CHAPTER 7

The Cost of Production

In the last chapter, we examined the firm's production technology-the relationship that shows how factor inputs can be transformed into outputs. Now we will see how the production technology, together with the prices of factor inputs, determine the firm's cost of production.

Given a firm's production technology, managers must decide *how* to produce. As we saw, inputs can be combined in different ways to yield the same amount of output. For example, one can produce a certain output with a lot of labor and very little capital, with very little labor and a lot of capital, or with some other combination of the two. In this chapter we see how the *optimal* (cost-minimizing) combination of inputs is chosen. We will also see how a firm's costs depend on its rate of output, and how these costs are likely to change over time.

We begin by explaining how cost is defined and measured, distinguishing between the concept of cost used by economists, who are concerned about the firm's performance, and by accountants, who focus on the firm's financial statements. We then examine how the characteristics of the firm's production technology affect costs, both in the short run when the firm can do little to change its capital stock, and in the long run when the firm can change all its factor inputs.

We then show how the concept of returns to scale can be modified to apply to the process of producing many different outputs. We also show how cost sometimes falls over time as managers and workers learn from experience, so that the production process becomes more efficient. Finally, we show how empirical information can be used to estimate cost functions and predict future cost.

7.1 *Measuring Cost: Which Costs Matter?*

Before we can analyze how cost is determined, we need to be clear about what we mean by cost and how we measure it. What items should be included as part of a firm's cost? Cost obviously includes the wages a firm pays its workers and the rent it pays for office space. But what if the firm already owns an office building and doesn't have to pay rent? And how should we treat money that the firm spent two or three years ago (and can't recover) for equipment or for research and development? We'll answer these questions in the context of the economic decisions that managers make.

Economic Cost versus Accounting Cost

An economist thinks of cost differently from an accountant, who is concerned with the firm's financial statements. Accountants tend to take a retrospective look at a firm's finances because they have to keep track of assets and liabilities and evaluate past performance. Accounting cost includes depreciation expenses for capital equipment, which are determined on the basis of the allowable tax treatment by the Internal Revenue Service.

Economists-and we hope managers-take a forward-looking view of the firm. They are concerned with what cost is expected to be in the future, and with how the firm might be able to rearrange its resources to lower its cost and improve its profitability. They must therefore be concerned with *opportunity cost,* the cost associated with opportunities that are foregone by not putting the firm's resources to their highest value use.

For example, consider a firm that owns a building, and therefore pays no rent for office space. Does this mean that the cost of office space is zero? Although an accountant would treat this cost as zero, an economist would note that the firm could have earned rent on the office space by leasing it to another company. This foregone rent is an opportunity cost of utilizing the office space and should be included as part of the cost of doing business.

Accountants and economists both include actual outlays, called *explicit costs,* in their calculations. Explicit costs include wages, salaries, and the cost of materials and property rentals. For accountants, explicit costs are important because they involve direct payments by a company to other firms and individuals that it does business with. These costs are relevant for the economist because the cost of wages and materials represents money that could have usefully been spent elsewhere.

Let's take a look at how economic cost can differ from accounting cost in the treatment of wages and economic depreciation. For example, consider an owner who manages her own retail store but chooses not to pay herself a salary. Although no monetary transaction has occurred (and thus would not appear as an accounting cost), the business nonetheless incurs an opportunity cost because the owner could have earned a competitive salary by working elsewhere.

Accountants and economists also treat depreciation differently. When estimating the future profitability of a business, an economist or manager is concerned with the capital cost of plant and machinery. This involves not only the explicit cost of buying and then running the machinery, but also the cost associated with wear and tear. When evaluating past performance, accountants use tax rules that apply to broadly defined types of assets to determine allowable depreciation in their cost and profit calculations. But these depreciation allowances need not reflect the actual wear and tear on the equipment, which is likely to vary asset by asset.

Sunk Costs

Although opportunity cost is often hidden, it should be taken into account when making economic decisions. Just the opposite is true of sunk cost-it is usually visible, but after it has been incurred, it should always be ignored when making future economic decisions.

A *sunk cost* is an expenditure that has been made and cannot be recovered. Because it cannot be recovered, it should not influence the firm's decisions. For example, consider the purchase of specialized equipment designed to order for a plant. We assume the equipment can be used to do only what it was originally designed for and can't be converted for alternative use. The expenditure on this equipment is a sunk cost. Because it has no alternative use, its opportunity cost is zero. Thus it shouldn't be included as part of the firm's costs.[1] The decision to buy this equipment may have been good or bad. It doesn't matter. It's water under the bridge and shouldn't affect the firm's current decisions.

For example, suppose a firm is considering moving its headquarters to a new city. Last year it paid $500,000 for an option to buy a building in the city; the option gives it the right to buy the building at a cost of $5,000,000, so that its total expenditure will be $5,500,000 if it indeed buys the building. Now it finds that a comparable building has become available in the same city at a price of $5,250,000. Which building should it buy? The answer is the original building. The $500,000 option is a cost that has been sunk and that should not affect the firm's current decision. The economic cost of the original property is $5,000,000 to the firm (because the sunk cost of the option is not part of the economic cost), while the newer property has an economic cost of $5,250,000. Of course, if the new building cost $4,750,000, the firm should buy it, and forgo its option,

[1] If, on the other hand, the equipment could be put to other use, or be sold or rented to another firm, its current economic cost would be measured by the value from its next most profitable use.

EXAMPLE 7.1 CHOOSING THE LOCATION FOR A NEW LAW SCHOOL BUILDING

The Northwestern University Law School has long been located in Chicago, along the shores of Lake Michigan. However, the main campus of the university is located in the suburb of Evanston. In the mid-1970s, the law school began planning the construction of a new building and needed to decide on an appropriate location. Should it be built on the current site in the city, where it would remain near the downtown law firms? Or should it be moved to Evanston, where it would become physically integrated with the rest of the university?

The downtown location had many prominent supporters. They argued in part that it was cost-effective to locate the new building in the city because the university already owned the land, whereas a large parcel of land would have to be purchased in Evanston if the building were to be built there. Does this argument make economic sense?

No. It makes the common mistake of failing to distinguish accounting costs from economic costs. From an economic point of view, it is very expensive to locate downtown because the opportunity cost of the valuable lakeshore location is high-that property could have been sold for enough money to buy the Evanston land with substantial funds left over.

In the end. Northwestern decided to keep the law school in Chicago. This was a costly decision. It may have been appropriate if the Chicago location was particularly valuable to the law school, but the decision was inappropriate if it was made on the presumption that the downtown land was without cost.

EXAMPLE 7.2 THE OPPORTUNITY COST OF WAITING IN A GASOLINE LINE

As a result of gasoline price controls in the spring of 1980, Chevron gasoline stations in California were required to lower their prices substantially below those of other major gasoline companies.[2] This allowed an experiment to be conducted in which consumers revealed information about the opportunity cost of their time.

In this experiment, 109 customers at one Chevron station and 61 customers at two competing stations nearby were surveyed.[3] The consumers could ei-

[2] This special treatment for Chevron stations occurred because these stations were owned and operated by Standard Oil of California. Stations owned by integrated oil companies such as SoCal were affected by the ceiling, but those operated by franchised dealers were not.

[3] The survey was by Robert T. Deacon and Jon Sonstelie, "Rationing by Waiting and the Value of Time: Results from a Natural Experiment," *Journal of Political Economy* 93 (1985): 627-647.

ther buy high-priced gasoline with little or no wait,or wait almost 15 minutes longer to buy lower-priced Chevron gasoline.

Many respondents chose to wait in line for the lower-priced Chevron gasoline, presumably because they valued their time less than the savings they could obtain when they bought the lower-priced gasoline. Suppose, for example, that a motorist could save $0.25 per gallon by waiting for 20 minutes at the Chevron station, and that there would be no wait at the other stations. If she bought ten gallons of gasoline, the total savings would be $2.50. Because she chose to wait in line, the opportunity cost of her time must be less than $2.50 per 20 minutes, or $7.50 per hour. Suppose another person chose to buy gasoline at one of the stations where there was no waiting. Then the opportunity cost of his time must be at least $7.50 per hour. By using this general approach, and by noting that Chevron patrons bought 53 percent more gasoline than the patrons of the other two stations, we can estimate the opportunity cost of waiting time.

Table 7.1 provides some lower- and upper-bound estimates of the opportunity cost of time, in dollars per hour, obtained from the study. Part-time workers displayed the lowest value of time. They could earn additional money working, but that did not conflict with waiting in gas lines because their schedules were flexible. Students' opportunity costs are relatively high because class work is time consuming and because those students who work part time have relatively inflexible schedules and could be working more rather than waiting in gas lines. For all groups, the opportunity cost of time was found to increase with income. This is not surprising; we would expect that the higher the wage one can earn, the greater the opportunity cost of waiting in line to buy lower-priced gas.

This example shows that consumers' as well as firms' decisions are typically based on economic or opportunity cost and not on accounting cost. Everyone would have saved money at the Chevron gas station, and thus made an accounting profit, but many people chose not to because the opportunity cost was too high.

TABLE 7.1 Opportunity Cost of Time

Category	Lower Bound	Upper Bound
Students	$7.15	$10.96
Part-time workers	3.52	5.39
Income $20,000–$30,000	6.51	9.44
Income $30,000–$40,000	8.93	13.70
Income over $40,000	11.26	17.26

7.2 *Cost in the Short Run*

In the short run, some of the firm's inputs to production are fixed, while others can be varied as the firm changes its output. Various measures of the cost of production can be distinguished on this basis.

Total Cost (*TC*) *The total cost of production* has two components: *the fixed cost* FC, which is borne by the firm whatever level of output it produces, and the *variable cost* VC, which varies with the level of output. Depending on circumstances, fixed cost may include expenditures for plant maintenance, insurance, and perhaps a minimal number of employees-this cost remains the same no matter how much the firm produces. Variable cost includes expenditures for wages, salaries, and raw materials-this cost increases as output increases.

Fixed cost does not vary with the level of output-it must be paid even if there is no output. It can be eliminated only by shutting down altogether. (In Chapter 8 we will see that a firm may decide to go out of business and thereby forgo its (future) outlays on fixed costs.)

To decide how much to produce, managers of firms need to know how variable cost increases with the level of output. To address this issue, we need to develop some additional cost measures. We will use a specific example that typifies the cost situation of many firms. After we explain each of the cost concepts, we will describe how they relate to our previous analysis of the firm's production process.

The data in Table 7.2 describe a firm with a fixed cost of $50. Variable cost increases with output, as does total cost. The total cost is the sum of the fixed cost in column (1) and the variable cost in column (2). From the figures given in columns (1) and (2), a number of additional cost variables can be defined.

