.

The Cobb-Douglas Production Function Predicts Constant Factor Shares

Real labor income (Y_L) is the real wage rate times the aggregate number of hours worked:

$$Y_L \equiv (w/p)L.$$

The **labor share** in real GDP is defined as

$$(6.11) \quad L\text{-share} \equiv Y_L/Y \equiv (w/p)L/Y.$$

Similarly, the **capital share** is defined as

$$(6.12) \quad K\text{-share} \equiv Y_K/Y \equiv (v/p)K/Y.$$

To move beyond these definitions, we use the assumption that the economy is approximately perfectly competitive. This allows us to use the profit-maximization rules from equations (6.3) and (6.5), replacing the lower-case y , l , and k with their macroeconomic analogues (upper-case Y , L , and K). The rules state that each real factor price is equal to the corresponding marginal product. Substituting the marginal products for the factor prices in (6.11) and (6.12) yields:

$$(6.13) \quad L\text{-share} = Y_L/Y = (mp_L)L/Y$$

and

$$(6.14) \quad K\text{-share} = Y_K/Y = (mpK)K/Y.$$

On the assumption that GDP is produced according to a Cobb-Douglas production function, we can substitute the specific form of the mpL from (6.9) into (6.13) and for the mpK from (6.10) into (6.14) to yield:

$$(6.15) \quad L\text{-share} = Y_L/Y = (\alpha Y/L)L/Y = \alpha$$

and

$$(6.16) \quad K\text{-share} = Y_K/Y = [(1 - \alpha)Y/K]K/Y = 1 - \alpha.$$

Since α is a constant, these equations show that the labor and capital shares are constant and equal to the exponents on L and K in the production function. (Henceforth, we shall refer to α as the *labor share of the Cobb-Douglas production function* and to $1 - \alpha$ as the *capital share*.)

Equations (6.15) and (6.16) demonstrate a key prediction: *if the economy is approximately perfectly competitive and if a Cobb-Douglas production function*

approximately characterizes aggregate production, then the shares of labor and capital in real GDP should be approximately constant. Are they?

Are Factor Shares Constant?

Our Cobb-Douglas model assumes that GDP is paid exclusively to owners of labor and owners of capital. The actual national accounts provide a more complex picture of the distribution of factor incomes (see Chapter 3, section 3.7). Table 3.7 and Figure 3.5 divide income into a number of categories. Of the major categories, it is reasonable to classify profits, rents, and interest as payments to capital. Similarly, compensation of employees is a payment to labor. National income itself differs from GDP by certain business-related items including consumption of fixed capital, which are reasonably assigned to capital.

Two items are harder to assign: proprietors' income and taxes on production (sometimes called "indirect business taxation"). Recall that proprietors' income accrues to the owners of small businesses, who supply both capital and labor. Similarly, some part of taxes on production falls on profits and some part is passed on to workers in the form of lower wages. To distribute the payments in these two categories, we assume that labor receives the same proportion of them as it does of other income (see the *Guide*, section G.16.2, for the exact formula).

Figure 6.10 plots the values of these shares for the United States in the post-World War II period.¹ The striking fact is that, while there is some variation, the

¹ Students who did problem 3.3 will already have seen that employee compensation as a share of GDP has no discernible trend in this period. But employee compensation does not capture all payments to labor.

variation is small, and there is no trend. The approximate constancy of the labor share confirms the prediction of our model and provides a good reason to take the Cobb-Douglas production function as a reasonable approximation of aggregate supply in the United States economy.

The average labor share for the postwar period is $\alpha = 0.69$. Throughout this book we will use this value of the labor share as the key parameter in the Cobb-Douglas production function.

6.2.4 A COBB-DOUGLAS PRODUCTION FUNCTION FOR THE UNITED STATES

For any quarter or year, we can formulate a Cobb-Douglas production function for the United States using readily available data on labor and capital, setting the labor share to its average value. The one unknown is the value of total factor productivity (A), which is not directly observable. One way to estimate A is set it to whatever level makes the two sides of the Cobb-Douglas production function equal. Using equation (6.8),

$$(6.17) \quad A = \frac{Y}{L^\alpha K^{1-\alpha}}.$$

For the U.S. economy in 1998, equation (6.17) can be filled in to yield

$$(6.17') \quad A = \frac{\$7,552 \text{ billion}}{(236,882 \text{ million worker - hours per year})^{0.69} (\$19,317 \text{ billion})^{0.31}}$$
$$= 8.15.$$

So that equation (6.8) can be written as

$$(6.18) \quad Y = 8.15L^{0.69}K^{0.31}.$$

Equation (6.18) can be used to answer questions such as the one in the introduction to this chapter: what would GDP in 1998 have been if the United States had permitted immigration that expanded the workforce by 5 percent but made no additional investment?

Answer: If the new workers worked similar hours to existing workers, the labor input would grow by 5 percent to 248,726 million worker-hours per year. Then, GDP would be:

$$Y = 8.15(248,726 \text{ million})^{0.69}(\$19,317 \text{ billion})^{0.31} = \$7,813 \text{ billion},$$

an increase of 3.5 percent. In the rest of this chapter and in future chapters, we shall use quantified versions of the Cobb-Douglas production function to answer many other questions about the economy.

6.3 Productivity

Another thought experiment using the Cobb-Douglas production function highlights the importance of technology to aggregate supply. What would happen to GDP if technological improvements raised total factor productivity 5 percent?

Answer: A 5 percent increase in total factor productivity multiplies A by 1.05.

Looking at equation (6.8) that would obviously raise Y by a factor of 1.05 as well. GDP in 1998 data, for example, would then rise by 5 percent from \$7,552 billion to \$7,930 billion.

Notice that an increase in total factor productivity is much more effective point for point in raising GDP than an increase in labor, for which a 5 percent increase resulted in only a 3.5 percent increase in GDP. Though numerically powerful, total factor productivity as represented by A does little to reveal the mechanisms of technological progress. And other measures of productivity are more commonly cited in economic and political discussions. Claims that productivity slowed markedly in 1970s compared the period immediately after World War II or that it rose considerably in the late 1990s are usually supported by data on labor productivity rather than total factor productivity. In this section, we will examine productivity and the relationships among the various measures in more detail.

6.3.1 ALTERNATIVE MEASURES OF PRODUCTIVITY

Three Measures

Productivity measures try to answer the question, how much output results from a particular level of input. All take the form $productivity = output/input$. Two measures correspond to the two main factors of production:

- **LABOR PRODUCTIVITY** is the amount of GDP per unit of labor. It is the macroeconomic equivalent of the average product of labor. We indicate it by the Greek letter “theta”: $\theta = Y/L = apL$.
- **CAPITAL PRODUCTIVITY** is the amount of GDP per unit of capital. It is the macroeconomic equivalent of the average product of capital. We indicate it by the Greek letter “phi”: $\phi = Y/K = apK$.
- **TOTAL FACTOR PRODUCTIVITY** is, as we have already seen (equation (6.17)) the amount of GDP per units of a bundle of capital labor: $A = \frac{Y}{L^\alpha K^{1-\alpha}}$. The numerator in this expression can be regarded as the weighted geometric average of labor and capital (see the *Guide*, section G.4.2).

The units of capital productivity are straightforward: dollars of GDP per dollar of capital. The units of labor productivity depend how labor is measured. If labor is measured in hours of work, then labor productivity is measured as dollars of GDP per hour. If labor is measured in number of workers, then labor productivity is measured as dollars of GDP per worker. If the average hours per worker changes (for example, if overtime work increases), then labor productivity per worker may rise, even though productivity per hour remains constant.

The units of total factor productivity are not very intuitive and depend on the units in which the variables from which it is calculated are measured – particularly, whether labor is measured in hours or workers. As we can see in (6.16), A takes whatever value is needed to make the two sides of the equation equal.

Although we can easily calculate productivity in dollars of real GDP per unit of input, the official productivity data of the Bureau of Labor Statistics (BLS) are presented as index numbers with a base of 1996 or 1997 = 100, depending on the series. The BLS does not compute the overall productivity data as we have. Instead, it computes indices for various subdivisions of the economy: the business sector, the manufacturing sector (overall, durable and nondurable), nonfinancial corporations, nonfarm business sector, private businesses, private nonfarm business, as well as for a small group of particular industries. The coverage among these categories is incomplete and overlapping. In particular, productivity measures for the ever more important service and financial sectors are not widely available.

One reason that productivity data are not complete is that they pose difficult measurement issues. The output measure that matters for any industry is the contribution to *final* goods and services. The BLS adjusts the data for interindustry trade to account for that part of the product that is an input into other production. But this is tricky. It is also often difficult to measure the real output of an industry at all. We saw in Chapter 3, section 3.5.2, that the output of government is generally measured by the inputs (wages paid and goods purchased) rather than by the value of what it produces. The reason, as we noted previously, is that final government services do not generally have market prices. Another important reason is it is hard to measure the intangible services that do

not result in storable, easily countable physical products. The same problem applies to many service outputs in industry as well. Some economists have speculated that the reason that productivity measures show a slowdown in the late 1960s and early in 1970s is that the economy was becoming more dominated by the hard-to-measure service industries. The slowdown might not have been genuine at all, but rather the result of poor measurement.

International Comparisons

Which countries are the most productive? Since capital stock data is generally more poorly developed, most cross-country comparisons rely on labor productivity rather than capital productivity or total factor productivity. Comparisons for 2001 among the G-7 and selected other countries for labor productivity measured both as GDP per worker and GDP per hour are shown in Figure 6.11. Data are reported as a percentage of labor productivity in the United States. For the most part, rankings based on the two measures agree, but the Netherlands, Norway, France, and Belgium all show productivity levels per hour above those in the United States, even while their productivity per worker remains substantially lower.

How are the Productivity Measures Related?

It is easy to see the relationship among the three productivity measures. Notice that $Y = Y^\alpha Y^{1-\alpha}$, for any α . Substituting this fact into equation (6.17), rearranging, and using the definitions of θ and ϕ yields:

$$(6.19) \quad A = \frac{Y^\alpha Y^{1-\alpha}}{L^\alpha K^{1-\alpha}} = \left(\frac{Y}{L}\right)^\alpha \left(\frac{Y}{K}\right)^{1-\alpha} = \theta^\alpha \phi^{1-\alpha}.$$

In words: total factor productivity is a weighted geometric average of labor productivity and capital productivity, where the weights are the labor and capital shares in GDP.

Total factor productivity cannot change unless at least one of labor or capital productivity changes. However, either measure of individual factor productivity could change without changing total factor productivity, provided the other factor productivity changed sufficiently in the opposite direction.

Diagrams may clarify the point. Figure 6.12 presents the production function in equation (6.8) graphically. Let us revisit the thought experiment of a 5 percent increase in labor *ceteris paribus* (see section 6.2.4). Initially, the economy uses inputs L_0 and K_0 to produce Y_0 . Labor productivity (θ_0) is shown as the slope of the ray from the origin to the point (L_0, Y_0) . Similarly, capital productivity (ϕ_0) is shown as the slope of the ray from the origin to (K_0, Y_0) . The increase of labor to L_1 raises GDP to Y_1 – less than proportionally because of diminishing returns to labor. Clearly, labor productivity falls to θ_1 . What happens to capital productivity? The capital production function must shift upwards, since it was drawn on the assumption that labor was at L_0 and now labor is higher. Since capital has not changed, to be compatible with the labor production function, the curve must pass through (K_0, Y_1) . At this point, capital productivity has risen to ϕ_1 .

The thought experiment illustrates a general rule:

When one factor of production is used more intensely ceteris paribus, its own productivity falls and the productivity of the other factors of production rise.

Notice also that, since the capital production function is steeper, its marginal product has also risen – see property 4 (section 6.2.2 and Box 6.1).

What happens to total factor productivity? . . . Nothing. Figure 6.12 graphed equation (6.8) and the thought experiment assumed that other things were equal, including A constant. This means that the rise in capital productivity was exactly offset by the fall in labor productivity, keeping A constant.

We can distinguish changes in factors of production from changes in technology with a graphical analysis of the thought experiment that heads section 6.3: the effect of a 5 percent increase in technology (A) on GDP. In Figure 6.13, the economy initially uses inputs L_0 and K_0 to produce Y_0 . Labor productivity is θ_0 and capital productivity, ϕ_0 . The improvement in technology means that more GDP (Y_1) can be produced for the same level of inputs. The labor production function must, therefore, shift upwards to pass through (L_0, Y_1) . Labor productivity rises to θ_1 . Similarly, the capital production function shifts upwards to pass through (K_0, Y_1) , and capital productivity rises to ϕ_1 .

Again, the thought experiment illustrates a general rule:

When technology improves ceteris paribus, the productivity of all factors of production rises.

Notice also that, since both the labor and capital production functions are steeper, the marginal products of both factors of production have also risen – see property 5 (section 6.2.2 and Box 6.1).

6.3.2 TECHNOLOGICAL PROGRESS

Factor Productivity Over Time

One lesson of the last section is that any change in the relative use of labor and capital will alter factor productivities, even when technology is constant. Since relative use of labor and capital fluctuates with the business cycle, we should also expect factor productivity to fluctuate cyclically – which Figure 6.14 shows they in fact do. At the end of the 1990s policymakers in the Clinton administration and the Federal Reserve debated endlessly whether sharp increases in productivity were permanent, secular changes or were the unsustainable result of using capital and labor too intensely. The one would have been cause for celebration, the other a rationale for restrictive monetary and fiscal policy. How can we tell the difference?

The thought experiments analyzed the last section, might be thought to suggest a test: a rise in total factor productivity is attributable to technology (as in Figure 6.13) and other changes are cyclical (Figure 6.12). But that does not work. Except in the cases in which changes in the relative use of labor and capital lead to exactly offsetting changes in factor productivities, equation (6.19) shows that any change in factor productivities also changes total factor productivity. Figure 6.13 shows that an increase in A *ceteris paribus* must increase θ and ϕ . But we cannot demonstrate the converse: that only if both θ and ϕ increase can A increase – sometimes other things are not equal.

Indeed, Figure 6.14 shows that labor productivity more than doubled (from \$14 per worker-hour in 1948 to \$32 per worker-hour in 1998), capital productivity showed virtually no secular growth. (The slight upward trend in capital productivity is less than 1/3 percent per year – at that rate capital productivity would take more than 200 years to double. In contrast, the trend in labor productivity is 1.47 percent per year, leading to doubling in a little more than 45 years.) While both series are variable, we will not go very far wrong in thinking about permanent technological change to regard capital productivity as constant and labor productivity as growing strongly on trend.

Factor-augmenting Technological Progress

According to (6.19), total factor productivity must grow to reflect the growth in labor productivity. But how can we reconcile the rise in A with the long-term constancy of ϕ in light of the thought experiment analyzed in Figure 6.13? Unlike the assumption of the thought experiment, capital in the United States did not stay constant, but in fact grew at a rate at which diminishing returns just offset the tendency of improving technology to increase capital productivity. The data suggest that technological progress does not affect factors of production equally.

Let us analyze this somewhat further. Technology is said to be **FACTOR-AUGMENTING** when technological progress has exactly the same effect on GDP as having more of the factor input. For example, a 2 percent increase in labor productivity, holding L constant, would have exactly the same on GDP as a 2 percent increase in the labor input, holding θ , constant. More specifically,

- Technology is said to be **labor-augmenting** when *an increase in technology is exactly the same as a proportional increase in the labor input.*
- Technology is said to be **capital-augmenting** when *an increase in technology is exactly the same as a proportional increase in the capital input.*

Suppose that all technology is factor-augmenting. Then we can rewrite the Cobb-Douglas production function as

$$(6.20) \quad Y = (A_L L)^\alpha (A_K K)^{1-\alpha},$$

where A_L measures labor-augmenting technological progress and A_K measures capital-augmenting technological progress. Rearranging (6.20), we get

$$(6.20') \quad Y = (A_L^\alpha A_K^{1-\alpha}) L^\alpha K^{1-\alpha}.$$

Comparing (6.20') to (6.8), it is clear that

$$(6.21) \quad A = A_L^\alpha A_K^{1-\alpha}.$$

In words, total factor productivity is the geometrically weighted average of labor-augmenting and capital-augmenting technological progress.

How does (6.21) related to (6.19)? Where labor and capital productivity may change because of changes in the relative use of the factors of production (as in Figure 6.12), A_L and A_K change only as a result of technological change. Over long periods,

when cyclical fluctuations cancel out, the average growth of labor productivity should equal the average rate of labor-augmenting technological progress ($\bar{\theta} \approx \bar{\hat{A}}_L$), and similarly for capital productivity and capital-augmenting technological progress ($\bar{\phi} \approx \bar{\hat{A}}_K$). The data for the United States show that the average growth of capital productivity is zero, so that $\bar{\hat{A}}_K \approx \bar{\phi} \approx 0$.

6.4 Short-run and Long-run Aggregate Supply

6.4.1 FLEXIBLE AND INFLEXIBLE PRODUCTION FUNCTIONS

Each point on the production function represents a different technique, using a different mix of labor and capital. Up to now we have assumed implicitly that firms were able to choose the technique and the level of production that maximized their profits for given real factor prices (w/p and v/p). We assumed that demand for the products of individual firms would always be sufficient to purchase whatever they found profitable to produce and that aggregate demand would always be sufficient to purchase planned or *ex ante* aggregate supply (see Chapter 2, section 2.7). We now consider what might happen if plans to purchase ($C + I + G + NX$) fall short of plans to produce.

In reality, falling aggregate demand is likely to be associated with changing relative prices – both of goods and of factors of production. For example, the oil shocks of the 1970s and 1980s, echoed by high oil prices in 2004, increased the value of imports thereby lowering net exports and aggregate demand. Firms had to adjust both to lower sales and to the higher prices of oil products. Figure 6.15 shows the labor production function. The economy uses L_0 (and capital, not shown) optimally to produce Y_0 . Now

suppose that an oil price increase reduced aggregate demand to Y_1 . It would be silly for firms to keep producing at the old level and building up stocks of unsold goods. The efficient course would be to move down the production function from point A to B. In addition, they would want to reduce their use of oil, and substitute the other now relatively cheaper energy inputs (such as coal or natural gas), as well as capital and labor for oil. This represents a change in technique. If technology were perfectly flexible, then this would be quick and easy.

Some capital is— like a child's Lego blocks — quite flexible. A power drill, a screwdriver, a pickup truck, or a personal computer are easily turned from one task to another. But other capital is highly inflexible — for example, a hydroelectric dam. Other capital may fall in between — for example, an oil-fired power plant might be converted with some time and effort to use coal. Similarly, labor inputs may not be flexible: an airliner designed with a flight crew of three will not be easily run by two when wage rates are relatively high and will not benefit from a third member when they are relatively low. An airline facing higher wage bills might demand planes designed for smaller crews as it adds to or replaces its existing fleet, but that too takes time.

And that is the essential point: neither technology nor the organization of production can turn on a dime. When demand falls or relative prices change, it is not possible to immediately move down the production function to a new optimum. The production function should be regarded as the long-run production possibilities for the economy. Here **LONG RUN** refers to a planning horizon — a time sufficiently far in the future that a firm can regard its choice of technique as open and not fettered by past decisions about capital and organization. Like the horizon, the long-run always recedes

as we approach it. In the face of higher oil prices, a trucking firm may decide that a more fuel-efficient fleet is optimal. In the **SHORT RUN**, the here and now, it begins to scrap its oldest trucks and makes sure that new trucks are of the fuel-efficient type. If nothing else changed it would eventually converge to its new preferred fleet. But of course, before it is already there, other things are likely to change. In the short run, it must always make decisions aimed at the long run. But the long run is moving target.

6.4.2 PRODUCTIVITY AND RESOURCE USE IN THE SHORT RUN

What difference does it make to aggregate supply if technology is inflexible in the short run. Consider an extreme case, in which production techniques are absolutely fixed and no adjustments are possible. In Figure 6.15, the economy would have to keep exactly the same technique with exactly the same labor productivity (shown as the ray with slope θ^*). It could operate this technique at a lower level, but it could not choose another technique with another labor productivity. In effect, the segment of the productivity ray from the origin to point A (shown as a solid line) would become the short-run production function.

What happens then if aggregate demand falls to Y_1 ? In the extreme short-run a firms may simply hold onto their labor and produce at the lower level – producing at point C, which is off both the short- and long-run production functions. This is known as **LABOR HOARDING**. It might happen if the fall in demand is expected to be short-lived and if the costs of firing, rehiring, and training workers is substantial. Actual labor productivity would fall from θ^* to θ_1 . Productivity falls not because firms have not adopted new techniques, but because some workers are idle. The degree of labor

hoarding is measured by $LF - L_1$, where L_1 is the amount of labor truly needed to produce Y_1 using existing techniques.

If the fall is not expected to be short-lived, then firms would reduce their costs by laying off redundant workers. They would then produce at point D. Here labor productivity remains at θ^* but some workers are unemployed. If labor is measured by the number of workers, then $LF - L_1$ now represents **UNEMPLOYMENT**, since L_1 is now the amount of labor actually employed.

In reality, the economy would not remain at a point like D. If demand recovered quickly it might move back to point A on the production function and full employment. If demand remained low, firms would begin the process of adjusting to the new circumstances. For example, with demand at Y_1 , point B on the production function, which uses labor of L_2 , represents efficient techniques. Labor productivity is higher at point B (θ_2). But point B still represents a mismatch between the supply plans of firms and aggregate demand. Ideally, firms would find the new point on the production function that maximizes their profits in the face of new factor prices.

The situation with respect to capital in the face of the same fall in aggregate demand is a little different. Figure 6.16 shows the capital production function. Initially, the economy produces Y_0 optimally using \bar{K} (and labor). Again, if technique is fixed in the short run, then we can regard the segment of the ray from the origin to point A with a capital productivity of ϕ^* as the short-run production function.

Capital is not like labor – firms own it or lease it from other firms. Labor can be laid off – the resulting unemployment is a problem for the worker, but not for the firm. Capital cannot be laid off by the firm sector as a whole. It can be scrapped, but that

would not be an economically sensible reaction in most cases, as it will be needed in future. Only the least productive capital is likely to be permanently scrapped when aggregate demand falls temporarily. Instead, what is not needed will sit idle. (When air travel fell dramatically after September 11th 2001, airlines and leasing agencies mothballed hundreds of aircraft in the deserts of California and Arizona. As demand picked up again, the idled planes were brought back into service.) In Figure 6.16, only K_1 units of capital are needed to produce Y_1 . But since \bar{K} units are actually available, measured capital productivity falls to ϕ_1 . Unused (or excess) capacity is measured by $\bar{K} - K_1$.

As with labor, the extreme short run does not last long. If demand is not restored then firms begin adjusting their capital to suit the new optimal techniques. Over time, through purchases of new types of machinery and depreciation, the capital stock is readjusted to move the economy back toward the new optimal point on the production function.

6.4.3 MEASURES OF RESOURCE USE

Labor Utilization

Economists have developed measures of the degree of idleness of the factors of production. Start with labor. The **labor force (LF)** is *the number of people who wish to be in paid employment*. If L is the actual number employed, then

- the **employment rate (EMP)** = L/LF is *the fraction of the labor force employed*;
- and

- the **unemployment rate** (U) = $\frac{LF - L}{LF} = 1 - EMP$, is *the fraction of the labor force not employed*.

Employment and unemployment are examined in much greater detail in Chapters 8 and 9.

Capital Utilization

What unemployment is to labor, **UNUSED, SPARE, OR EXCESS CAPACITY** is to capital. In the United States, the Federal Reserve collects data from which it computes the **capacity utilization rate**, which is the analogue of the employment rate.

Consider capacity at the level of the firm. Imagine that both capital and labor are used optimally according to a long-run production function. The optimal output on this production function point is the theoretical capacity of the firm's capital. In practice, a firm's decision about optimal production is a complex one. To collect data, the Federal Reserve asks managers of individual plants to estimate the

sustainable practical capacity [of the plant], which is defined as the greatest level of output that a plant can maintain within the framework of a realistic work schedule, taking account of normal downtime, and assuming sufficient availability of inputs to operate the machinery and equipment in place.²

Data for individual plants are combined into a large number of indices for more specific industrial groupings. These are, in turn, combined into various indices for larger industrial groupings and an overall *industrial capacity index*. The Federal Reserve also collects data on actual production for these same industries and combines them into an **index of industrial production** and indices for various industrial groupings. Both the

² Board of Governors of the Federal Reserve System, "About the G.17: Industrial Production and Capacity Utilization," <http://www.bog.frb.fed.us/releases/G17/About.htm>.

capacity index and the index of industrial production are based so that their values in 1997 equal 100.

The capacity index does not cover all production in the economy, but concentrates on the physical output of the manufacturing, mining, and electric and gas utilities. Actual and potential production is more difficult to measure in other sectors of the economy – especially in the increasingly important service sectors. Even so, as we saw in Chapter 5 (Figure 5.9), the index of industrial production is one of the best coincident indicators of the economy: it tracks real GDP very closely.

The *capacity utilization rate* (which we will call *CU*) is defined as

$$CU = \frac{\text{index of industrial production}}{\text{capacity index}};$$

that is, the capacity utilization rate expresses the level of output as a fraction of the economically optimal level of output of the physical plant and equipment in the economy. Figure 6.17 plots the capacity utilization rate for manufacturing against the business cycle.

Just as the unemployment rate does not fall to zero even in the best of times, the capacity utilization rate never reaches 100 percent. Just as the troughs of the unemployment rate gradually increased between the 1950s and the early 1990s, the peaks of capacity utilization have fallen from highs around 92 percent in the 1950s to the peak in the last complete business cycle of 85 percent. Effectively full capacity utilization is probably close to the most recent peak value.

The unemployment and capacity utilization rates fluctuate with the business cycle. Yet, even in the best of times, the unemployment rate never falls to zero, and the capacity utilization rate never reaches 100 percent. The average levels of each measure may differ significantly among countries. In recent years, Germany and France, for instance have had much higher average unemployment rates than the United States. These differences reflect differences in national labor, industrial, tax, and social welfare policies.

[Place Box 6.2 near here.]

6.5 Potential Output

6.5.1 THE CONCEPT OF ECONOMIC POTENTIAL

The fact that the unemployment rate is greater than zero and the capacity utilization rate is less than 100 percent, suggests that the economy produces less than it could. A natural notion of **POTENTIAL OUTPUT** answers the question, *how much real GDP would be produced if the entire labor force and the entire capital stock were fully employed at their economically optimal levels?* The Cobb-Douglas production function provides an easy way to address this question. Substitute the whole labor force (LF) for (L) in equation (6.8) and assume that capital is fully employed to get:

$$(6.22) \quad Y^{pot} = ALF^{\alpha} K^{1-\alpha},$$

where Y^{pot} is potential GDP.

Previously, we had assumed that capital was always fully employed. But if (6.22) defines potential GDP, then it would make more sense to adjust the capital stock to take account of unused capacity: capital in effective use is $CU \times K$. The production of real GDP could then be represented by the Cobb-Douglas production function written as:

$$(6.23) \quad Y = AL^\alpha (CU \times K)^{1-\alpha} .$$

Practically, the main difference between this equation and equation (6.8) is that the variability of capacity utilization was previously hidden in the variability of A , while now it is explicitly represented in its interaction with the capital stock.