Marginal Cost (*MC*) *Marginal cost-sometimes called incremental cost-is the increase in cost that results from producing one extra unit of output.* Because fixed cost does not change as the firm's level of output changes, marginal cost is just the increase in variable cost that results from an extra unit of output. We can therefore write marginal cost as

$$\mathrm{MC} = \Delta\mathrm{VC}/\Delta Q$$

Marginal cost tells us how much it will cost to expand the firm's output by one unit. In Table 7.2, marginal cost is calculated from either the variable cost (column 2) or the total cost (column 3). For example, the marginal cost of increasing output from 2 to 3 units is $20 because the variable cost of the firm increases from $78 to $98. (Total cost of production also increases by $20, from $128 to $148. Total cost differs from variable cost onlyby the fixed cost, which by definition does not change as output changes.)

Average Cost (*AC*) *Average cost is the cost per unit of output.* Average total cost (ATC) is the firm's total cost divided by its level of output TC/Q. Thus, the

TABLE 7.2 A Firm's Short-Run Costs ($)

Rate of Output	Fixed Cost	Variable Cost	Total Cost	Marginal Cost	Average Fixed Cost	Average Variable Cost	Average Total Cost
	(FC)	(VC)	(TC)	(MC)	(AFC)	(AVC)	(ATC)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
0	50	0	50	—	—	—	—
1	50	50	100	50	50	50	100
2	50	78	128	28	25	39	64
3	50	98	148	20	16.7	32.7	49.3
4	50	112	162	14	12.5	28	40.5
5	50	130	180	18	10	26	36
6	50	150	200	20	8.3	25	33.3
7	50	175	225	25	7.1	25	32.1
8	50	204	254	29	6.3	25.5	31.8
9	50	242	292	38	5.6	26.9	32.4
10	50	300	350	58	5	30	35
11	50	385	435	85	4.5	35	39.5

average total cost of producing at a rate of five units is$36, $180/5. Basically, average total cost tells us the per-unit cost of production. By comparing the average total cost to the price of the product, we can determine whether production is profitable.

ATC has two components. *Average fixed cost* AFC is the fixed cost (column 1) divided by the level of output, FC/Q. For example, the average fixed cost of producing four units of output is $12.50 ($50/4). Because fixed cost is constant, average fixed cost declines as the rate of output increases. *Average variable cost* (AVC) is variable cost divided by the level of output VC/Q. The average variable cost of producing five units of output is $26, $130 divided by 5.

The Determinants of Short-Run Cost

Table 7.2 shows that variable and total costs increase with output. The rate at which these costs increase depends on the nature of the production process, and in particular on the extent to which production involves diminishing returns to variable factors. Recall from Chapter 6 that diminishing returns to labor occurs when the marginal product of labor is decreasing. If labor is the only variable factor, what happens as we increase the firm's rate of output? To produce more output, the firm has to hire more labor. Then, if the marginal product of labor decreases rapidly as the amount of labor hired is increased (owing to diminishing returns), greater and greater expenditures must be made to produce output at the faster rate. As a result, variable and total costs increase rapidly as the rate of output is increased. On the other hand, if the marginal product of labor decreases only slightly as the amount of labor is increased, costs will not rise so fast when the rate of output is increased.[4]

Let's look at the relationship between production and cost in more detail by concentrating on the costs of a firm that can hire as much labor as it wishes at a fixed wage w. Recall that marginal cost MC is the change in variable cost for a one-unit change in output (i.e., $\Delta VC/\Delta Q$). But the variable cost is the per-unit cost of the extra labor w times the amount of extra labor ΔL. It follows, then, that

$$MC = \Delta VC/\Delta Q = w\Delta L/\Delta Q$$

The marginal product of labor MP_L is the change in output resulting from a one-unit change in labor input, or $\Delta Q/\Delta L$. Therefore, the extra labor needed to obtain an extra unit of output is $\Delta L/\Delta Q$ = I/MP_L,. As a result,

$$MC = w/MP_L \tag{7.1}$$

Equation (7.1) states that in the short run, marginal cost is equal to the price of the input that is being varied divided by its marginal product. Suppose, for example, that the marginal product of labor is 3 and the wage rate is $30 per hour. Then, one hour of labor will increase output by 3 units, so that 1 unit of output will require ⅓ hour of labor, and will cost $10. The marginal cost of producing that unit of output is $10, which is equal to the wage, $30, divided by the marginal product of labor, 3. A low marginal product of labor means that a large amount of additional labor is needed to produce more output, which leads to a high marginal cost A high marginal product means that the labor requirement is low, as is the marginal cost. More generally, whenever the marginal product of labor decreases, the marginal cost of production increases, and vice versa.[5]

The effect of the presence of diminishing returns in the production process can also be seen by looking at the data on marginal costs in Table 7.2. The marginal cost of additional output is high at first because the first few inputs to production are not likely to raise output much in a large plant with a lot of equipment. However, as the inputs become more productive, the marginal cost decreases substantially. Finally, marginal cost increases again for relatively high levels of output, owing to the effect of diminishing returns.

The law of diminishing returns also creates a direct link between the average variable cost of production and the average productivity of labor. Average variable cost AVC is the variable cost per unit of output, or VC/Q. When L units of labor are used in the production process, the variable cost is wL. Thus,

$$AVC = wL/Q$$

[4] We are implicitly assuming that labor is hired in competitive markets, so that the payment per unit of factor used is the same no matter what the firm's output.

[5] With two or more variable inputs, the relationship is more complex, but still the greater the productivity of factors/the less the variable cost that the firm must incur to produce any given level of output

Recall from Chapter 6 that the average product of labor APL is given by the output per unit of input *Q/L*. As a result

$$\text{AVC} = w/\text{AP}_L \tag{7.2}$$

Since the wage rate is fixed from the firm's perspective, there is an inverse relationship between average variable cost and the average product of labor. Suppose, for example, that the average product of labor is 5 and the wage rate is $30 per hour. Then, each hour of labor will increase output on average by 5 units, so that each unit of output will require ½ hour of labor, and will cost $6. The average variable cost of producing each unit of output is $6, which is equal to the wage, $30, divided by the average product of labor, 5. A lower marginal product of labor means that a lot of labor is needed to produce the firm's output, which leads to a high average variable cost. A high average product of labor means that the labor required for production is low, as is the average variable cost.

We have seen that with both marginal cost and average variable cost, there is a direct link between the productivity of factors of production and the costs of production. Marginal and average product tell us about the relationship between inputs and output. The comparable cost variables tell us about the budgetary implications of that production information.

The Shapes of the Cost Curves

Figure 7.1 shows two sets of continuous curves that approximate the cost data in Table 7.2. The fixed cost, variable cost, and total cost curves are shown in Figure 7.1a. Fixed cost FC does not vary with output and is shown as a horizontal line at $50. Variable cost VC is zero when output is zero, and then increases continuously as output increases. The total cost curve TC is determined by vertically adding the fixed cost curve to the variable cost curve. Because fixed cost is constant, the vertical distance between the two curves is always $50.

Figure 7.1b shows the corresponding set of marginal and average variable cost curves.[6] Since total fixed cost is $50, the average fixed cost curve AFC falls continuously from $50 toward zero. The shape of the remaining short-run cost curves is determined by the relationship between the marginal and average cost curves. Whenever marginal cost lies below average cost, the average cost curve falls. Whenever marginal cost lies above average cost, the average cost curve rises. And when average cost is at a minimum, marginal cost equals average cost. Marginal and average costs are another example of the average-marginal relationship described in Chapter 6 (with respect to marginal and

[6] Because the marginal cost represents the change in cost associated with a change in output, we have plotted the marginal cost curve associated with the first unit of output by setting output equal to ½, for the second unit by setting output equal to 1½, and so on.

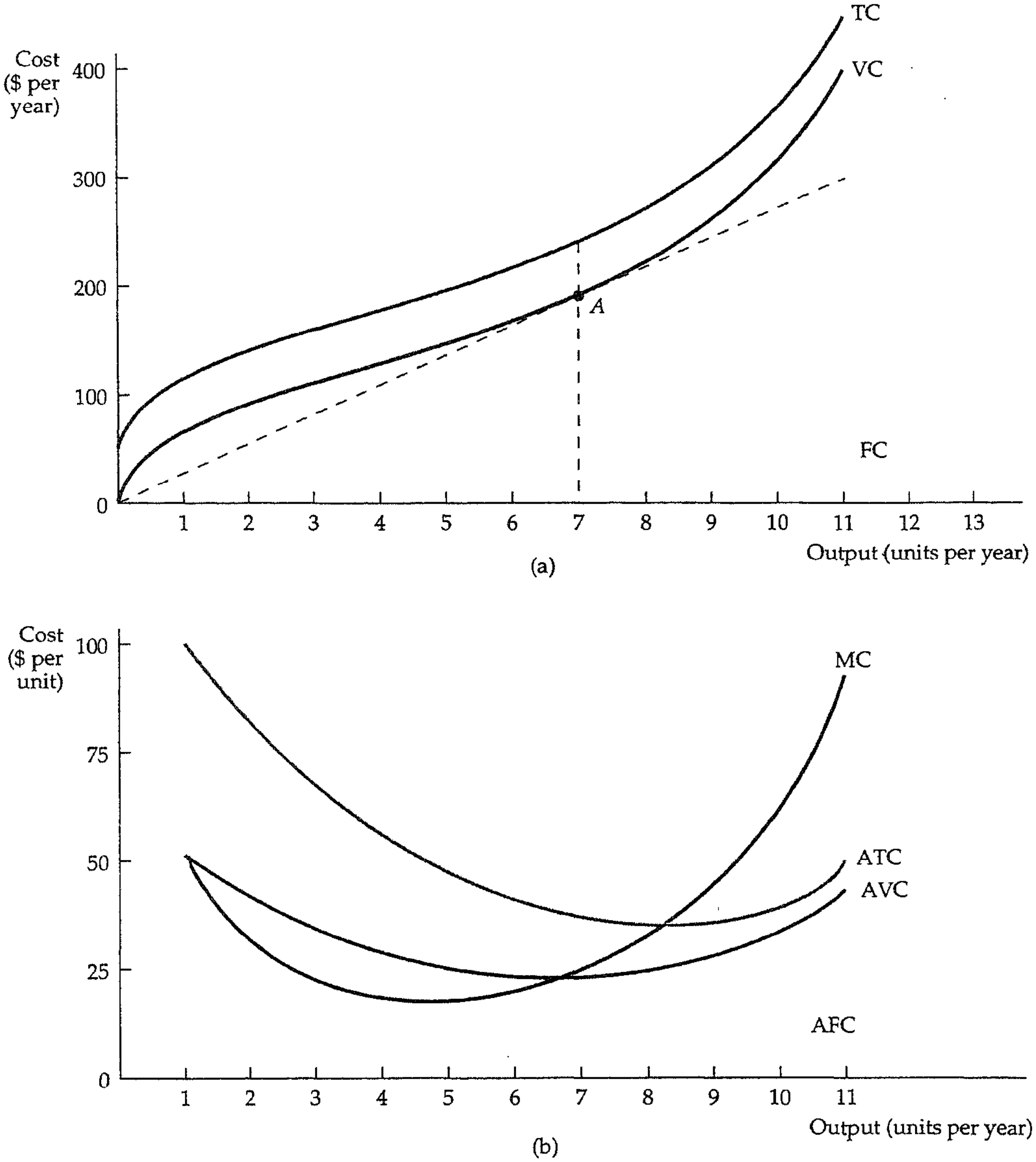

FIGURE 7.1 Cost Curves for a Firm. In (a) total cost TC is the vertical sum of fixed cost FC and variable cost VC. In (b) average total cost ATC is the sum of average variable cost AVC and average fixed cost AFC. Marginal cost MC crosses the average variable cost and average total cost curves at their minimum points.

average product). For example, at $20 marginal cost is below the average variable cost of $25, and the average is lowered. But when marginal cost is $30, which is greater than average variable cost ($25), the average increases. Finally, when marginal cost ($25) and average cost ($25) are the same, the average variable cost remains unchanged (at $25).