6.5.2 SCALED OUTPUT

The measure of potential output based on full employment of labor and capital has a nice interpretation. Begin with real GDP as a fraction of potential output, calculated as

$$(6.24) \quad \tilde{Y} = \frac{Y}{Y^{pot}} .$$

Let us call \tilde{Y} *output scaled by potential output* or just **SCALED OUTPUT** for short.³ Scaled output is a kind of generalized measure of capacity utilization that takes the rate of labor use, as well as the rate of capital use, into account. Scaled output can be expressed in

³ The symbol “~” is known as a *tilde*, and the variable \tilde{Y} can be read as “*Y-tilde*.”

terms of the Cobb-Douglas production function by substituting equations (6.22) and (6.23) into the definition:

$$(6.25) \quad \tilde{Y} = Y / Y^{pot} = \frac{A L^\alpha (CU \times K)^{1-\alpha}}{A L F^\alpha K^{1-\alpha}} = (L / LF)^\alpha CU^{1-\alpha} .$$

To simplify the notation, define $\tilde{L} = L/LF$ and $\tilde{K} = CU$. (We can call \tilde{L} **scaled labor** and \tilde{K} **scaled capital**. \tilde{L} is, of course, another name for the employment rate, *EMP*, and \tilde{K} is another name for the capacity utilization rate, *CU*. Neither new name is necessary, but they help to distinguish the more generalized notions of factor utilization needed here from the actual data, which never correspond precisely to the ideal concepts of economic theory.

Equation (6.25) can be rewritten as

$$(6.26) \quad \tilde{Y} = \tilde{L}^\alpha \tilde{K}^{1-\alpha} .$$

Equation (6.26) has the same form as a Cobb-Douglas production function without the level of technological progress (total factor productivity), usually indicated by *A*. Technological progress is not omitted. Rather, it is incorporated into potential output in the calculation of \tilde{Y} .

Figure 6.18 shows potential output, actual output and scaled output against the business cycle. The usefulness of scaled output is already evident in the figure. Both actual and potential output trend upwards as the economy grows. Similarly, labor and

capital trend upwards over time. And technology (measured by total factor productivity, *A*) also advances. But the calculation of scaled output uses factor inputs expressed as fractions of the available inputs, which can never exceed 100 percent, and technology cancels out. As a result, scaled output does not trend. Since the growth of the economy is factored into its calculation, it is easy to judge the state of the economy at different times, even though the economy is now much bigger than in the past.

The fluctuations in scaled output follow the business cycle closely. Scaled output is a leading-to-coincident indicator of the business-cycle peak and a clearly coincident indicator of the trough.

6.6 Aggregate Supply: Questions Answered, Questions Raised

At the head of this chapter we posed some questions – three examples of the type of question that falls under the general heading of aggregate supply. This chapter showed that questions of that type can be answered using the aggregate production function and the assumption that firms are profit maximizers. While these assumptions are only approximate, we saw that the Cobb-Douglas production function provides a good approximation. It allowed us to answer the most specific question (“What would happen to GDP if immigration rules were relaxed at 5 percent more workers joined the U.S. labor force?”) with reasonable precision: *ceteris paribus* GDP would rise by 3.5 percent (see section 6.2.4).

The more general questions (“Why is GDP in the United States vastly higher in 2005 than in 1955 or 1905?” and “Why is GDP in the United States higher than in any other country in the world?”) received only an implicit answer: GDP is higher when

labor or capital inputs are higher or when technology is better. Our analysis would help us to calculate just how much of a change in labor, capital, or technology it would take to account for the differences in GDP over time or between countries. These questions are addressed in more detail in Chapter 7. Our analysis raises new questions: If factor inputs help to determine output, what determines factor inputs? This question is addressed in Chapters 8 and 9. If cyclical fluctuations in aggregate demand result in aggregate supply less than full potential, what determines aggregate demand? We examine that question in Chapters 12-15.

Summary

1. Aggregate supply is the combined effect of the production decisions of individual firms.
2. Production converts inputs (factors of production), broadly classified as labor and capital (non-labor) into outputs using particular techniques. Technology is the set of available techniques.
3. The production function represents the set of the best available techniques for producing each possible level of production. An acceptable production function shows no output when there are no inputs; increases as inputs increase; and displays diminishing returns to each factor of production (that is, output increases at a decreasing rate as each factor increases *ceteris paribus*).
4. When all factors of production increase by the same proportion, the production function shows constant returns to scale if it increases by the same proportion;

- diminishing returns to scale, if it increases by a lesser proportion; and increasing returns to scale, if it increases by a greater proportion.
5. The firm maximizes profits when they set the level of production such that marginal cost equal marginal revenue ($MC = MR$). When firms are perfectly competitive, this rule can be expressed as: a firm maximize profits when it uses labor up to the point that the marginal product of labor equals the real wage ($mpl = w/p$) and the marginal product of capital equals the real (implicit) rental price of capital ($mpk = v/p$).
 6. The aggregate production function is the macroeconomic analogue of the firm's production function in which GDP (Y) replaces the output of a particular product, and aggregate labor and capital replace the individual firm's factors of production.
 7. The Cobb-Douglas production function ($Y = AL^\alpha K^{1-\alpha}$) provides a useful representation of aggregate supply. Its properties are similar to those of microeconomic production functions. In addition, it shows other useful properties: constant returns to scale; an increase in a factor of production decreases the productivity of that factor but raises the productivity of other factors; and technological progress raises the productivity of all factors.
 8. Under the assumption that firms are approximately perfectly competitive and that it describes the economy reasonably well, the Cobb-Douglas production function predicts constant labor and capital shares in GDP. This prediction agrees to a good approximation with the facts about the U.S. economy in the long run.
 9. Productivity is measured as units of output per unit of input. The most common measures are labor productivity, capital productivity, and total factor productivity, which is the weighted geometric average of the first two.

10. Factor-augmenting technological progress occurs when technological progress can be represented as equivalent to adding more of the factor of production. Since labor productivity in the United States shows a strong trend, and capital productivity almost no trend, technological progress can be represented as labor-augmenting.
11. In the short run, it is not possible to adjust the choice of technique to changes in the level of demand or factor prices – technology is inflexible. But over time, if changes are long-lasting, techniques adapt. Short-run inflexibility leads to underutilization of factors of production (labor hoarding and unemployment and excess capacity).
12. The unemployment rate measures the underutilization of workers as a percentage of the labor force. The capacity utilization rate measures the utilization of capital as the ratio of industrial production to industrial capacity, so that a capacity utilization rate less than 100 percent shows underutilization.
13. Potential output is the amount of GDP that would be produced if both labor and capital were fully utilized with the best choice of technique. The ratio of actual to potential output (scaled output) measures how intensely the economy is supplying GDP. Fluctuations in scaled output closely track the business cycle.

Key Concepts

inputs	aggregate production function
output	Cobb-Douglas production function
technique	total factor (or multifactor) productivity
technology	(TFP)
factors of production	labor productivity
labor	capital productivity
production function	factor-augmenting technological
capital	progress
diminishing returns to factors of	long run
production	short run
constant returns to scale	labor hoarding
perfect competition	unemployment
marginal product of labor (mpl)	unused, spare, or excess capacity
real wage rate (w/p)	potential output
implicit rental rate (v/p)	scaled output
marginal product of capital (mpk)	

Problems

Data for this exercise are available on the course website under the link for Chapter 6 (**insert web link here**). Before starting these exercises, the student should review the relevant portions of the *Guide to Working with Economic Data*: sections G.12, G.13, and G.16.

Problem 6.1. Imagine that Florida was an independent country and that a series of hurricanes destroyed 10 percent of its capital stock (while miraculously not harming any person). Use production function diagrams to show what this disaster would imply for (i) the level of GDP; (ii) labor productivity; and (iii) capital productivity. Assume that the labor force is fixed and fully employed.

Problem 6.2. Using a Cobb-Douglas production function for the United States, calibrated to the values of 2003, what would you predict *ceteris paribus* would be the quantitative effects of a 10 percent fall in the capital stock on (i) the level of GDP; (ii) labor productivity; and (iii) capital productivity.

Problem 6.3. Using a diagram similar to Figure 6.4 and 6.5 analyze the effects of a 20 percent increase in both capital and labor when production is subject to decreasing returns to scale. Particularly, note the effects on labor and capital productivity.

Problem 6.4. Consider a production which is like the Cobb-Douglas production function except that the exponents on the factors of production need not add up to one:

$$Y = AL^\alpha K^\beta$$

(Notice that if $\alpha + \beta = 1$, this is exactly the same as a Cobb-Douglas production function.) Show that, if $\alpha + \beta > 1$, this function displays *increasing returns to scale* (that is, output rises more than proportionately to equiproportional increases in factor inputs), and that, if $\alpha + \beta < 1$, it displays *decreasing returns to scale*. Think of economic situations in which each of these cases seems natural.

Problem 6.5. Using appropriate annual data for the United States, compute the time series (1948-2002) for total factor productivity (A), labor productivity, and capital productivity. (Assume that the labor share is constant at $\alpha = 0.69$.) Compute the average rate of growth of each (compound annual rates). What do your calculations suggest about the sources of productivity growth?

Problem 6.6.

- (a) Using the same data and calculations from Problem 6.4, compute the implied real wage in 1948, 1978, and 2002.
- (b) Compute the average growth rates of the real wage in (a) for 1948-2002, 1948-1978, 1978-2002.
- (c) Compare the result for 1948-2002 to your computation of the growth rate of the average product of labor in Problem 6.4. Explain your finding.
- (d) The U.S. economy is often characterized favorably compared to European economies as a “job creation machine.” What light do the data in parts (a) and (b) shed on the costs of this success. (Be careful to relate your conclusions to specific data.)

Problem 6.7. Use quarterly data for labor productivity in the United States from 1948 to the present. Detrend the series (explain your choice of detrending method.)

- (a) Before attempting any data analysis. Think about the business cycle using the models of this chapter. Explain how you might expect labor and capital productivity to vary with the business cycle (for example, do you expect productivity to be a leading, lagging, or coincident indicator; pro-cyclical or countercyclical). State your reasoning carefully.
- (b) Plot the detrended labor productivity against the NBER business-cycle dates (shading the recessions).

- (c) Create a graph showing the typical cyclical behavior of labor productivity. First identify the segment of the data that run from 12 quarters before to 36 quarters after each of the last seven business cycle peaks (excluding the most recent if there are not enough data.) Convert each segment into an index taking the value 100 at the business cycle peak. Average across all seven cycles. Plot the resulting series on the vertical axis against the values -12, -11, . . . , 0, 1, 2, . . . 36 (horizontal axis) – i.e., against the number of quarters after the peak. (Your graph should be similar to Figures 5.8 and 5.9.)
- (d) Based on your graphs is there a clear cyclical pattern of either productivity series? Comment on the nature of the pattern. How well do these patterns agree with those you hypothesized in (a)? What might account for the differences between your expectations and the data?

Problem 6.8. Using quarterly data for the United States from 1948 to the present. Compute capital productivity. Using this series repeat steps (a)-(c) in Problem 6.7, substituting capital productivity for labor productivity where appropriate. (Since capital productivity does not have a significant trend, there is no need to use a detrended series in (b).)

Problem 6.9. Assume that the economy was fully employed and efficiently operating at the peak of the business cycle in 1990.3.

- (a) Write down the equation of the short-run labor production function that corresponds to the segment of the ray from the origin to point A in Figure 6.14.
- (b) Noting how far real GDP actually fell from the peak to the trough in 1991.1, use the function from (a) to compute the amount of labor needed to produce the 1991.1 level of output (i.e., the amount of labor that corresponds to point B in Figure 6.14. What level of capital productivity corresponds to that level of output assuming that capital had remained fixed at the 1990.3 level?
- (c) What were the actual levels of labor, labor productivity, and capital productivity in 1991.1? Comparing these to the data and the computations from (b), how well does the short-run production function (Figures 6.14 and 6.14) describe the recession of 1990-91? What might account for the difference between what it predicted and the actual outcomes?

Problem 6.10. Using quarterly data on real GDP and worker-hours plot labor productivity calculated as Y/L . On the same graph, using a separate scale, plot the government's index of labor productivity. Take first differences of each series (to remove the trends), then calculate the correlation coefficient for the two series. Comment on how closely the two estimates of productivity are related. What might explain the differences between them?

Problem 6.11. Economists widely believe that sometime in the 1960s or 1970s productivity growth slowed down and that sometime in the mid 1990s it speeded up again. Using either of the two productivity series in Problem 6.6 (yours or the government's) and any appropriate statistical tool determine the date or dates, if any of a slowdown or speedup in labor productivity.

Problem 6.12. Using the dates determined in the Problem 6.7, calculate the average rates of growth of labor productivity for each of the periods that you have identified. Describe the pattern of your findings.

Problem 6.13. Is Belgium more productive than the United States? Repeat the analysis in Box 6.2 for Belgium, using the U.S. data presented there and the following Belgian data for 2003: productivity per hour, 120.1 percent of the European Union average (Source: Eurostat); average hours of work per year, 1,542 hours per year; unemployment rate, 8.1 percent (Source: OECD).

Box 6.1. Working with Economic Data: The Cobb-Douglas Production Function

In the *Guide*, section G.16, the properties of the Cobb-Douglas production function are derived using calculus. They are also easily illustrated with simple numerical examples.

To apply the Cobb-Douglas production function in equation (6.8) to the data of the actual economy, we would have to have estimates of the key parameters, A and α . To illustrate its properties, however, we can just make up some numbers. Let $A = 1$ and $\alpha = 0.75$. Then the production function can be written as

$$Y = L^{0.75} K^{0.25}.$$

To illustrate the properties of this function, we calculate the values of Y for different levels of the inputs, L and K . The table below shows sixteen different combinations – four levels of L matched with four levels of K .

		L			
		0.0	1.0	1.1	1.2
K	0.0	0.0	0.0	0.0	0.0
	1.0	0.0	1.0000	1.0741	1.1465
	1.1	0.0	1.0241	1.1000	1.1742
	1.2	0.0	1.0466	1.1242	1.2000

The student should plug some of the values of L and K into the production function above to check that these values are correct.

Using the table, we can prove the six properties in Section 6.2.2.

➤ **Property 1** (*GDP increases with each factor of production ceteris paribus.*):

Each row shows what happens to Y as L is increased, holding K constant. As we read across any row, Y increases. Similarly, each column shows what happens to Y as K is increased, holding L constant. The values of Y increase in both directions.

➤ **Property 2** (*The production function goes through the origin*): GDP is zero for every entry in the first row (capital zero) and the first column (labor zero) – “there’s no such thing as a free lunch.”

➤ **Property 3** (*diminishing returns to each factor of production*): Looking at the first row ($K = 1.0$), Y increases by 0.0741 as L increases from 1 to 1.1 and by a smaller amount, 0.0724, as L increases from 1.1 to 1.2. The small increase in Y is the marginal product of labor. The student can confirm that the same pattern of the second equal increase in L corresponding to a smaller increase in Y is true for the other rows as well. Similarly, holding L constant and looking down the columns, the same pattern holds for equal increases in K .

➤ **Property 4** (*an increase in one factor of production ceteris paribus raises the marginal product of the other factors*): Looking at the row with $K = 1$, the marginal effect of adding 0.1 units of labor (from $L = 1$ to $L = 1.1$) is $\Delta Y = 1.0741 - 1.0000 = 0.0741$, so that the marginal product of labor is $\Delta Y/\Delta L = 0.741$. Looking at the row with $K = 1.1$, the same change in labor yields a higher $\Delta Y = 1.1000 - 1.0241 = 0.0759$, so that the marginal product of labor is also higher ($\Delta Y/\Delta L = 0.759$). The student can confirm that the same sort of result applies to capital holding labor constant by looking down the rows.

- **Property 5** (*an increase in technology raises the marginal product of each factor of production*). The student should Recompute the values in the table on the assumption that $A = 2$ instead of $A = 1$. Comparing the new table to the old, every non-zero cell will have a higher value. (See Problem 6.
- **Property 6** (*constant returns to scale*): The last two cells along the main diagonal correspond to equal percentage increases in each of the factors (10, and 20 percent) compared to the second cell ($K = 1, L = 1$). Notice that Y increases by the same percentage in each case.

Box 6.2. Is France More Productive Than the United States?

Figure 6.11 shows that the United States has the highest productivity in the world as measured by GDP per worker but that several countries, including France, have higher productivity than the United States as measured by GDP per hour worked. Using more recent Eurostat data (2003) GDP per hour in the United States is 116.4 percent of the European Union average, while French GDP per hour is 123.8 – that is, 7.4 percentage points higher. The French also work fewer hours per year on average than Americans. Some commentators have interpreted this to mean that France has an intrinsically more productive economy: the French are rich enough to enjoy leisure while the Americans have to work extra hard to keep up. Is this the right way to look at it?

We know from diminishing returns to a factor of production (property 3 of the production function) that a reduction in an input raises the productivity of that input. If the French work fewer hours, then some increase in their productivity is not the result of having a better production function, but simply the result of diminishing returns. And not only do the French work fewer hours (1,453 per year versus 1,792 for the United States in 2003, according to Eurostat), a smaller proportion of their workforce is actually employed (91.6 percent versus 94 percent for the United States, according to the OECD). One way to address the question is to ask, how much would productivity in the United States change if the United States had the same rate of employment and average hours as France?

Actual output in the United States in 2003 can be described by a Cobb-Douglas production function:

$$Y_{2003} = A_{2003}L^{\alpha}K^{1-\alpha} = 9.27(237,375.6 \text{ million})^{0.69}(9,562.9 \text{ billion})^{0.31} = \$_{96}9,562.9 \text{ billion.}$$

Actual labor productivity is \$40.28 per hour ($= \$_{96}9,562.9 \times 10^9 / 237,375.6 \times 10^6$).

For every 100 workers employed out of the U.S. labor force only 97.4 ($= (91.6/94) \times 100$) are employed out of the French labor force. And on average employed French workers work only 81.1 percent ($= 1,453/1,792$) as many hours. Therefore, a labor input 79 percent ($= 0.974 \times 0.811$) of the actual input would give the equivalent of the French labor input. The counterfactual question is: what would GDP in 2003 have been if Americans worked according to French patterns? The answer:

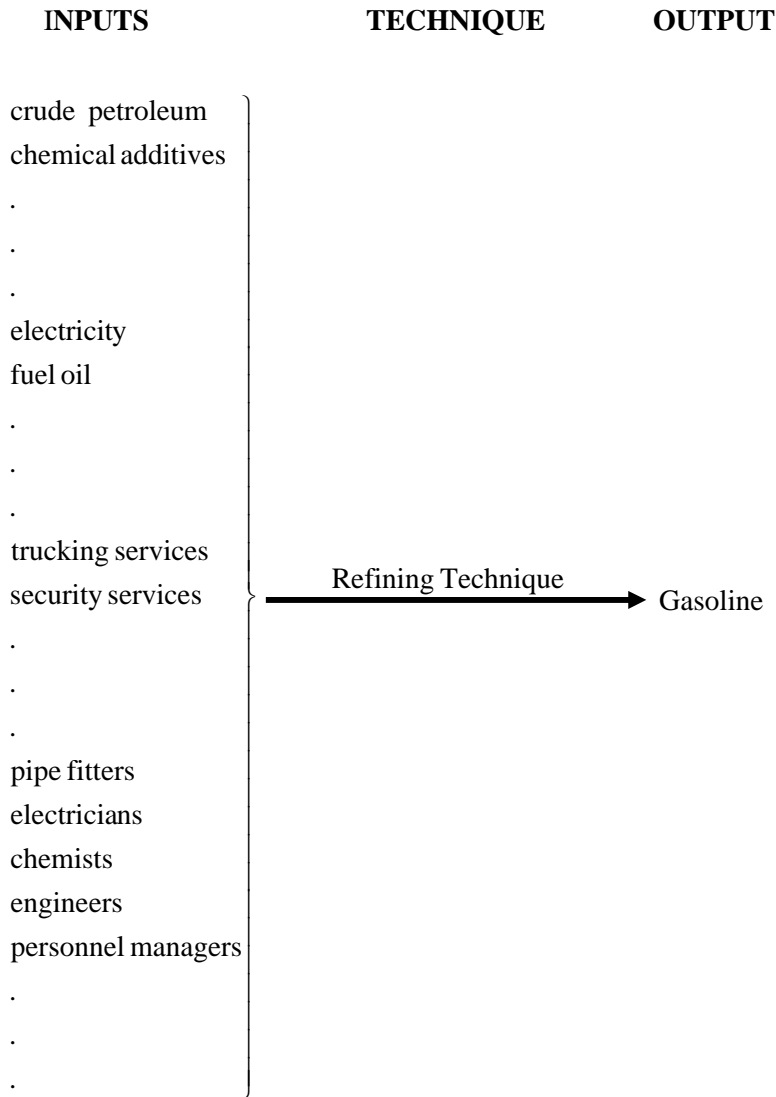
$$\begin{aligned} Y_{2003} &= A_{2003}L^{\alpha}K^{1-\alpha} \\ &= 9.27(0.79 \times 237,375.6 \times 10^6)^{0.69}(9,562.9 \times 10^9)^{0.31} = \$_{96}8,125.9 \text{ billion.} \end{aligned}$$

Counterfactual labor productivity would be \$43.33 per hour ($= 8,125.9 \times 10^9 / (0.79 \times 237,375.6 \times 10^6)$), which is 7.6 percent higher than actual.

Therefore, if Americans worked like the French, productivity in the United States would be 125.2 ($= 1.076 \times 116.4$), which is 1.4 percentage points *greater* than French productivity.

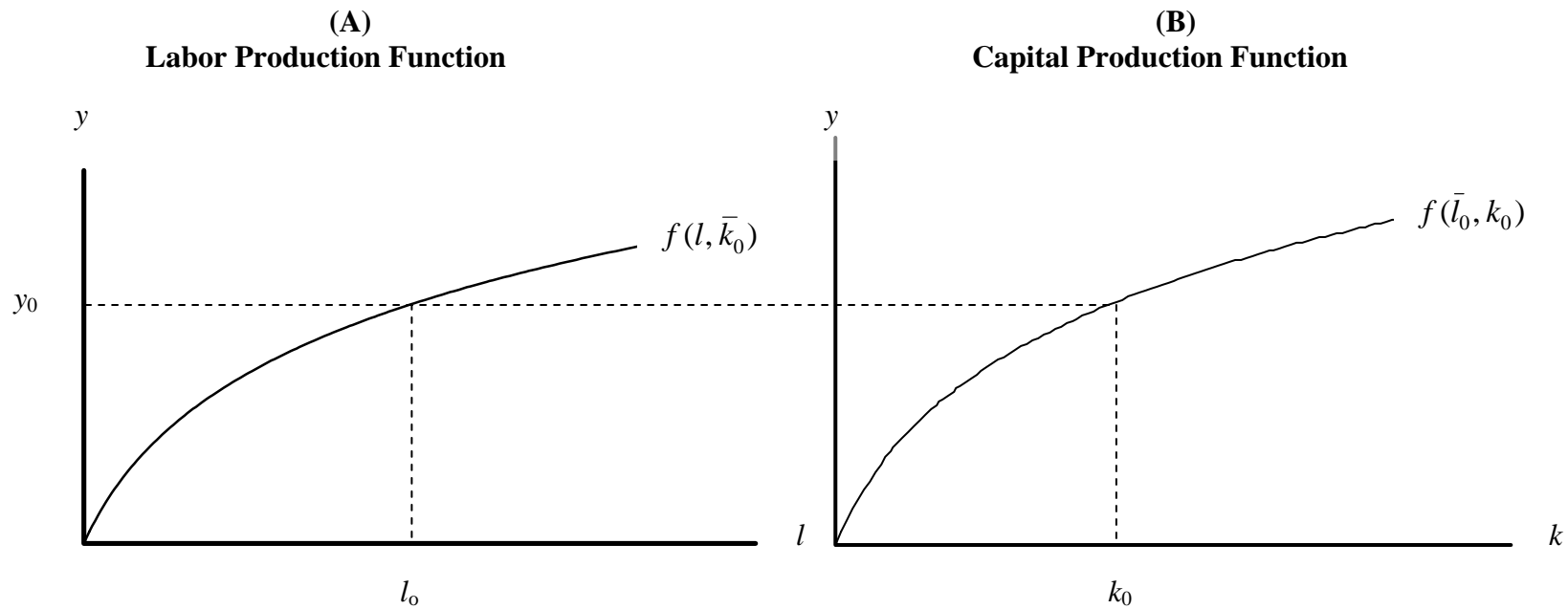
It is also the case that France has much higher capital per worker than the United States. How qualitatively would these results have changed if the United States had as much capital per worker as the French?

Figure 6.1
A Schematic View of a Production Process



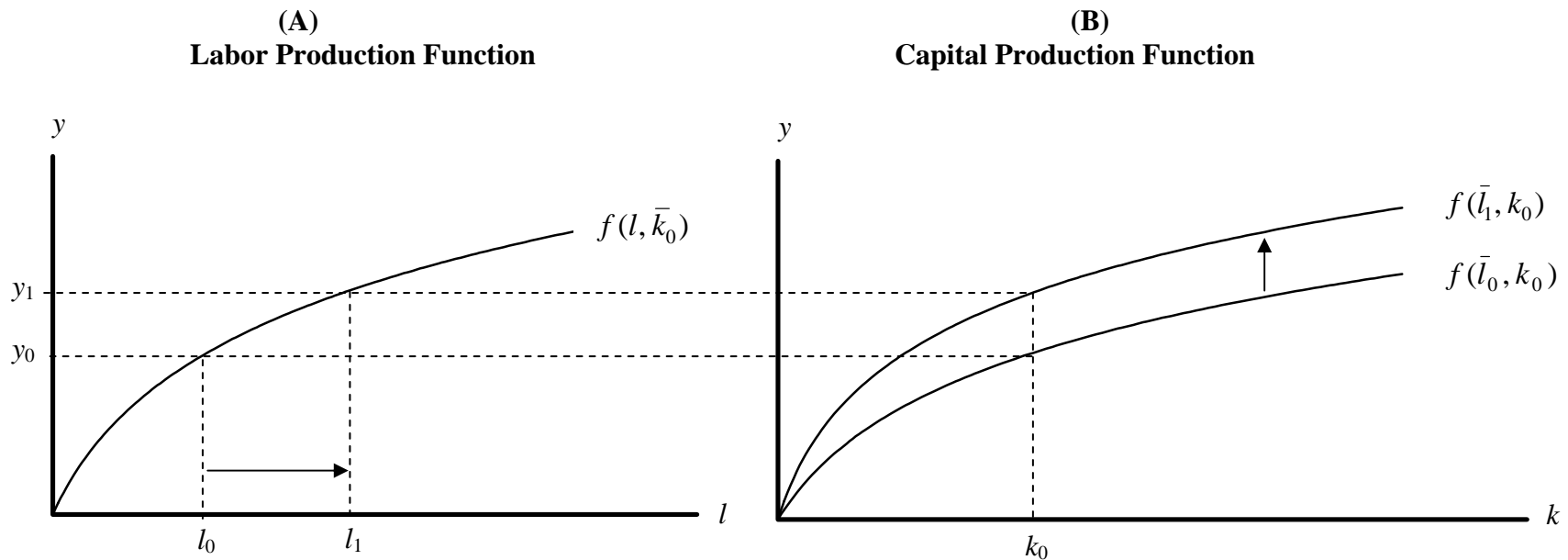
Production of individual commodities such as gasoline requires the use of particular techniques drawn from the available technology to convert (frequently large number of) inputs into an output.