The ATC curve shows the average total cost of production. Since average total cost is the sum of average variable cost and average fixed cost and the AFC curve declines everywhere, the vertical distance between the ATC and AVC curves decreases as output increases. The AVC cost curve achieves its minimum point at a lower output than the ATC curve. This follows because MC = AVC at its minimum point, and MC = ATC at its minimum point. Since ATC is always greater than AVC and the marginal cost curve MC is rising, the minimum point of the ATC curve must lie above and to the right of the minimum point of the AVC curve.

Another way to see the relationship between the total cost curves and the average and marginal cost curves is- to consider the ray from origin to point *A* in Figure 7.1a. In that figure the slope of the ray measures average variable cost (a total cost of $175 divided by an output of 7, or a cost per unit of $25). Since the slope of the VC curve is the marginal cost (it measures the change in variable cost as output increases by one unit), the tangent to the VC curve at A is the marginal cost of production when output is 7. At A this marginal cost of $25 is equal to the average variable cost of $25, since average variable cost is minimized at this output.[7]

Note that the *firm's* output is measured as a flow; the firm produces a certain number of units *per year.* Hence its total cost is a flow, for example, some number of dollars per year. (Average and marginal costs, however, are measured in dollars' *per unit.)* For simplicity, we will often drop the time reference, and refer to total cost in dollars and output in units. But you should remember that a firm's production of output and expenditure of cost occur over some time period. Also, for simplicity, we will often use *cost* (C) to refer to total cost. Likewise, unless noted otherwise, we will use *average cost* (AC) to refer to average total cost.

Marginal and average cost are important concepts. As we will see in Chapter 8, they enter critically into the firm's choice of output level. Knowledge of short-run costs is particularly important for firms that operate in an environment in which demand conditions fluctuate considerably. If the firm is currently producing at a level of output at which marginal cost is sharply increasing, and demand may increase in the future, the firm might want to expand its production capacity to avoid higher costs.

[7] The relationships just described hold only approximately when we are describing discrete rather than infinitesimal changes in output. Thus, at an output of 8, average total cost is approximately, but not identically, equal to marginal cost.

7.3 *Cost in the Long Run*

In the long run, the firm can change all its inputs. In this section we show how to choose the combination of inputs that minimizes the cost of producing a given output. We will also examine the relationship between long-run cost and the level of output.

The Cost-Minimizing Input Choice

Let's begin by considering a fundamental problem that all firms face: *how to select inputs to produced given output at minimum cost.* For simplicity, we will work with two variable inputs: labor (measured in hours of work per year) and capital (measured in hours of use of machinery per year). We assume that both labor and capital can be hired (or rented) in competitive markets. The price of labor is the wage rate *w,* and the price of capital is the rental rate for machinery *r.* We assume that capital is rented rather than purchased, so that we can put all business decisions on a comparable basis. For example, labor services might be hired at a wage of $12,000 per year, or capital might be "rented"[7] for $75,000 per machine per year.

Because capital and labor inputs are hired in competitive factor markets, we can take the price of these inputs as fixed. We can then focus on the firm's optimal combination offactors, without worrying about whether large purchases will cause the price of an input to increase.[8]

The Isocost Line

We begin by looking at the cost of hiring factor inputs, which can be represented by a firm's isocost lines. An *isocost line* includes all possible combinations of labor and capital that can be purchased for a given total cost. To see what an isocost line looks like, recall that the total cost C of producing any particular output is given by the sum of the firm's labor cost wL and its capital cost rK:

$$C = wL + rK \tag{7.3}$$

For each different level of total cost, equation (7.3) describes a different isocost line. For example, in Figure 7.2, the isocost line Co describes all possible combinations of inputs that cost Co to purchase.

If we rewrite the total cost equation (7.3) as an equation for a straight line, we get:

[8] This might happen because of overtime or a relative shortage of capital equipment. We discuss the possibility of a relationship between the prices of factor inputs and the quantities demanded by a firm in Chapter 14.

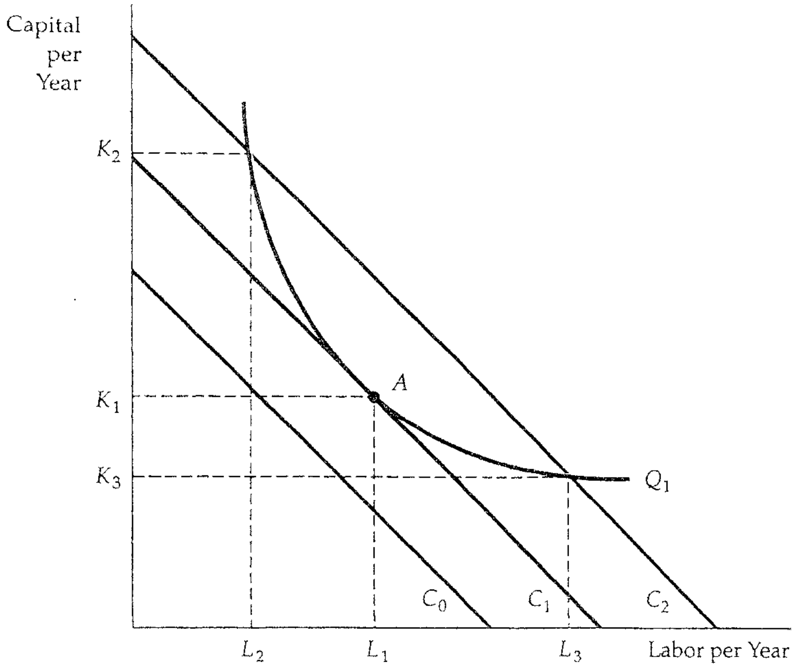

FIGURE 7.2 Producing a Given Output at Minimum Cost. Isocost curves describe the combination of inputs to production that cost the same amount to the firm. Isocost curve C1 is tangent to isoquant Q1 at A and shows that output Q1 can be produced at minimum cost with labor input K1 and capital input L1. Other input combinations-L2, K2 and L3, K3-yield the same output at higher cost.

$$K = C/r - (w/r)L$$

It follows that the isocost line has a slope of $\Delta K/\Delta L = -(w/r)$, which is the ratio of the wage rate to the rental cost of capital. This slope is similar to the slope of the budget line that the consumer faces (because it is determined solely by the prices of the goods in question, whether inputs or outputs). It tells us that if the firm gave up a unit of labor (and recovered w dollars in cost) to buy w/r units of capital at a cost of r dollars per unit, its total cost of production would remain the same. For example, if the wage rate were $10 and the rental cost of capital $5, the firm could replace one unit of labor with two units of capital, with no change in total cost.

Choosing Inputs

Suppose we wish to produce output level Q_1. How can we do this at minimum cost? Look at the firm's production isoquant, labeled Q_1, in Figure 7.2. The problem is to choose the point on this isoquant that minimizes total costs.

Figure 7.2 illustrates the solution to this problem. Suppose the firm were to spend C_0 on inputs. Unfortunately, no combination of inputs can be purchased for expenditure C_0 that will allow the firm to achieve output Q_1. Output Q_1 can be achieved with the expenditure of C_2, however, either by using K_2 units of capital and L_2 units of labor, or by using K_3 units of capital and 13 units of labor. But C_2 is not the minimum cost. The same output Q_1 can be produced more cheaply than this, at a cost of C_1, by using K_1 units of capital and L_1 units of labor. In fact, isocost line C_1 is the lowest isocost line that allows output Q_1 to be produced. The point of tangency of the isoquant Q_1 and the isocost line C_1 at point *A* tells us the cost-minimizing choice of inputs, L_1 and K_1, which can be read directly from the diagram. At this point, the slopes of the isoquant and the isocost line are just equal.

When the expenditure on all inputs increases, the slope of the isocost line does not change (because the prices of the inputs have not changed), but the intercept increases. Suppose, however, that the price of one of the inputs, such as labor, were to increase. Then, the slope of the isocost line -(*w*/*r*) *would* increase in magnitude, and the isocost line would become steeper. Figure 7.2

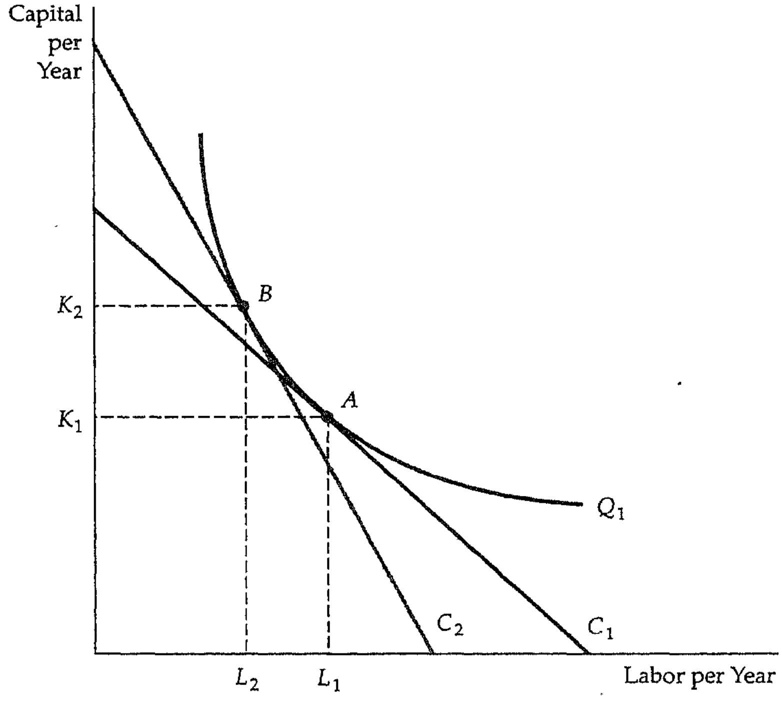

FIGURE 7.3 Input Substitution When an Input Price Changes. Facing an isocost curve C_1, the firm produces output Q_1 using L_1 units of labor and K_1 units of capital. TO When the price of labor increases, the isocost curves become steeper. Output Q_1 is now produced at *B* on isocost curve C_2, by using L_2 units of labor and K_2 units of capital.

shows this. Initially, the isocost line is C_1, and the firm minimizes its costs of producing output Q_1 at A by using L_1 units of labor and K_1 units of capital. When the price of labor increases, the isocost line becomes steeper. The isocost line C_2 reflects the higher price of labor. Facing this higher price of labor, the firm minimizes its cost of producing output Q_1 by producing at B, using L_2 units of labor and K_2 units of capital. The firm has responded to the higher price of labor by substituting capital for labor in the production process.