Figure 6.2
A Production Function



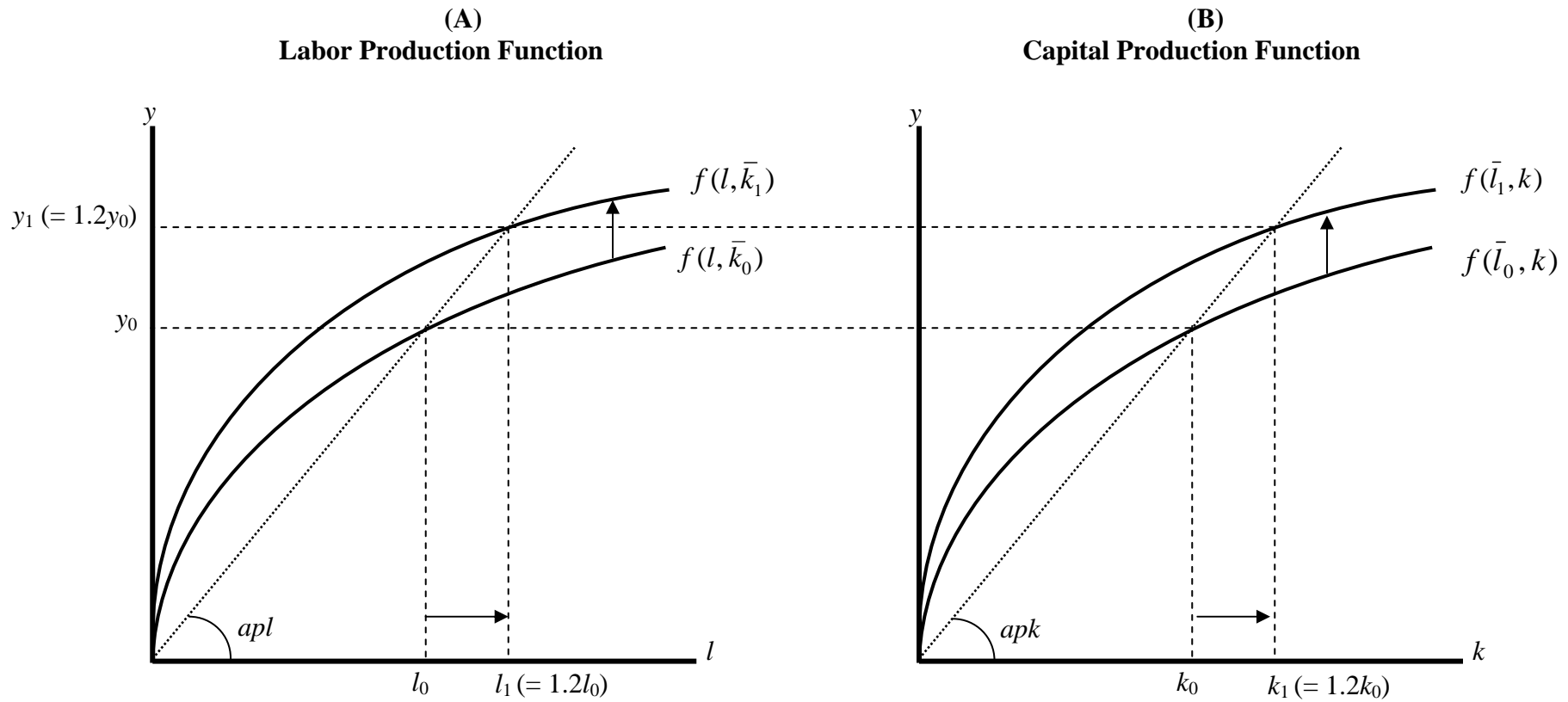
A production function in which output is related to two inputs, labor and capital, can be shown in two panels. In panel (A), the curve shows how output varies with the labor input provided that capital is fixed at \bar{k}_0 . The bar over k_0 is a reminder that the curve presupposes that capital is fixed at that level. In panel (B), the curve shows how output varies with the capital input provided that labor is fixed at \bar{l}_0 .

Figure 6.3
The Effect of an Increase in an Input



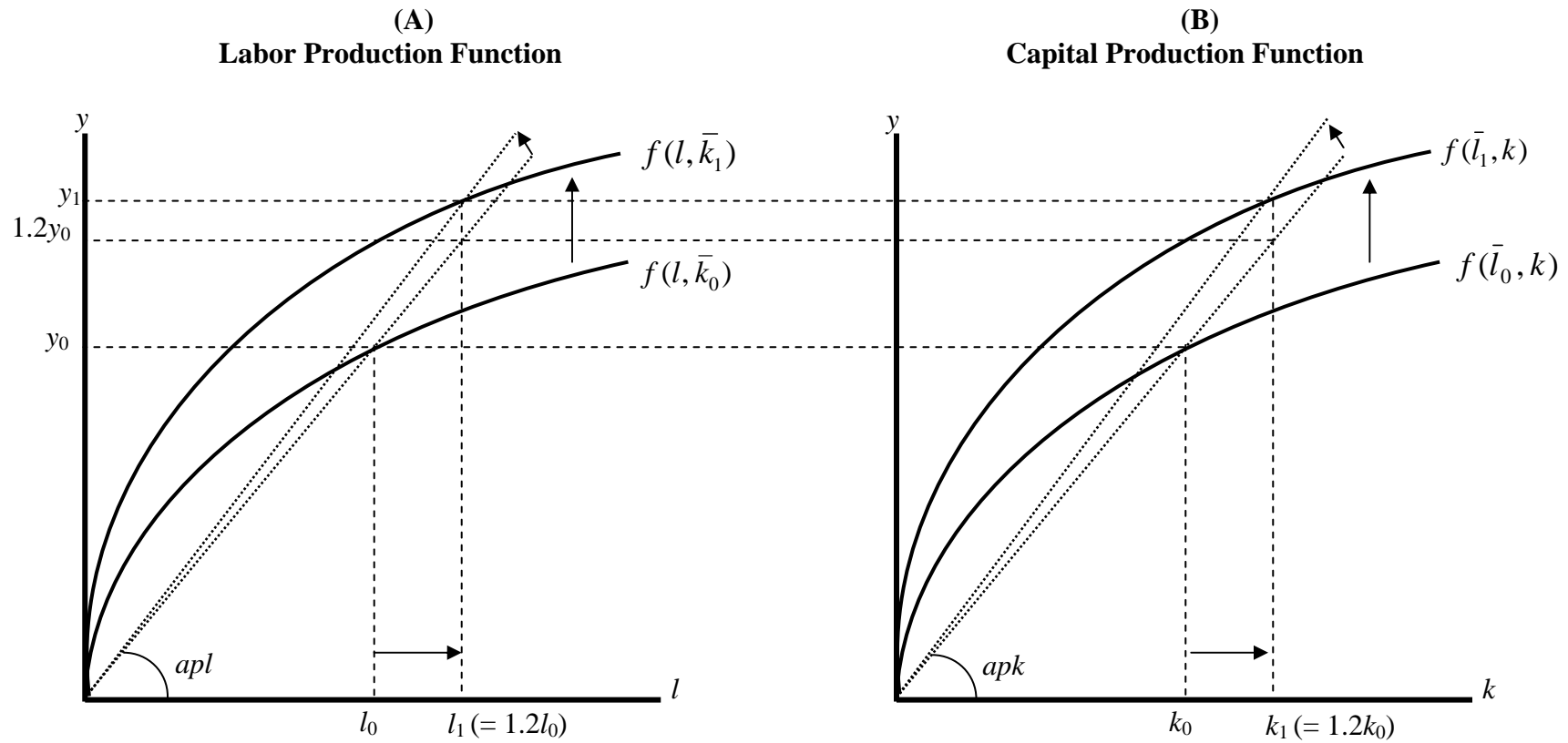
An increase in labor holding capital constant (panel (A)) is a movement along the labor production function. The increase of labor from l_0 to l_1 increases output from y_0 to y_1 . Since labor has increased, the original capital production function (panel (B)), which was drawn on the assumption of labor at l_0 must shift upwards so that it intersects the new level of output immediately above the unchanged level of capital.

Figure 6.4
Constant Returns to Scale



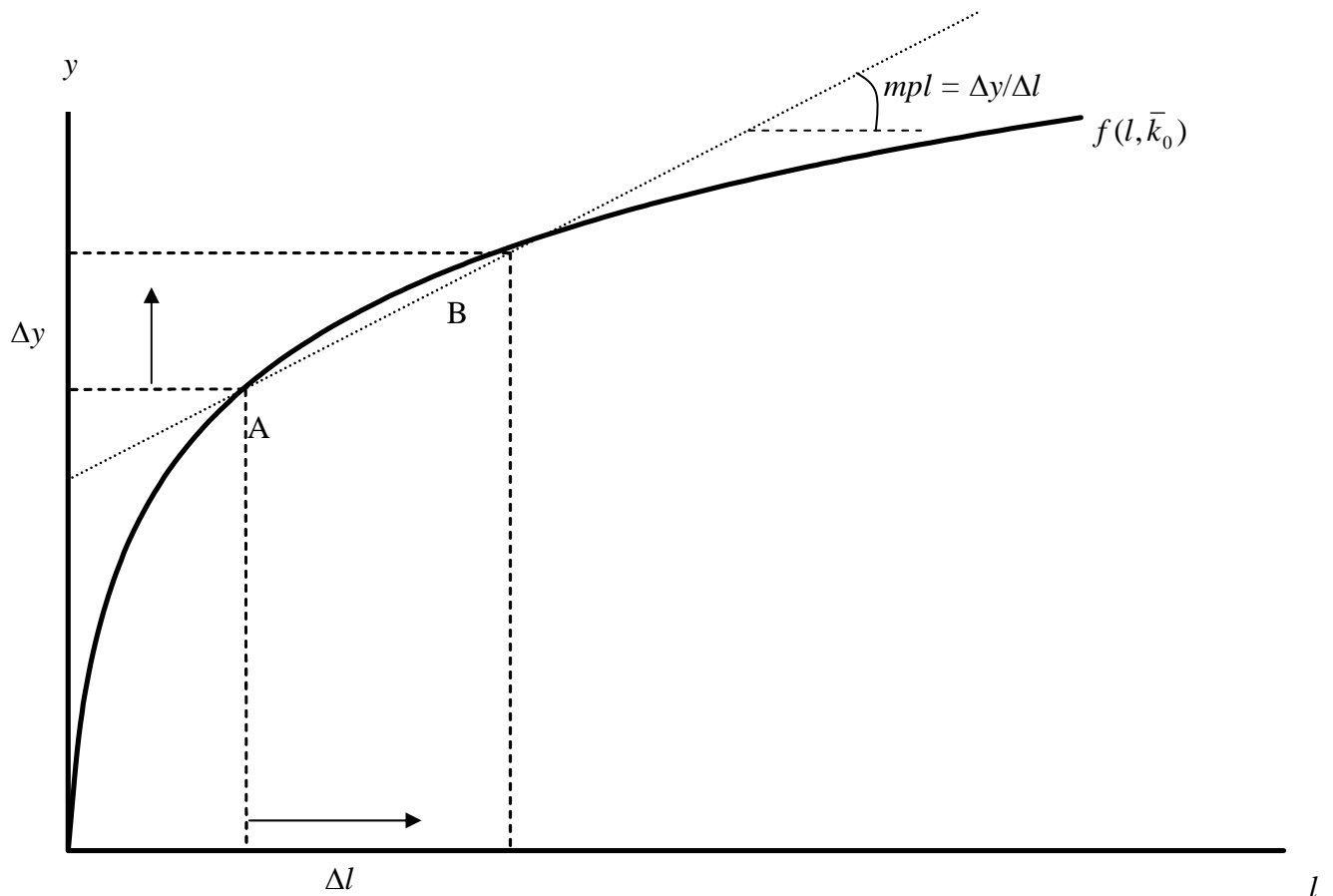
A production function displays constant returns to scale when equiproportional increases in all factors of production generate a proportionately equal increase in output. In this case, both labor and capital rise by 20 percent, as does output. The average product of labor ($apl = y/l$) and the average product of capital ($apk = y/k$) are indicated by the slope of a ray from the origin to the point of production. Constant returns to scale implies that the apl and apk remain constant as output expands: higher production points lie on the same ray from the origin as lower production points.

Figure 6.5
Increasing Returns to Scale



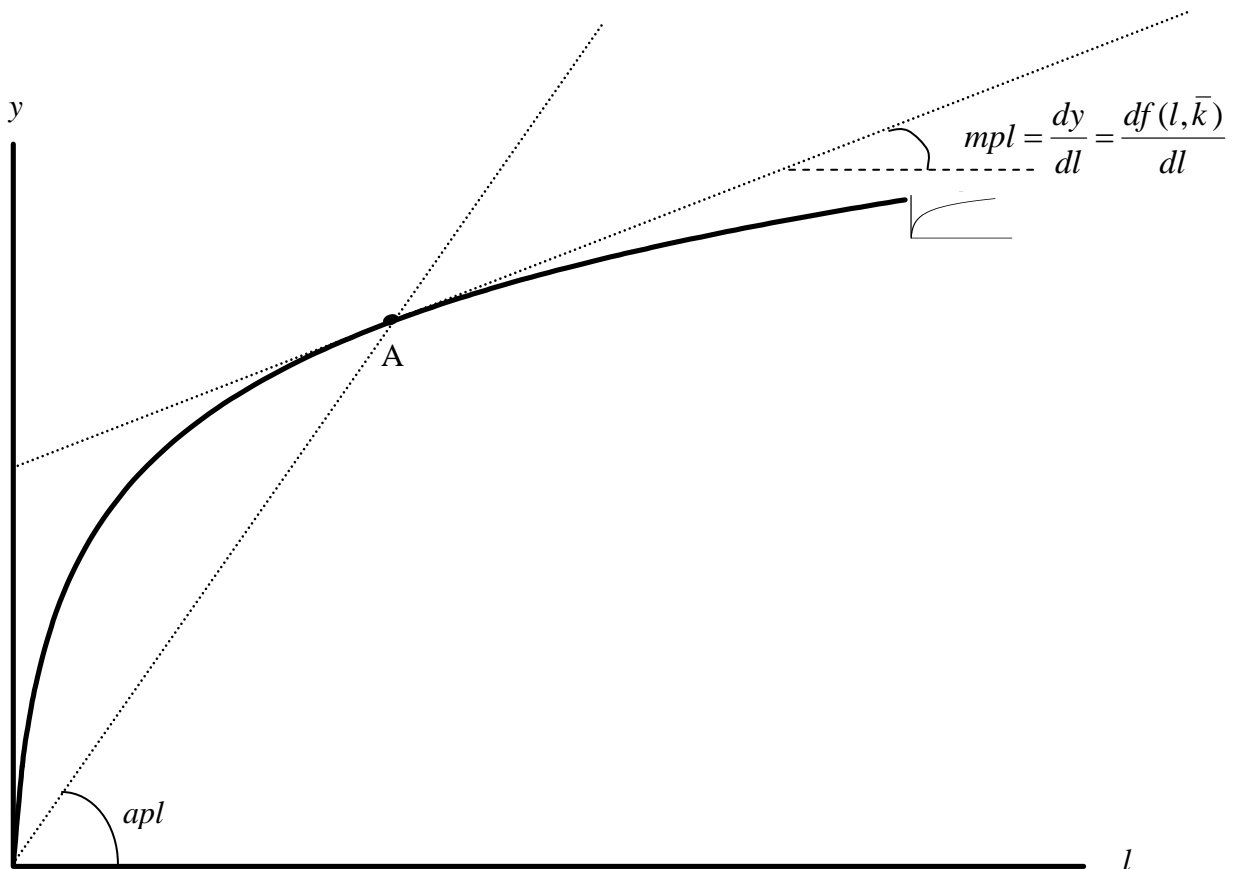
A production function displays increasing returns to scale when equiproportional increases in all factors of production generate a proportionately greater increase in output. In this case, both labor and capital rise by 20 percent, but output rises by more than 20 percent. Since output rose faster than either labor or capital, the slope of ray from the origin to the new production point in each panel is steeper than one to the old production point, indicating that the average products of labor (apl) and capital (apk) rise as inputs rise equiproportionately.

Figure 6.6
The Marginal Product of Labor for Discrete Changes



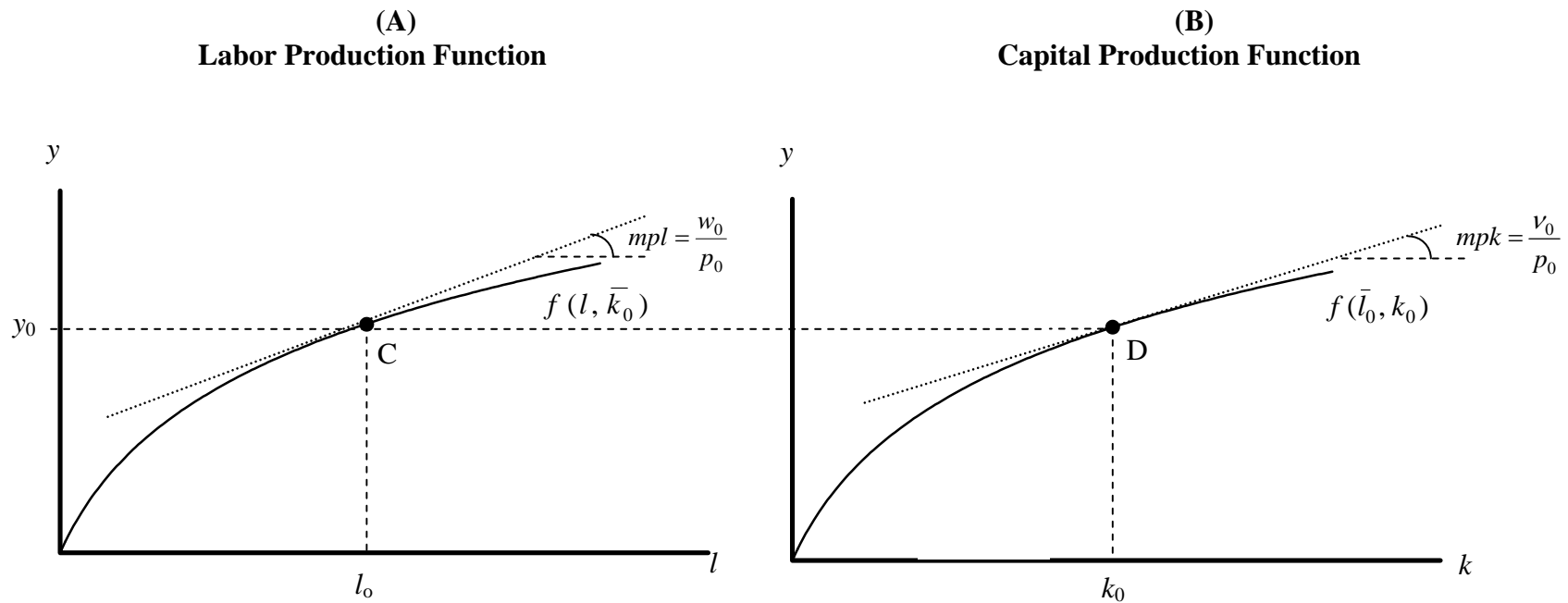
The marginal product of labor is the small increase in output (Δy) that results, holding other inputs to production constant, from a small increase in labor (Δl). The diagram shows only the labor production function. The slope of the line connecting the production points before after the increase in labor (that is, connecting point A to point B) measures the marginal product of labor ($mpl = \Delta y / \Delta l$).

Figure 6.7
The Marginal Product of Labor for Infinitesimal Changes



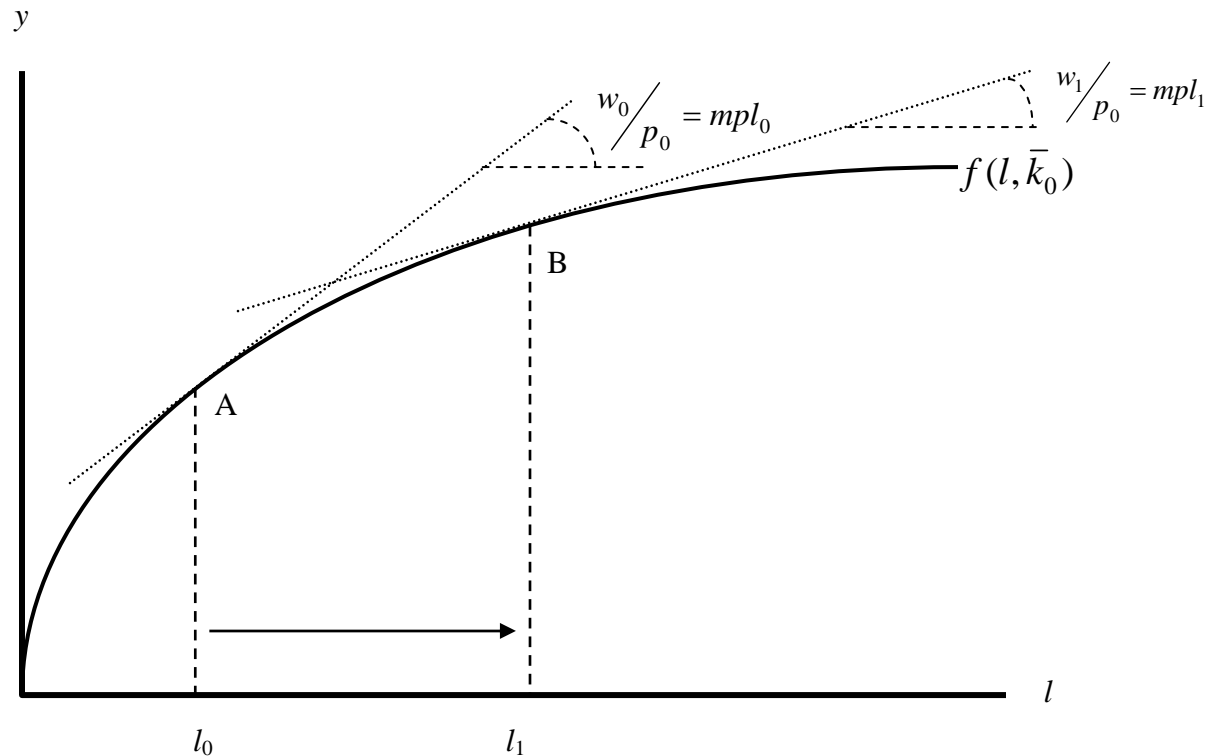
If we imagine the discrete change in labor (Δl) in Figure 6.6 becoming smaller and smaller without limit, the line running through points A and B will eventually become tangent to the production function at point A. The slope of the tangent is the derivative of the production function at point A, so that the marginal product of labor $mpl = dy/dl$. The slope of the tangent to the production function at any point is always less than the slope of ray from the origin to that point, indicating that always $mpl < apl$. This property reflects diminishing returns to a factor of production.

Figure 6.8
Optimal Output Adjusts to the Prices of Factors of Production



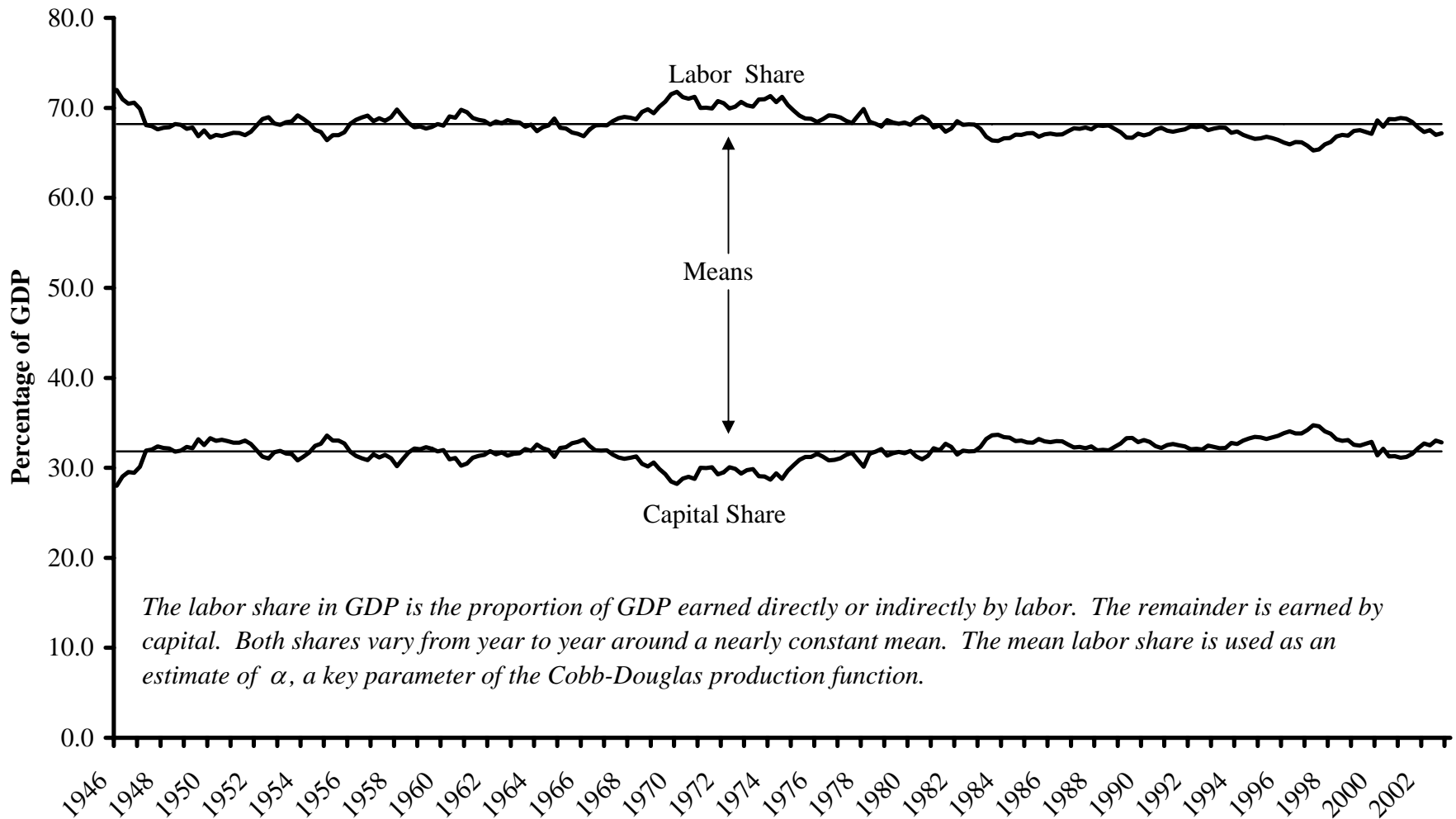
The firm maximizes its profits by hiring workers up to the point that the marginal product of labor is equal to the real price of each factor of production. Panel (A) shows that at the production point C (y_0, l_0), the marginal product of labor (the slope of a tangent to the production function at point C) is just equal to the real wage ($mpl = w_0/p_0$). Panel B shows that at the production point D (y_0, k_0), the marginal product of capital (the slope of a tangent to the production function at point D) is just equal to the real (implicit) rental rate ($mpk = v_0/p_0$).

Figure 6.9
The Consequences of a Change in the Real Price of a Factor of Production
Labor Production Function



A firm that faces a real wage rate of w_0/p_0 would produce at point A, where the marginal product of labor is equal to that wage rate ($mpl_0 = w_0/p_0$). If the wage rate falls to w_1/p_1 and labor becomes cheaper, the optimal production point would correspond to a lower marginal product of labor. Because of diminishing returns to labor, a lower mpl lies to the right on the production function at point B where $mpl_1 = w_1/p_1$. To reach this point, the firm must hire more labor, increasing the demand for labor from l_0 to l_1 .

Figure 6.10
U.S. Labor and Capital Shares



Source: Bureau of Economic Analysis and author's calculations.