How does this relate to the firm's production process? Recall that in our analysis of production technology, we showed that the marginal rate of technical substitution MRTS of labor for capital is the negative of the slope of the isoquant, and is equal to the ratio of the marginal products of labor and capital.

$$\text{MRTS} = -\Delta K/\Delta L = \text{MP}_L/\text{MP}_K \tag{7.4}$$

Above, we noted that the isocost line has a slope of $\Delta K/\Delta L = -w/r$. It follows that when a firm minimizes the cost of producing a particular output, the following condition holds:

$$\text{MP}_L/\text{MP}_K = w/r$$

Rewriting this condition slightly,

$$\text{MP}_L/w = \text{MP}_K/r \tag{7.5}$$

Equation (7.5) tells us that when cost is minimized, each dollar of input added to the production process will add an equivalent amount to output. Assume, for example, that the wage rate is $10 and the rental rate on capital is $2. If the firm chooses inputs so that the marginal product of labor and the marginal product of capital are equal to ten, it will want to hire less labor and rent more capital because capital is five times less expensive than labor. The firm can minimize its cost only when the production of an additional unit of output costs the same regardless of which additional input is used.

EXAMPLE 7.3 THE EFFECT OF EFFLUENT FEES ON FIRMS' INPUT CHOICES

Steel plants are often built on or near a river. A river offers readily available, inexpensive transportation for both the iron ore that goes into the production process and the finished steel itself. A river also provides a cheap method of disposing of by-products of the production process, called effluent. For example, a steel plant processes its iron ore for use in blast furnaces by grinding taconite deposits into a fine consistency. During this process, the ore is extracted by a magnetic field as a flow of water and fine ore passes through the plant. One by-product of this process-fine taconite particles-can be dumped in the river at relatively little cost to the firm, whereas alternative removal methods or private treatment plants are relatively expensive.

Because the taconite particles are a nondegradable waste that can harm vegetation and fish, the Environmental Protection Agency (EPA) has imposed an effluent fee-a per-unit fee that the steel firm must pay for the effluent that goes into the river. How should the manager of the firm respond to the imposition of this effluent fee to minimize the costs of production?

Suppose that without regulation the steel firm is producing 2000 tons of steel per month, while using 2000 machine-hours of capital and 10,000 gallons of water (which contains taconite particles when returned to the river). The manager of the firm estimates that a machine-hour costs $40, and dumping each gallon of waste water in the river costs the firm $10. (The total cost of production is therefore $180,000: $80,000 for capital and $100,000 for waste water.) How should the manager respond to an EPA-imposed effluent fee of $10 per gallon of waste water dumped?

Figure 7.4 shows the cost-minimizing response. The vertical axis measures the firm's input of capital in machine-hours per month, and the horizontal axis measures the quantity of waste water in gallons per month. First, consider how .the firm produces when there is no effluent fee. Point A represents

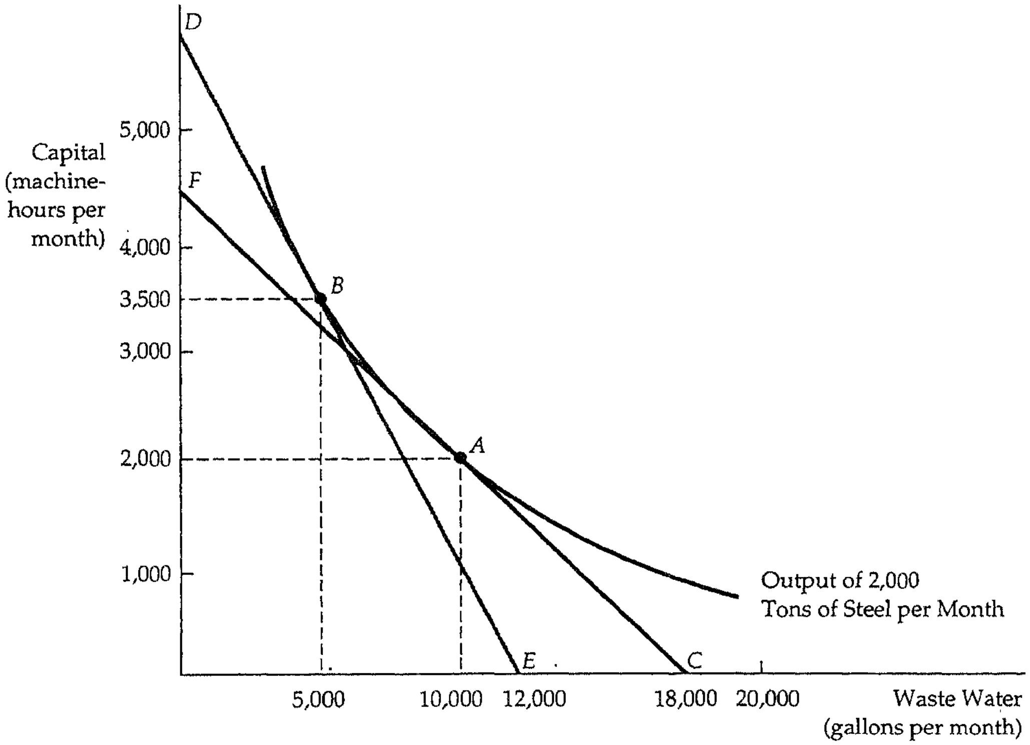

FIGURE 7.4 The Cost-Minimizing Response to an Effluent Fee. When the firm is not charged for dumping its waste water in a river, it chooses to produce a given output using 10,000 gallons of waste water and 2000 machine-hours of capital at A However, an effluent fee raises the cost of waste water, shifts the isocost curve from *FC* to *DE*, and causes the firm to produce at *B*, with much less effluent.

the input of capital and the level of waste water that allows the firm to produce its quota of steel at minimum cost. Because the firm is minimizing cost, A lies on the isocost line *FC*, which is tangent to the isoquant. The slope of the isocost line is equal to -$10/$40 = - 0.25 because a unit of capital costs four times more than a unit of waste water.

When the effluent fee is imposed, the cost of waste water increases from $10 per gallon to $20, because for every gallon of waste water (which costs $10), the firm has to pay the government an additional $10. The effluent fee increases the cost of waste water relative to capital. To produce the same output at the lowest possible cost, the manager must choose the isocost line with a slope of -$20/$40 = -0.5, which is tangent to the isoquant. In Figure 7.4, *DE* is the appropriate isocost line, and *B* gives the appropriate choice of capital and waste water. The move from *A* to *B* shows that with an effluent fee the use of an alternative production technology, which emphasizes the use of capital (3500 machine-hours) and uses less waste water (5000 gallons), is cheaper than the original process, which did not emphasize recycling. (The total cost of production has increased to $240,000: $140,000 for capital, $50,000 for waste water, and $50,000 for the effluent fee.)

We can learn two lessons from this decision. First, the more easily factors can be substituted in the production process, that is, the more easily the firm can deal with its taconite particles without using the river for waste treatment, the more effective the fee will be in reducing effluent. Second, the greater the degree of substitution, the less the firm will have to pay. In our example, the fee would have been $100,000 had the firm not changed its inputs. However, the steel company pays only a $50,000 fee by moving production from A to *B*.

Cost Minimization with Varying Output Levels

In the previous section we saw how a cost-minimizing firm selects a combination of inputs to produce a given level of output. Now we extend this analysis to see how the firm's costs depend on its output level. To do this we determine the firm's cost-minimizing input quantities for each output level, and then calculate the resulting cost.

The cost-minimization exercise yields a result such as that shown in Figure 7.5. Each of the points *A*, *B*, *C*, *D*, and *E* represents a tangency between an isocost curve and an isoquant for the firm. The curve which moves upward and to the right from the origin, tracing out the points of tangency, is the firm's *expansion path*. The expansion path describes the combinations of labor and capital that the firm will choose to minimize costs for every output level. So long as the use of both, inputs increases as output increases, the curve will look approximately as shown in Figure 7.5. The firm's expansion path shows the lowest long-run total cost of producing each level of output.

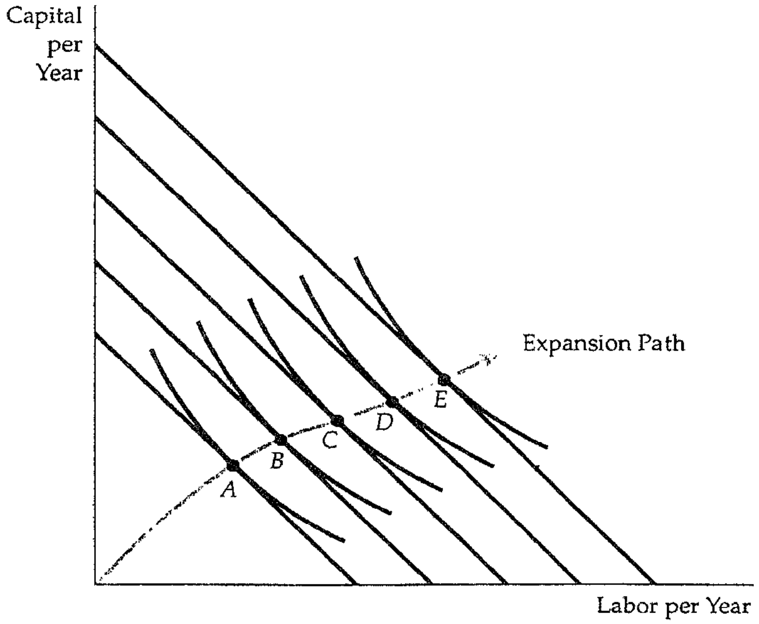

FIGURE 7.5 A Firms Expansion Path. The expansion path illustrates the least-cost rnmbinations oflabor and capital that can be used to produce each level of output in the long run when both inputs to production can be varied.

7.4 *Long-Run Versus Short-Run Cost Curves*

We saw earlier (see Figure 7.1) that short-run average cost curves are U-shaped. We will see that long-run average cost curves can also be U-shaped, but different economic factors explain the Shapes of these curves. In this section, we discuss long-run average and marginal cost curves and highlight the differences between these curves and their short-run counterparts.

The Inflexibility of Short-Run Production

Recall that in the long run all inputs to the firm are variable, because its planning horizon is long enough to allow for a change in plant size. This added flexibility allows the firm to produce at a lower average cost than in the short run. To see why, we might compare the situation in which capital and labor are both flexible to the case in which capital is fixed in the short run.

Figure 7.6 shows the firm's production isoquants. Suppose capital is fixed at a level K_1 in the short run. To produce output Q_1, the firm would minimize costs by choosing labor equal to L_1, corresponding to the point of tangency with the isocost line *AB*. The inflexibility appears when the firm decides to increase its output to Q_2. If capital were not fixed, it would produce this output with capital K_2 and labor L_2 Its cost of production would be reflected by isocost line CD. However, the fixed capital forces the firm to increase its output by using capital K_1 and labor L_3 at P. Point *P* lies on the isocost line *EF*, which represents a higher cost than isocost line *CD*. The cost of production is higher when capital is fixed because the firm is unable to substitute relatively inexpensive capital for more costly labor when it expands its production.

Long-Run Average Cost

In the long run, the ability to change the amount of capital allows the firm to reduce costs. To see how costs vary as the firm moves along its expansion path

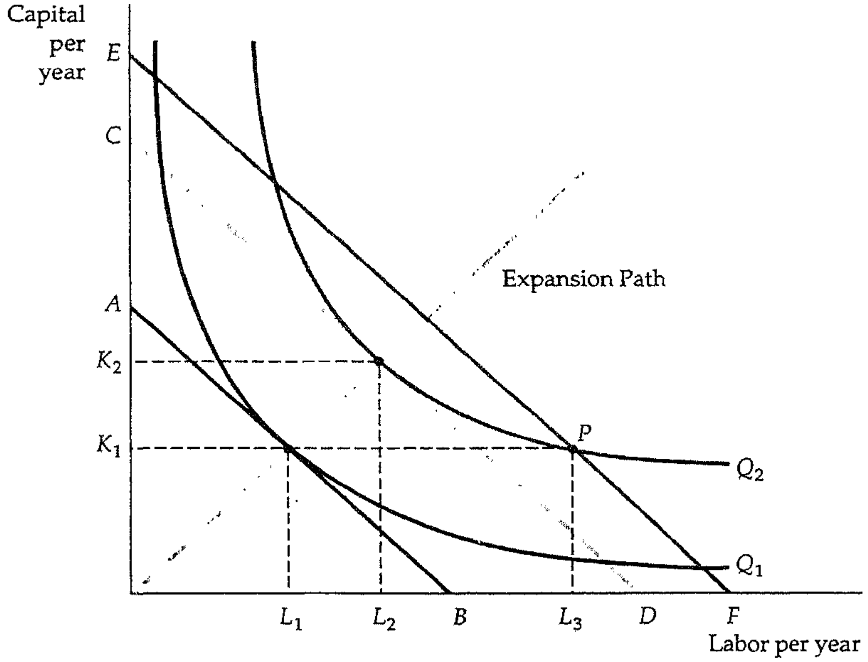

FIGURE 7.6 The Inflexibility of Short-Run Production. When a firm operates in me short run, its cost of production may not be minimized because of inflexibility in the use of capital inputs. Output is initially at level Q_1. In the short run, output Q_2 can be produced only by increasing labor from L_1 to L_3 because capital is fixed at K_1. In the long run, the same output can be produced more cheaply by increasing labor from L_1 to L_2 and capital from K_1 to K_2.

in the long run, we can look at the long-run average and marginal cost curves.[9] The most important determinant of the shape of the long-run average and marginal cost curves is whether there are increasing, constant, or decreasing returns to scale. Suppose, for example, that the firm's production process exhibits constant returns to scale at all levels of output. Then a doubling of inputs leads to a doubling of output. Because input prices remain unchanged as output increases, the average cost of production must be the same for all levels of output.

Suppose instead that the firm's production process is subject to increasing returns to scale. A doubling of inputs leads to more than a doubling of output. Then the average cost of production falls with output because a doubling of costs is associated with a more than twofold increase in output. By the same logic, when there are decreasing returns to scale, the average cost of production must be increasing with output.

In the last chapter, we saw that in the long run most firms' production technologies first exhibit increasing returns to scale, then constant returns to scale, and eventually decreasing returns to scale. Figure 7.7 shows a typical long-run average cost curve LAC consistent with this description of the production process. The long-run average cost curve is U-shaped, just like the short-run average cost curve, but the source of the U-shape is increasing and decreasing returns to scale, rather than diminishing returns to a factor of production.

The long-run marginal cost curve LMC is determined from the long-run average cost curve; it measures the change in long-run total costs as output is increased incrementally. LMC lies below the long-run average cost curve when LAC is falling, and above the long-run average cost curve when LAC is rising. The two curves intersect at *A*, where the long-run average cost curve achieves its minimum. And in the special case in which LAC is constant, LAC and LMC are equal.

Economies and Diseconomies of Scale

In the long run, it may be in the firm's interest to change the input proportions as the level of output changes. When input proportions do change, the concept of returns to scale no longer applies. Rather, we say that a firm enjoys *economies of scale* when it can double its output for less than twice the cost. Correspondingly, there are *diseconomies of scale* when a doubling of output requires more than twice the cost. The term *economies of scale* includes increasing returns to scale as a special case, but it is more general because it allows input combinations to be altered as the firm changes its level of production. In this more general setting, a U-shaped long-run average cost curve is con-

[9] We saw that in the short run the shapes of the average and marginal cost curves were determined primarily by diminishing returns. As we showed in Chapter 6, diminishing returns to each factor are consistent with constant (or even increasing) returns to scale.

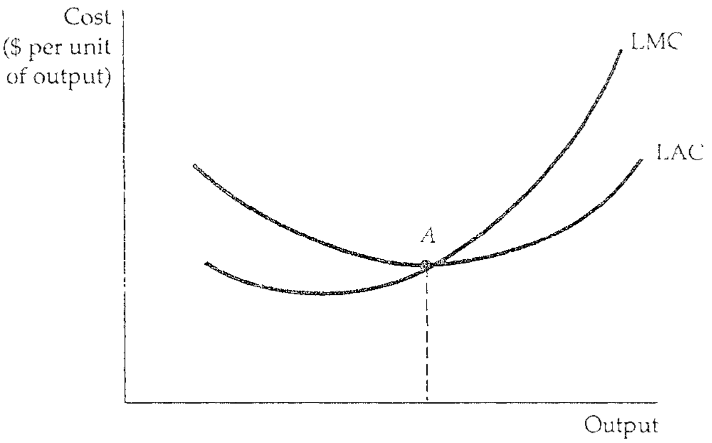

FIGURE 7.7 Long-Run Average and Marginal Cost. When a firm is producing at an output at which the long-run average cost LAC is falling, the long-run marginal cost LMC is less than long-run average cost. When long-run average cost is increasing, long-run marginal cost is greater than long-run average cost.

sistent with the firm facing economies of scale for relatively low output levels and diseconomies of scale for higher levels.

Economies of scale are often measured in terms of a cost-output elasticity, *E*c. Ec 1 S the percentage change in the average cost of production resulting from a one percent increase in output:

$$E_c = (\Delta C/C)/(\Delta Q/Q) \quad (7.6)$$

To see how *E*c relates to our traditional measures of cost, rewrite equation (7.6) as follows:

$$E_c = (\Delta C/\Delta Q)/(C/Q) = MC/AC \quad (7.7)$$

Clearly, *E*c is equal to one when marginal and average costs are equal; then costs increase proportionately with output, and there are neither economies nor diseconomies of scale (constant returns to scale would apply if input proportions were fixed). When there are economies of scale (costs increase less than proportionately with output), marginal cost is less than average cost (both are declining), and *E*c is less than one. Finally, when there are diseconomies of scale, marginal cost is greater than average cost, and Ec is greater than one.

The Relationship Between Short-Run and Long-Run Cost

Figures 7.8 and 7.9 show the relationship between short-run and long-run cost. Assume a firm is uncertain about the future demand for its product and is considering three alternative plant sizes. The short-run average cost curves for

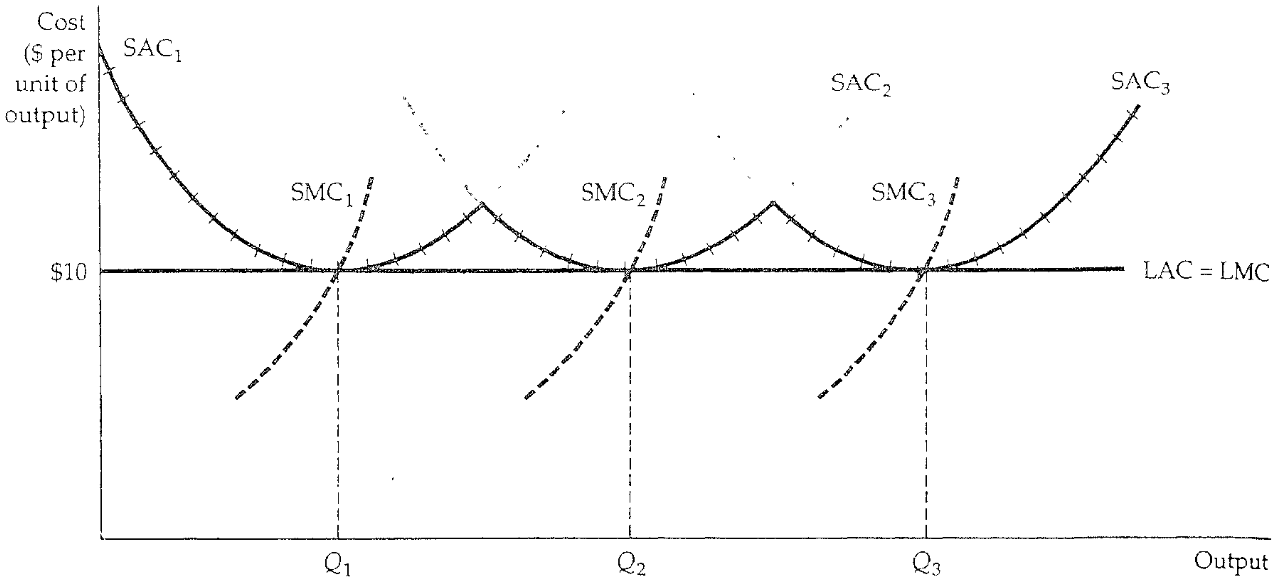

FIGURE 7.8 Long-Run Cost with Constant Returns to Scale. The long-run average cost curve LAC, which is identical to the long-run marginal cost curve LMC, is the envelope of the short-run average cost curves (SAC_1, SAC_2, and SAC_3 are shown). With constant returns to scale, the long-run average cost curve consists of the minimum points of the short-run average cost curves.

the three plants are given by SAC_1, SAC_2 and SAC_3 in Figure 7.8. The decision is important because, once built, the firm may not be able to change the plant size for some time.