Figure 6.11
Productivity in Selected Countries

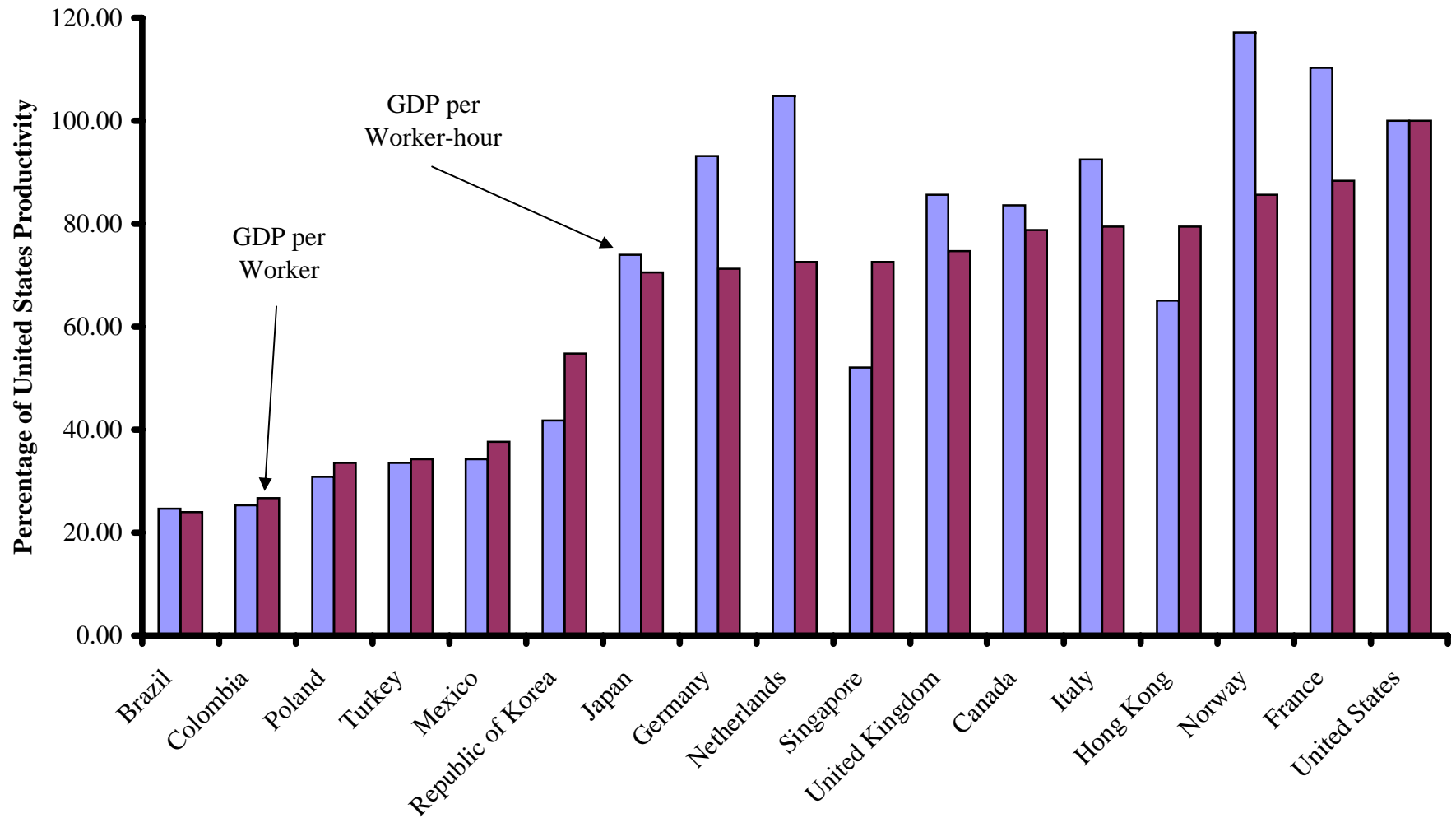
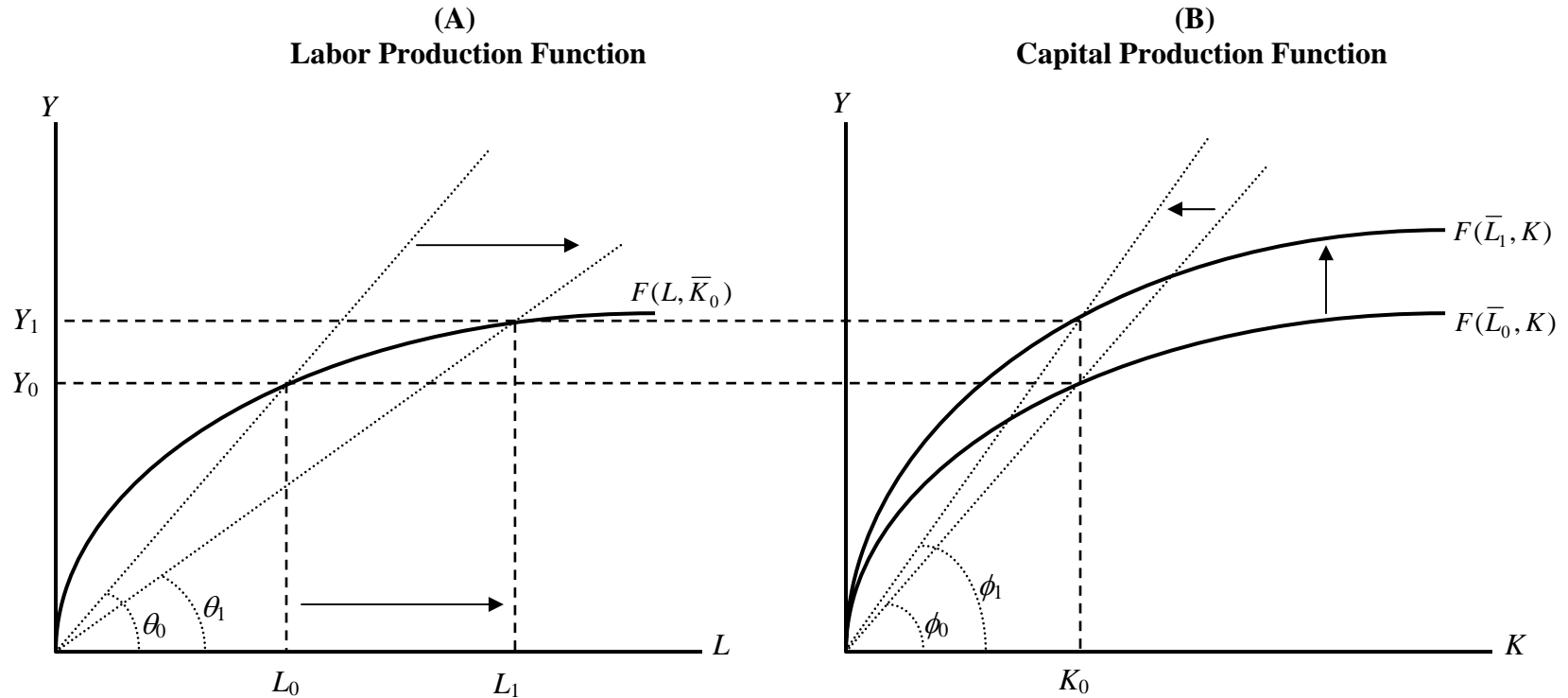
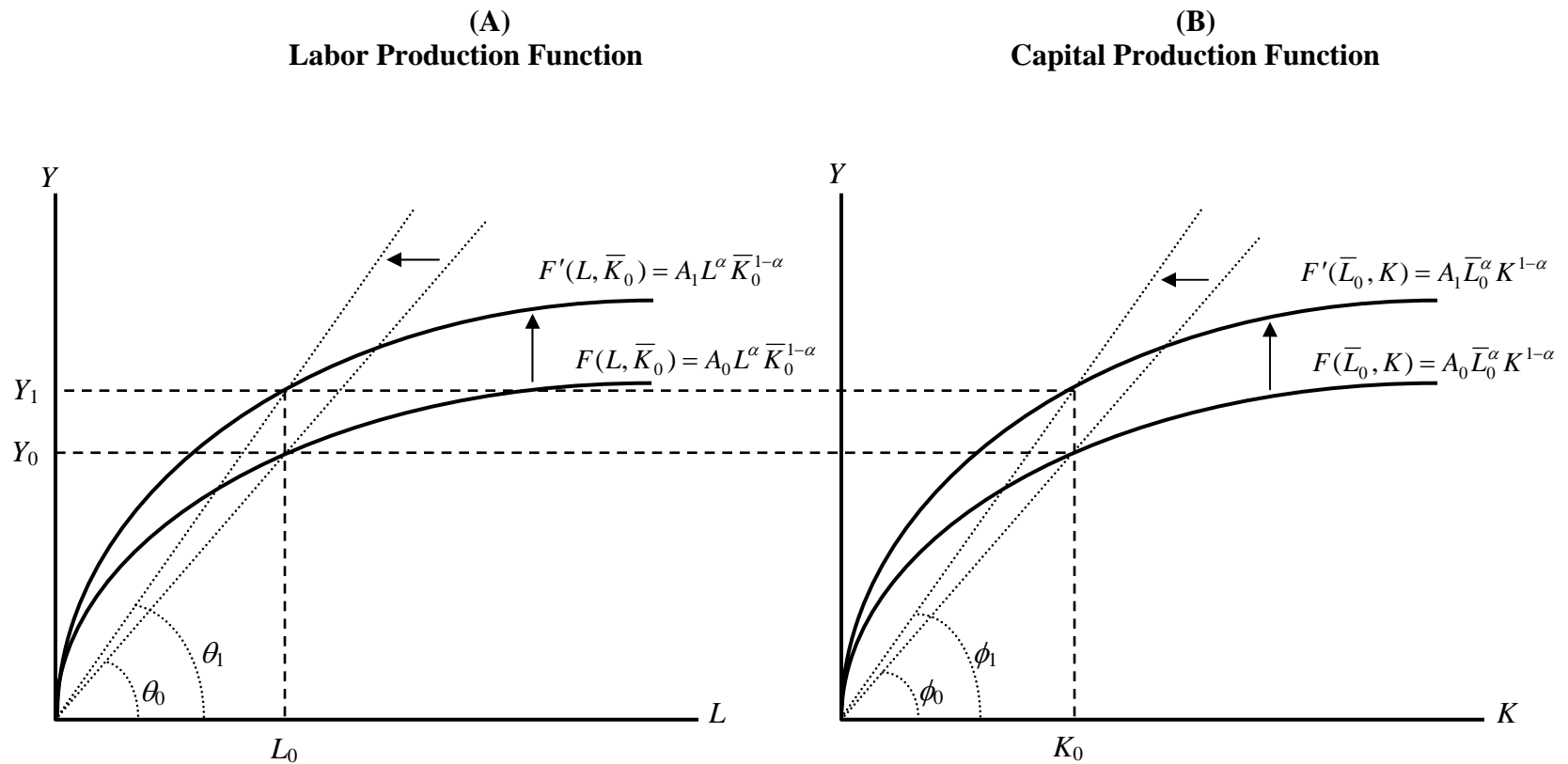


Figure 6.12
Factor Productivities Change in Opposite Directions When One Factor Production is Increased



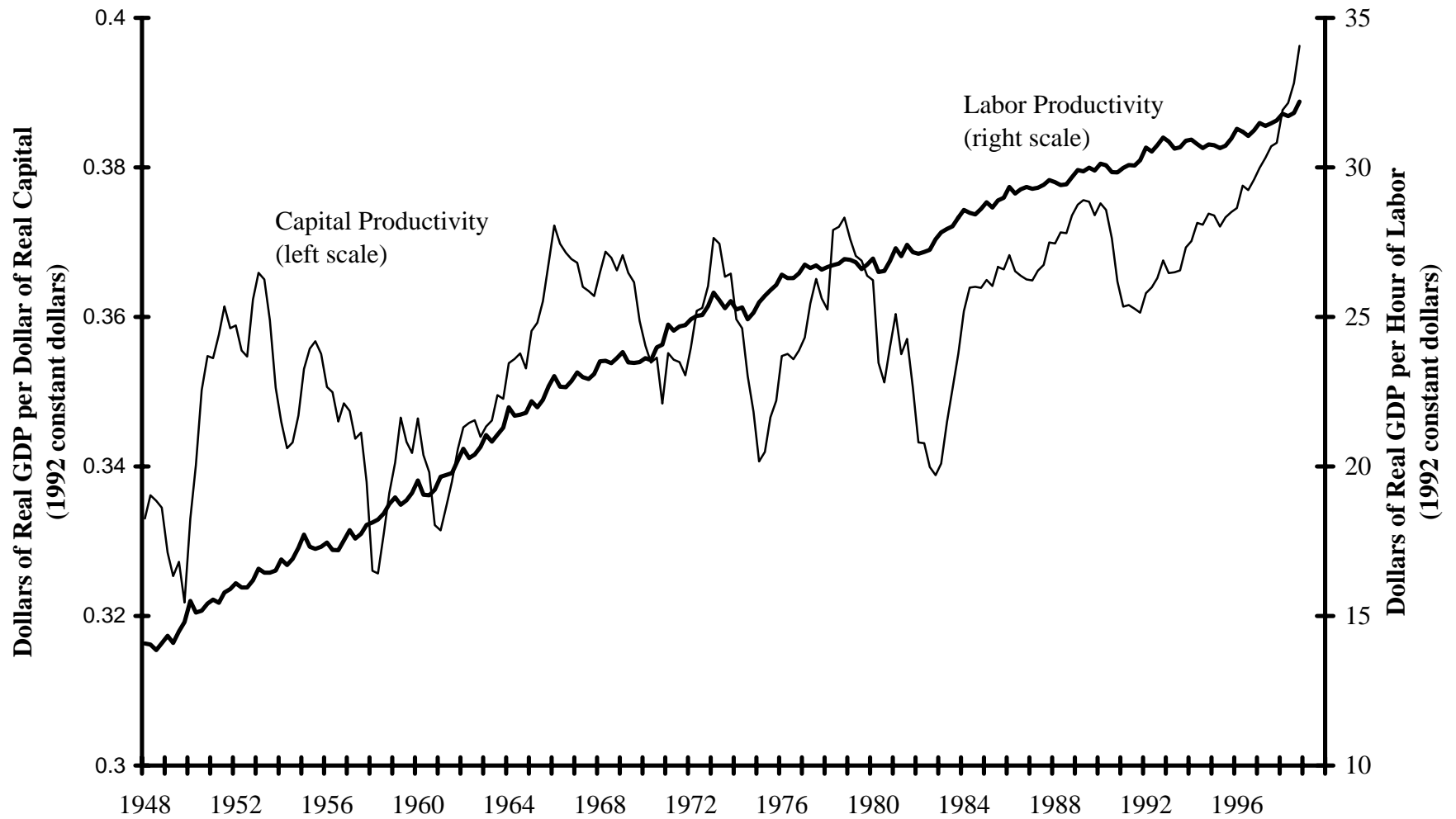
Production initial takes place at (Y_0, L_0, K_0) at which point labor productivity is given by θ_0 and capital productivity by ϕ_0 – each is the slope of a ray from the origin to the production point. The increase of labor to L_1 raises GDP to Y_1 as the production point moves out along the labor production function to (Y_1, L_1) . Diminishing returns to labor implies a fall in labor productivity to θ_1 . Since the capital stock is unchanged, the capital production function must shift upwards to intersect the new production point at (Y_1, K_0) . The ray from the origin through this new point is steeper ($\phi_1 > \phi_0$), indicating that capital productivity is now higher.

Figure 6.13
All Factor Productivities Rise When Total Factory Productivity Rises



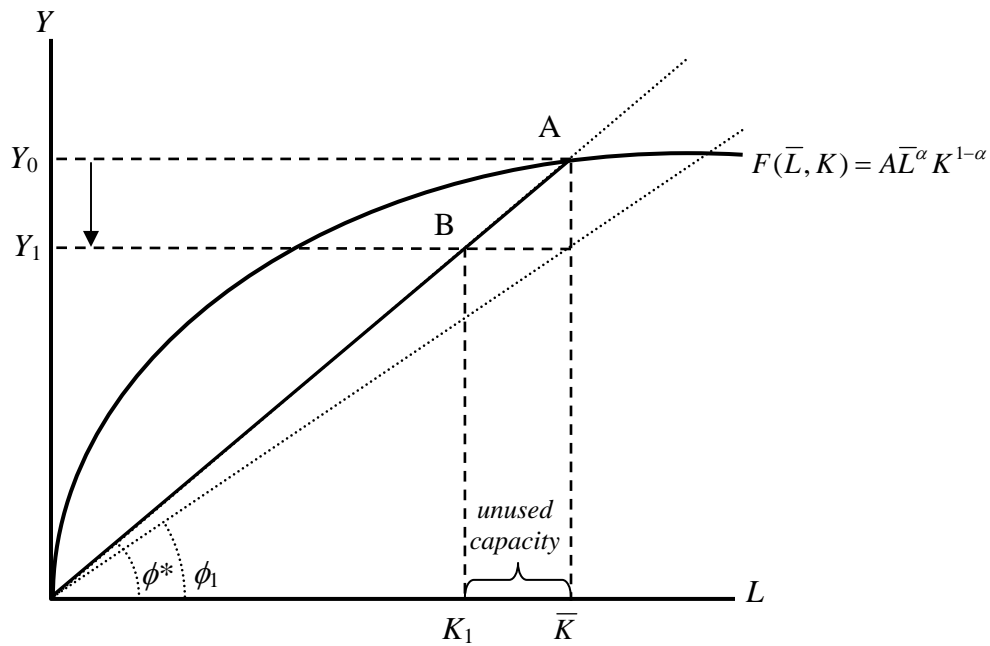
Production initial takes place at (Y_0, L_0, K_0) , where the production function is $F(L_0, K_0) = A_0 L_0^\alpha K_0^{1-\alpha}$. Initial labor productivity is given by θ_0 and capital productivity by ϕ_0 – each is the slope of a ray from the origin to the production point. The increase in total factor productivity to A_1 raises GDP to Y_1 at the original levels of labor and capital. The production function must shift up to intersect the new production point at (Y_1, L_0, K_0) . As a result the rays from the origin to the production point steepen, indicating a rise in labor productivity to θ_1 and in capital productivity to ϕ_1 .

Figure 6.14
U.S. Labor and Capital Productivity



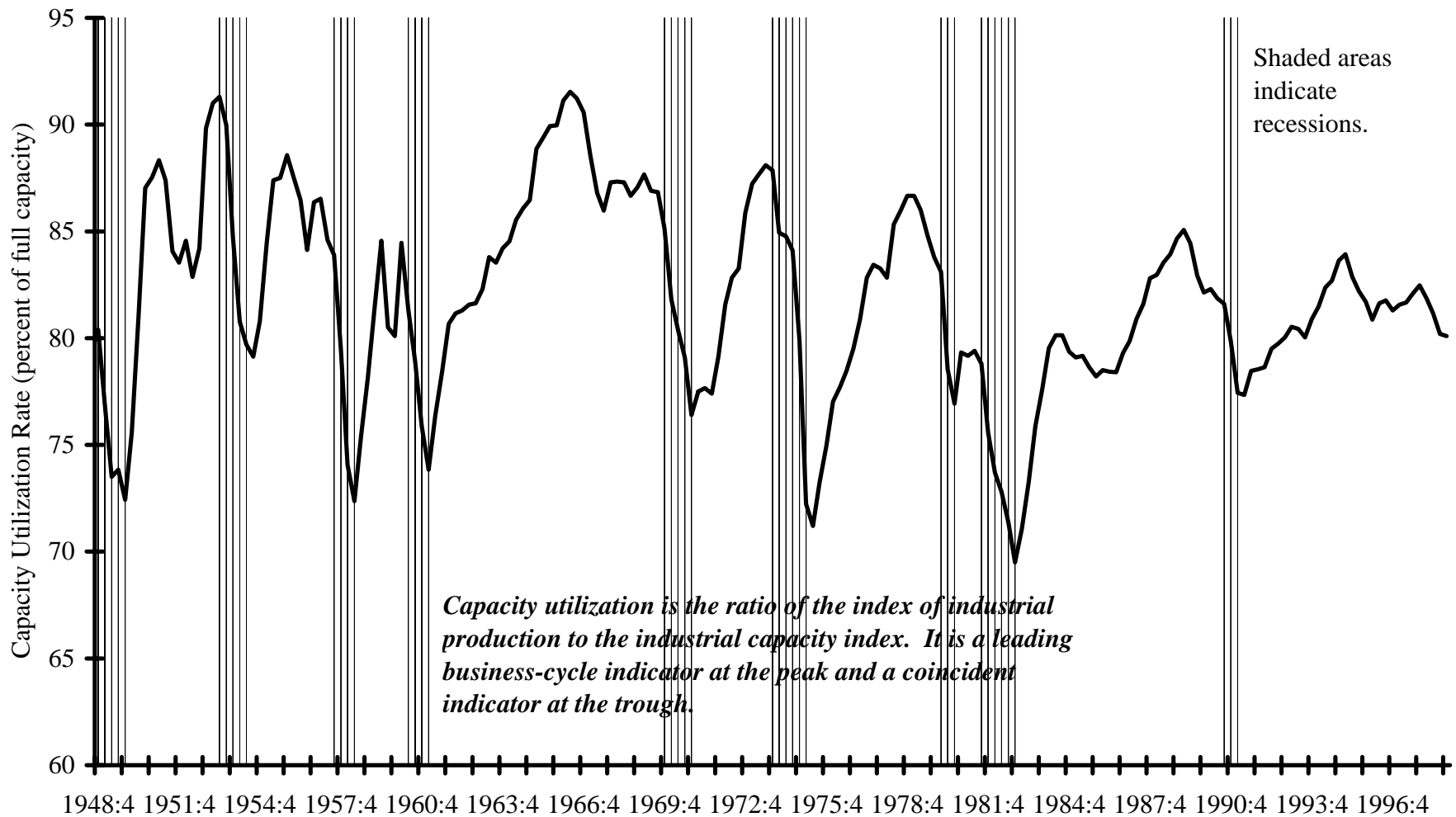
Source: Bureau of Economic Analysis, Bureau of Labor Statistics and author's calculations.

Figure 6.16
Short-run Adjustment to a Fall in Aggregate Demand:
Capital Productivity and Capacity Utilization



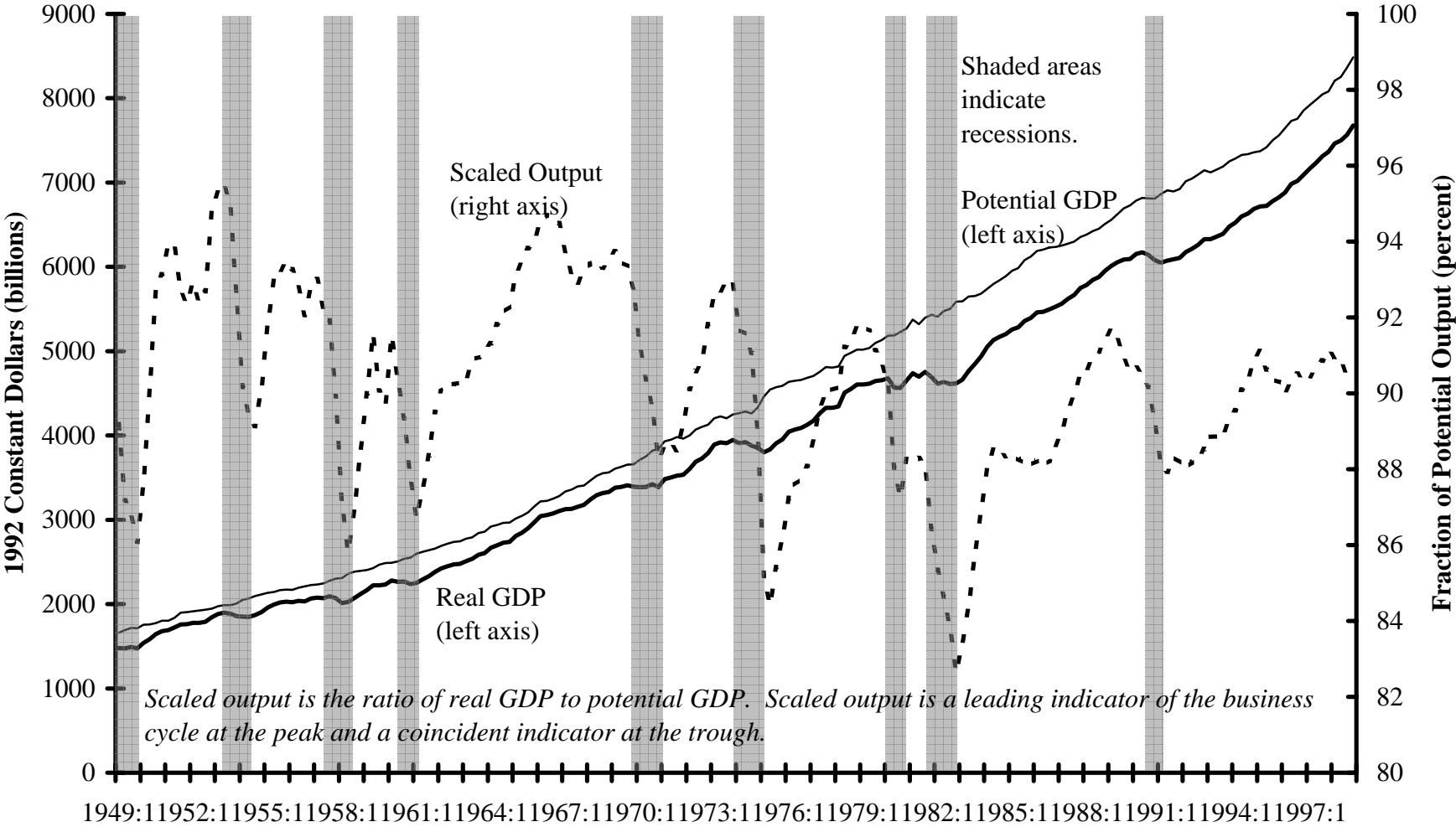
Initially production occurs at point A at where the whole capital stock (\bar{K}) is used to produce Y_0 . If aggregate demand falls to Y_1 but technology is not flexible, firms would continue to use the same technique, so that the heavy segment of the ray from the origin to point A would describe the short-run production function. Firms producing at point B would still possess the capacity to produce at a higher level, and the portion of capital actually used (K_1) would produce with the original productivity (ϕ^). Actual capital productivity (counting both the used and the unused portions of capital) would fall to ϕ_1 and $\bar{K} - K_1$ would be unused capacity. The existence of unused capacity permits a rapid rise in actual capital productivity when demand is restored to a higher level.*

Figure 6.17
Capacity Utilization and the Business Cycle



Source: Board of Governors of the Federal Reserve System and the National Bureau of Economic Research.

Figure 6.18
Output, Potential Output, and Scaled Output



Sources: Bureau of Economic Analysis, Bureau of Labor Statistics, the Board of Governors of the Federal Reserve System.

Table 6.1
Profit-maximization Decision for a Golf-ball Manufacturer

Workers	Production	Revenue: <i>p × production</i>	Cost: <i>w × workers</i>	Profit	Marginal Product: $\Delta y/\Delta l$	Marginal Cost: <i>w/p</i>
	<i>(balls per hour)</i>	<i>(dollars/hour)</i>	<i>(dollars per hour)</i>	<i>dollars</i>	<i>(balls per hour)</i>	<i>(balls per hour)</i>
99	4996	2498	1188	1310	–	–
100	5024	2512	1200	1312	28	24
101	5050	2525	1212	1313	24	24
102	5066	2533	1224	1309	16	24

Calculations assume that that the firm can hire workers at \$12 per hour and sell golf balls at 50¢ each.