Figure 7.8 shows the case in which there are constant returns to scale in the long run. If the firm were expecting to produce Q_1 units of output, then it should build the smallest plant. Its average cost of production would be $10; this is the minimum cost because the short-run marginal cost SMC crosses short-run average cost SAC when both equal $10. If the firm is to produce Q_2 units of output, the middle-sized plant is best, and its average cost of production is again $10. If it is to produce Q_3, it moves to the third plant. With only these plant sizes, any production choice between Q_1 and Q_2 will entail an increase in the average cost of production, as will any level of production between Q_2 and Q_3.

What is the firm's long-run cost curve? In the long run, the firm can change the size of its plant, so if it was initially producing Q_1 and wanted to increase output to Q_2 or Q_3, it could do so with no increase in average cost. The long-run average cost curve is therefore given by the cross-hatched portions of the short-run average cost curves because these show the minimum cost of production for any output level. The long-run average cost curve is the *envelope* of the short-run average cost curves-it envelops or surrounds the short-run curves.

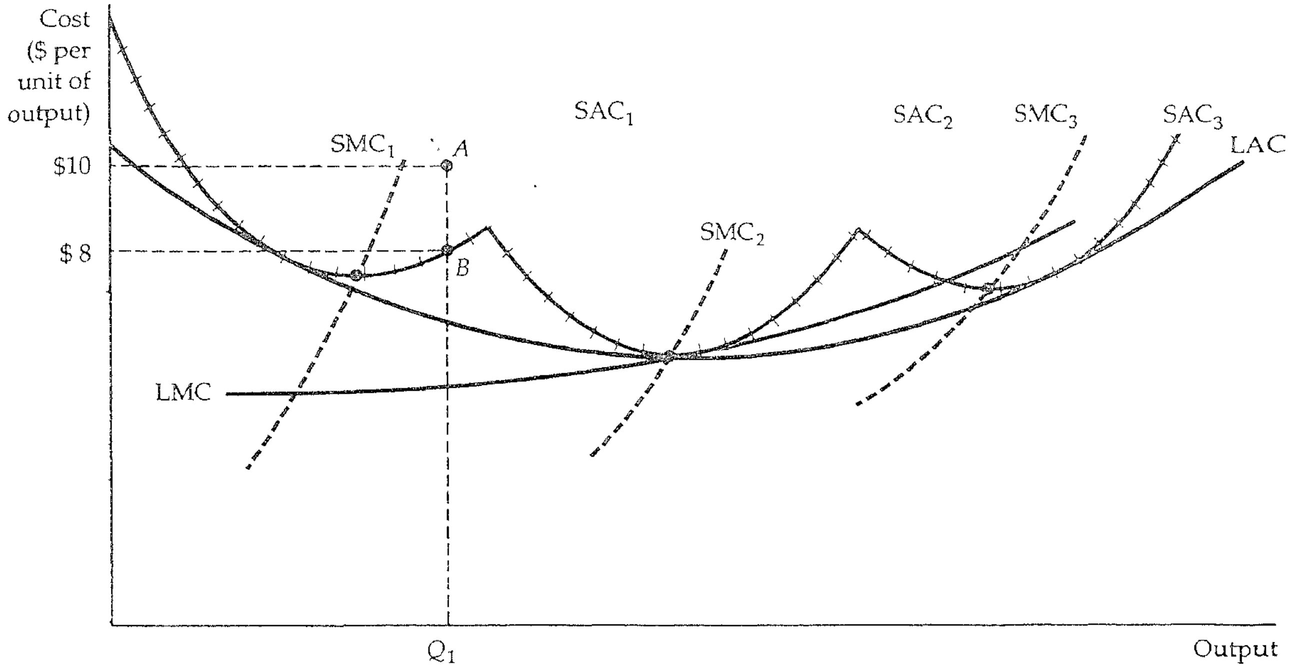

FIGURE 7.9 Long-Run Cost with Economies and Diseconomies of Scale. The long-run average cost curve LAC is the envelope of the short-run average cost curves (SAC1, SAC2, and SAC3). With economies and diseconomies of scale, the minimum points of the short-run average cost curves do not He on the long-run average cost curve.

Now suppose there are many choices of plant size, each of which has a short-run average cost curve that has its minimum at the $10 level. Again, the long-run average cost curve is the envelope of the short-run curves. In Figure 7.8 it is the straight line LAC. Whatever the firm wants to produce, it can choose the plant size (and the mix of capital and labor) that allows it to produce that output at the minimum average cost of $10.

With economies or diseconomies of scale, the analysis is essentially the same, but the long-run average cost curve is no longer a horizontal line. Figure 7.9 illustrates the typical case in which three plant sizes are possible; the minimum average cost is lowest for a medium-sized plant. The long-run average cost curve, therefore, exhibits economies of scale initially, but at higher output levels it exhibits diseconomies. Once again, the cross-hatched lines show the envelope associated with the three plants.

To clarify the relationship between the short-run and the long-run cost curves, consider a firm that wants to produce output Q_1 in Figure 7.9. If it builds a small plant, the short-run average cost curve SAC$_1$ is relevant, so that the average cost of production (at *B* on SAC$_1$) is $8. A small plant is a better choice than a medium-sized plant with an average cost of production of $10 (*A* on curve SAC$_2$). Point *B* would, therefore, become one point on the long-run cost function when only three plant sizes are possible. If plants of other sizes could be built, and at least one size allowed the firm to produce Q_1 at less than $8 per unit, then *B* would no longer be on the long-run cost curve.

In Figure 7.9, the envelope that would arise if plants of any size could be built is given by the LAC curve, which is U-shaped. Note, once again, that the LAC curve never lies above any of the short-run average cost curves- Also note that the points of minimum average cost of the smallest and largest plants do *not* lie on the long-run average cost curve because there are economies and diseconomies of scale in the long run. For example, a small plant operating at minimum average cost is not efficient because a larger plant can take advantage of increasing returns to scale to produce at a lower average cost.

Finally, note that the long-run marginal cost curve LMC is not the envelope of the short-run marginal cost curves. Short-run marginal costs apply to a particular plant; long-run marginal costs apply to all possible plant sizes. Each point on the long-run marginal cost curve is the short-run marginal cost associated with the most cost-efficient plant.

7.5 *Production with Two Outputs-Economies of Scope*

Many firms produce more than one product. Sometimes a firm's products are closely linked to one another-a chicken farm produces poultry and eggs, an automobile company produces automobiles and trucks, and a university produces teaching and research. Other times,firms produce products that are physically unrelated. In both cases, however, a firm is likely to enjoy production or cost advantages when it produces two or more products. These advantages could result from the joint use of inputs or production facilities, joint marketing programs, or possibly the cost savings of a common administration. In some cases, the production of one product gives an automatic and unavoidable by-product that is valuable to the firm. For example, sheet metal manufacturers produce scrap metal and shavings they can sell.

To study the economic advantages of joint production, let's consider an automobile company that produces two products, cars and tractors. Both products use capital (factories and machinery) and labor as inputs. Cars and tractors are not typically produced at the same plant, but they do share management resources, and both rely on similar machinery and skilled labor. The managers of the company must choose how much of each product to produce. Figure 7.10 shows two *product transformation curves. Each* curve shows the various combinations of cars and tractors that can be produced with a given input of labor and machinery. Curve O_1 describes all combinations of the two outputs that can be produced with a relatively low level of inputs, and curve O_2 describes the output combinations associated with twice the inputs.

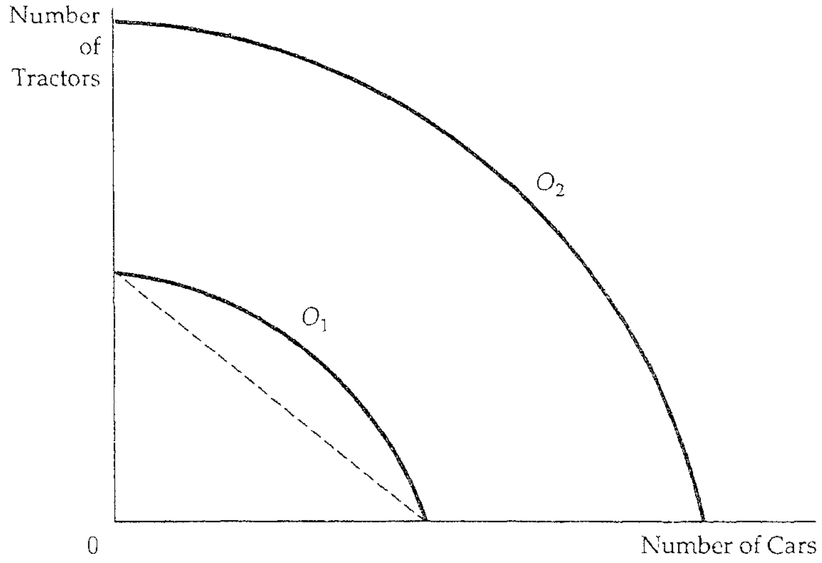

FIGURE 7.10 Product Transformation Curve. me product transformation curve describes the different combinations of two outputs that can be produced with a fixed amount of production inputs. The product transformation curves O_1 and O_2 are bowed out (or concave) because there are economies of scope in production.

The product transformation curve has a negative slope because to get more of one output, the firm must give up some of the other output. For example, a firm that emphasizes car production will devote less of its resources to producing tractors. In this case, curve O_2 lies twice as far from the origin as curve O_1, signifying that this firm's production process exhibits constant returns to scale in the production of both commodities.[10]

If curve O_1 were a straight line, joint production would entail no gains (or losses). One smaller company specializing in cars and another in tractors would generate the same output as the single company that produces both. However, the product transformation curve is bowed outward (or *concave)* because joint production usually has advantages that enable a single company to produce more cars and tractors with the same resources than would two companies producing each product separately. These production advantages involve the joint sharing of inputs. A single management is often able to schedule and organize production and to handle accounting and financial aspects more effectively than separate managements could.

[10] Our discussion would be more complex if it incorporated the possibility of diseconomies or economies of scale. For a more general analysis of economies of scope, see Elizabeth E. Bailey and Ann F. Friedlaender, "Market Structure and Multiproduct Industries: A Review Article," *Journal of Economic Literature* 20 (Sept. 1982): 1024-1048, or John C. Panzar and Robert D. Willig, "Economies of Scope," *American Economic Revieiw* 71 (May 1981): 268-272.

In general, *economies of scope* are present when *the joint output of a single firm is greater than the output that could be achieved by two different firms each producing a single product* (with equivalent production inputs allocated between the two firms). If a firm's joint output is *less* than that which could be achieved by separate firms, then its production process involves *diseconomies of scope.* This could occur if the production of one product somehow conflicted with the production of the second product.

There is no direct relationship between economies of scale and economies of scope. A two-output firm can enjoy economies of scope even if its production process involves diseconomies of scale. Suppose, for example, that manufacturing flutes and piccolos jointly is cheaper than producing both separately. Yet the production process involves highly skilled labor and is most effective if undertaken on a small scale. Likewise, a joint-product firm can have economies of scale for each individual product, yet not enjoy economies of scope. Imagine, for example, a large conglomerate that owns several firms that produce efficiently on a large scale but that do not take advantage of economies of scope because they are administered separately.

The extent to which there are economies of scope can also be determined by studying a firm's costs. If a combination of inputs used by one firm generates more output than two independent firms would produce, then it costs less for a single firm to produce both products than it would cost the independent firms. To measure the degree to which there are economies of scope, we should ask what percentage of the cost of production is saved when two (or more) products are produced jointly rather than individually. Equation (7.8) gives the *degree of economies of scope* (SC) that measures this savings in cost:

$$\mathrm{SC} = \frac{C(Q_1) + C(Q_2) - C(Q_1, Q_2)}{C(Q_1, Q_2)} \tag{7.8}$$

$C(Q_1)$ represents the cost of producing output Q_1, $C(Q_2)$ the cost of producing output Q_2, and $C(Q_1, Q_2)$ the joint cost of producing both outputs. (When the physical units of output can be added, as in the car-tractor example, the expression becomes $C(Q_1 + Q_2)$.) With economies of scope, the joint cost is less than the sum of the individual costs, so that SC is greater than 0. With diseconomies of scope, SC is negative. In general, the larger the value of SC, the greater the economies of scope.

EXAMPLE 7.4 ECONOMIES OF SCOPE IN THE TRUCKING INDUSTRY

Suppose that you are managing a trucking firm that hauls loads of different sizes between cities.[11] In the trucking business, several related but distinct products can be offered depending on the size of the load and the length of

[11] This example is based on Judy S. Wang Chiang and Ann R Friedlaender, "Truck Technology and Efficient Market Structure," *Review of Economics and Statistics* 67 (1985): 250-258.

the haul. First, any load, small or large, can be taken directly from one location to another without intermediate stops. Second, a load can be combined with other loads, which may go between different locations, and eventually be shipped indirectly from its origin to the appropriate destination. And each type of load, partial or full, may involve different lengths of haul.

This raises questions about both economies of scale and economies of scope. The scale question is whether large-scale, small, direct hauls are cheaper and more profitable than individual hauls by small truckers. The scope question is whether a large trucking firm enjoys cost advantages from operating both direct quick hauls and indirect, slower (but less expensive) hauls. Central planning and organization of routes could provide for economies of scope. The key to the presence of economies of scale is the fact that the organization of routes and the types of hauls we have described can be accomplished more efficiently when many hauls are involved. Then it will be more likely that hauls can be scheduled that allow most truckloads to be full, rather than half-full.

Studies of the trucking industry show that economies of scope are present. For example, an analysis of 105 trucking firms in 1976 looked at four distinct outputs: (1) short hauls with partial loads, (2) intermediate hauls with partial loads, (3) long hauls with partial loads, and (4) hauls with total loads. The results indicate that the degree of economies of scope SC was 1.576 for a reasonably large firm. However, the degree of economies of scope falls to 0.104 when the firm. becomes very large. Large firms carry sufficiently large truckloads, so there is usually no advantage to stopping at an intermediate terminal to fill a partial load. A direct trip from the origin to the destination is sufficient. Apparently, however, other disadvantages are associated with the management of very large firms, so the economies of scope get smaller as the firm gets bigger. In any event, the ability to combine partial loads at an intermediate location lowers the firm's costs and increases its profitability.

The study suggests, therefore, that to compete in the trucking industry a firm must be large enough to be able to combine loads at intermediate stopping points.

*7.6 Dynamic Changes in Costs— The Learning Curve

Our discussion has suggested one reason a large firm may have a lower long-run average cost than a small firm-increasing returns to scale in production. It is tempting to conclude that firms that enjoy lower average cost over time are growing firms with increasing returns to scale. But this need not be true. In some firms, long-run average cost may decline over time because workers and managers absorb new technological information as they become more experienced at their jdbs.

As management and labor gain experience with production, the firm's marginal and average cost of producing a given level of output falls for four reasons. First, workers often take longer to accomplish a given task the first few times they do it. As they become more adept, their speed increases. Second, managers learn to schedule the production process more effectively, from the flow of materials to the organization of the manufacturing itself. Third, engineers, who are initially very cautious in their product designs, may gain enough experience to be able to allow for tolerances in design that save cost without increasing defects. Better and more specialized tools and plant organization may also lower cost. Fourth, suppliers of materials may learn how to process materials required by the firm more effectively and may pass on some of this advantage to the firm in the form of lower materials cost.

As a consequence, a firm "learns" over time as cumulative output increases. Managers use this learning process to help plan production and to forecast future costs. Figure 7.11 illustrates this process, in the form of a learning curve. A *learning curve* describes the relationship between a firm's cumulative output and the amount of inputs needed to produce a unit of output.

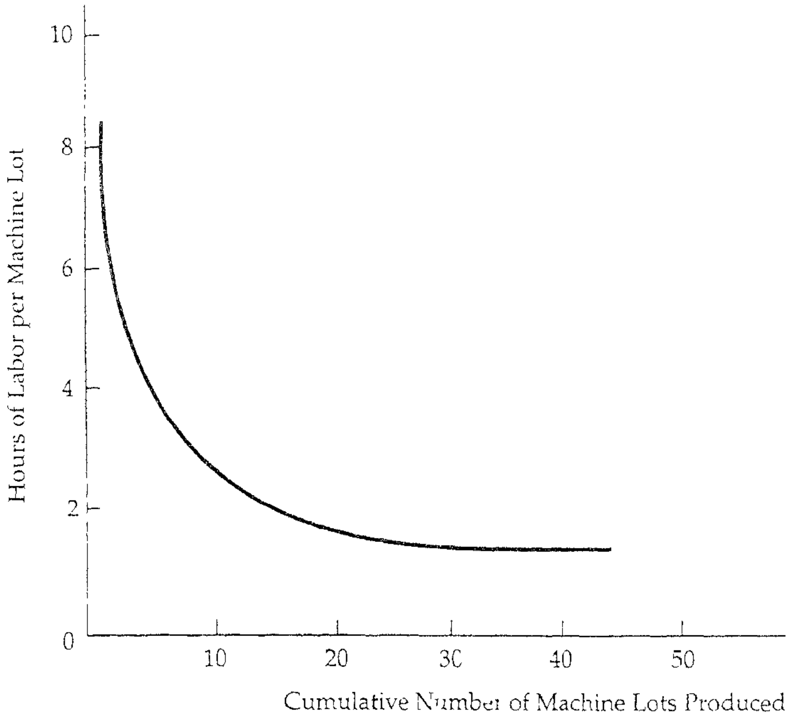

FIGURE 7.11 The Learning Curve. A firm's cost of production may fall over time as the managers and workers become more experienced and more effective at using the available plant and equipment. The learning curve shows the extent to which the hours of labor needed per unit of output (a machine in this case) fall as the cumulative output (number of machines) produced increases.

Figure 7.11 shows a learning curve for the production of machine tools by a manufacturer.[12]The horizontal axis measures the *cumulative* number of lots of machine tools that the firm has produced (a lot is a group of approximately 40 machines), and the vertical axis the number of hours of labor needed to produce each lot. Labor input per unit of output directly affects the firm's cost of production because the fewer the hours of labor needed, the lower the marginal and average cost of production.

The learning curve in the figure is based on the relationship

$$L = A + BN^{-\beta} \tag{7.9}$$

where, N is the cumulative units of output produced, L is the labor input per units of output, and A, B, and β are constants, with A and B positive, and β between 0 and 1. When N is equal to 1, L is equal to $A + B$, so that $A + B$ measures the labor input required to produce the first unit of output. When β equals 0, labor input per unit of output remains the same as the cumulative level of output increases, so there is no learning. When β is positive and N gets larger and larger, L becomes arbitrarily close to A, so that A represents the minimum labor input per unit of output after all learning has taken place.

The larger is β, the more important is the learning effect. With β equal to 0.5, for example, the labor input per unit of output falls proportionally to the square root of the cumulative output. This degree of learning can substantially reduce the firm's production costs as the firm becomes more experienced.

In this machine tool example, the value of β is 0.31. For this particular learning curve, every doubling in cumulative output causes the difference between the input requirement and the minimum attainable input requirement to fall by about 20 percent.[13] As Figure 7.11 shows, the learning curve drops sharply as the cumulative number of lots produced increases to about 20. Beyond an output of 20 lots, the cost savings are relatively small.

Once the firm has produced 20 or more machine lots, the entire effect of the learning curve would be complete, and the usual analysis of cost could be employed. If, however, the production process were relatively new, then relatively high cost at low levels of output (and relatively low cost at higher levels) would indicate learning effects, and not economies of scale. With learning, the cost of production for a mature firm is relatively low irrespective of the scale of the firm's operation. If a firm that produces machine tools in groups (or "lots") knows that it enjoys economies of scale, it should produce its machines in very large lots to take advantage of the lower cost associated with size. If there is a learning curve, the firm can lower its cost by scheduling the production of many lots irrespective of the individual lot size.

Figure 7.12 shows this phenomenon. AC_1 represents the long-run average cost of production of a firm that enjoys economies of scale in production. Thus, the change inl production from A to B along AC_1 leads to lower

[12] See Werner Z. Hirsch, "Manufacturing Progress Functions,"' Review of Economics and Statistics 34 (May 1952): 143-155.

[13] Because $(L - A) = BN^{-.31}$, one can check that $0.8(L - A)$ is approximately equal to $B(2N)^{-.31}$.

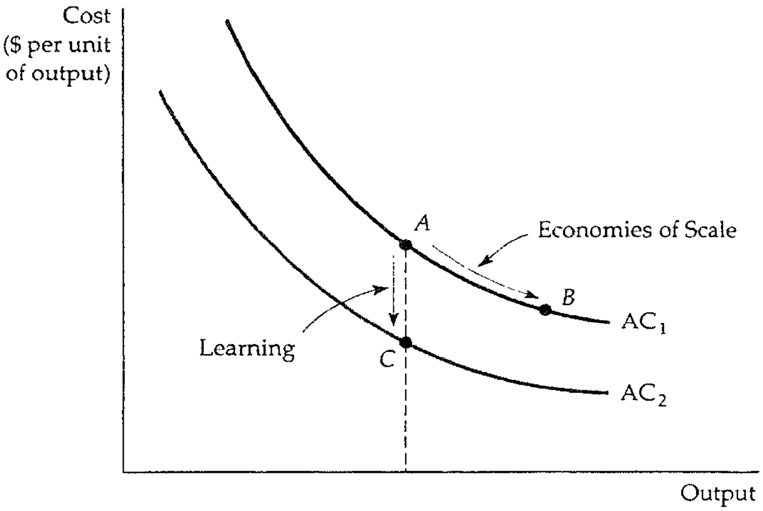

FIGURE 7.12 Economies of Scale Versus Learning. A firm's average cost of production can decline over time because of growth of sales when increasing returns are present (a move from *A* to *B* on curve AC1), or it can decline because there is a learning curve (a move from *A* on curve AC1 to *C* on curve AC2).

cost due to economies of scale. However, the move from *A* on AC1 to *C* on AC2 leads to lower cost due to learning, which shifts the average cost curve downward.

The learning curve is crucial for a firm that wants to predict the cost of producing a new product. Suppose, for example, that a firm producing machine tools knows that its labor requirement per machine for the first 10 machines is 1.0, the minimum labor requirement A is equal to zero, and β is approximately equal to 0.32. Table 7.3 calculates the total labor requirement for producing 80 machines.

Because there is a learning curve, the per-unit labor requirement falls with increased production. As a result, the total labor requirement for producing more and more output increases in smaller and smaller increments. Therefore,. a firm looking at the high initial labor requirement will obtain an overly pessimistic view of the business. Suppose the firm plans to be in business for a long time and the total labor requirement for each year's product is 10. In the first year of production, the labor requirement is 10, so the firm's cost will be high as it learns the business. But once the learning effect has taken place, production costs will be lower. After 8 years, the labor requirement will be only 0.51, and per-unit cost will be roughly half what it was in the first year of production. Thus, learning curve effects can be important for a firm deciding whether it is profitable to enter an industry.

TABLE 7.3 Predicting the Labor Requirements of Producing a Given Output

Cumulative Output (*N*)	Per-Unit Labor Requirement for each 10 units of Output (*L*)[14]	Total Labor Requirement
10	1.00	10.0
20	.80	18.0 (10.0 + 8.0)
30	.70	25.0 (18.0 + 7.0)
40	.64	31.4 (25.0 + 6.4)
50	.60	37.4 (31.4 + 6.0)
60	.56	43.0 (37.4 + 5.6)
70	.53	48.3 (43.0 + 5.3)
80 and over	.51	53.4 (48.3 + 5.1)

EXAMPLE 7.5 THE LEARNING CURVE IN THE CHEMICAL PROCESSING INDUSTRY

Suppose that as the manager of a firm that has just entered the chemical processing industry you face the following problem: Should you produce a relatively low level of output (and sell at a high price), or should you price your product lower and increase your rate of sales? The second alternative is particularly appealing if there is a learning curve in this industry. Then the increased volume will lower your average production costs over time and increase the firm's profitability.

To decide what to do, you can examine the available statistical evidence that distinguishes the components of the learning curve (learning new processes by labor, engineering improvements, etc.) from increasing returns to scale. A study of 37 chemical products from the late 1950s to 1972 reveals that cost reductions in the chemical processing industry were directly tied to the growth of cumulative industry output, to investment in improved capital equipment, and to a lesser extent to economies of scale.[15] In fact,for the entire sample of chemical products, average costs of production fell at 5.5 percent per year.[16] The study reveals that for each doubling of plant scale, the average cost of production falls by 11 percent. For each doubling of cumulative output, how-

[14] The numbers in this column were calculated from the equation log(L) = -0.322 log(*N/10*), where *L* is the unit labor input and *N* is cumulative output.

[15] The study was by Marvin Lieberman, "The Learning Curve and Pricing in the Chemical Processing Industries," *RAND Journal of Economics* 15 (1984): 213-228.

[16] The author used the average cost AC of the chemical products, the cumulative industry output X, and the average scale of a production plant Z and estimated the relationship log (AC) = - 0.387 log (X) - 0.173 log (Z). The -0.387 coefficient on cumulative output tells us that for every 1 percent increase in cumulative output, average cost decreases 0.387 percent. The -0.173 coefficient on plant size tells us that for every 1 percent increase in plant size,cost decreases 0.173 percent.

ever, the average cost of production falls by 27 percent. The evidence shows clearly that learning effects are more important than economies of scale in the chemical processing industry.[17]

Learning curve effects can be important in determining the shape of long-run cost curves and can thus help guide the firm's manager. The manager can use learning curve information to decide whether a production operation is profitable, and if it is, to plan how large the plant operation and the volume of cumulative output need be before a positive cash flow will result.

*7.7 Estimating and Predicting Cost

A business that is expanding or contracting its operation needs to predict how costs will change as output changes. Estimates of future costs can be obtained from a *cost function,* which relates the cost of production to the level of output and other variables that the firm can control.

Suppose we wanted to characterize the short-run cost of production in the automobile industry. We could obtain data on the number of automobiles Q produced by each car company and relate this information to the variable cost of production VC. The use of variable cost, rather than total cost, avoids the problem of trying to allocate the fixed cost of a multiproduct firm's production process to the particular product being studied.[18]

Figure 7.13 shows a typical pattern of cost and output data. Each point on the graph relates the output of an auto company to that company's variable cost of production. To predict cost accurately, we need to determine the underlying relationship between variable cost and output. Then, if a company expands its production, we can calculate what the associated cost is likely to be. The curve in the figure is drawn with this in mind-it provides a reasonably close fit to the cost data. (Typically, least-squares regression analysis would be used to fit the curve to the data.) But what shape of curve is the most appropriate, and how do we represent that shape algebraically?

One cost function that might be chosen is

$$\mathrm{VC} = \alpha + \beta Q \tag{7.10}$$

[17] By interpreting the two coefficients in footnote 16 in light of the levels of the output and plant size variables, one can allocate about 15 percent of the cost reduction to increases in the average scale of plants, and 85 percent to increases in cumulative industry output. (Suppose plant scale doubled, while cumulative output increased by a factor of 5 during the study. Then costs would fall by 11 percent from the increased scale and by 62 percent from the increase in cumulative output.)

[18] If an additional piece of equipment is needed as output increases, then annual rental cost of the equipment should be counted as a variable cost. If, however, the same machine can be used at all output levels, then its cost is fixed and should not be included.

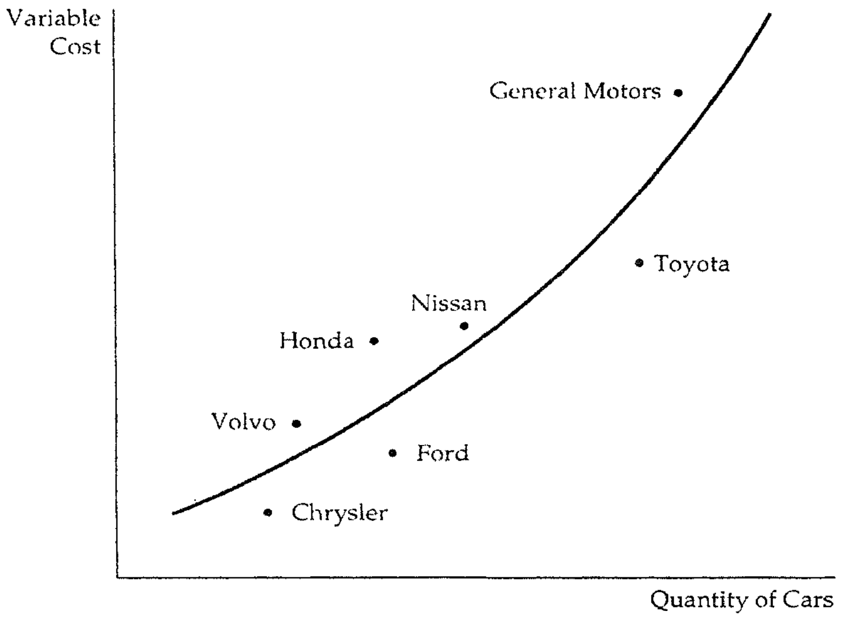

FIGURE 7.13 Total Cost Curve for the Automobile Industry. An empirical estimate of the total cost curve can be obtained by using data for individual firms in an industry. The total cost curve for automobile production is obtained by determining statistically the curve that best fits the points that relate the output of each firm to the total cost of production.

This *linear* relationship between cost and output is easy to use but is applicable only if marginal cost is constant.[19] For every unit increase in output, variable cost increases by β, so marginal cost is constant and equal to β. (α is also a component of variable cost but it varies with factors other than output.)

If we wish to allow for a U-shaped average cost curve and a marginal cost that is not constant, we must use a more complex cost function. One possibility, shown in Figure 7.14, is the *quadratic* cost function, which relates variable cost to output and output squared:

$$VC = \alpha + \beta Q + \gamma Q^2 \tag{7.11}$$

This implies a straight-line marginal cost curve of the form MC = $\beta + 2\gamma Q$.[20] Marginal cost increases with output if γ is positive, and decreases with output if γ is negative. Average cost, given by $AC = \alpha/Q + \beta + \gamma Q$, is U-shaped when γ ispositive.

[19] In statistical cost analyses, other variables might be added to the cost function to account for differences in input costs, production processes, product mix, etc., among firms.

[20] Short-run marginal cost is given by $\Delta TVC/\Delta Q = \beta + \gamma\Delta(Q^2)/\Delta Q$. But $\Delta(Q^2)/\Delta Q = 2Q$. Check this using calculus or by numerical example.) Therefore, $\beta + 2\gamma Q$.

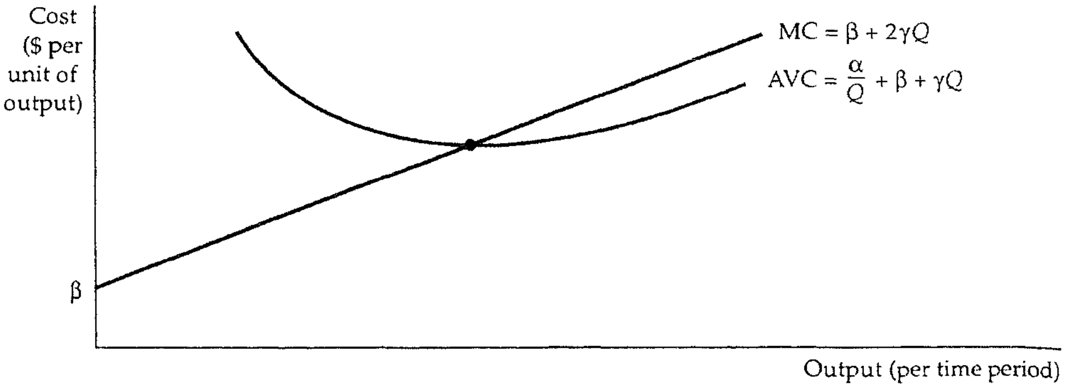

FIGURE 7.14 Quadratic Cost Function. A quadratic function is useful for either short-run or long-run cost functions when the average cost curve is U-shaped and the marginal cost curve is linear.

If the marginal cost curve is not linear, we might use a *cubic* cost function:

$$VC = \alpha + \beta Q + \gamma Q^2 + \delta Q^3 \tag{7.12}$$

Figure 7.15 shows this cubic cost function. It implies U-shaped marginal as well as average cost curves.

Cost functions can be difficult to measure. First, output data often represent an aggregate of different types of products. Total automobiles produced by General Motors, for example, involves different models of cars. Second, cost data are often obtained directly from accounting information that fails to re-

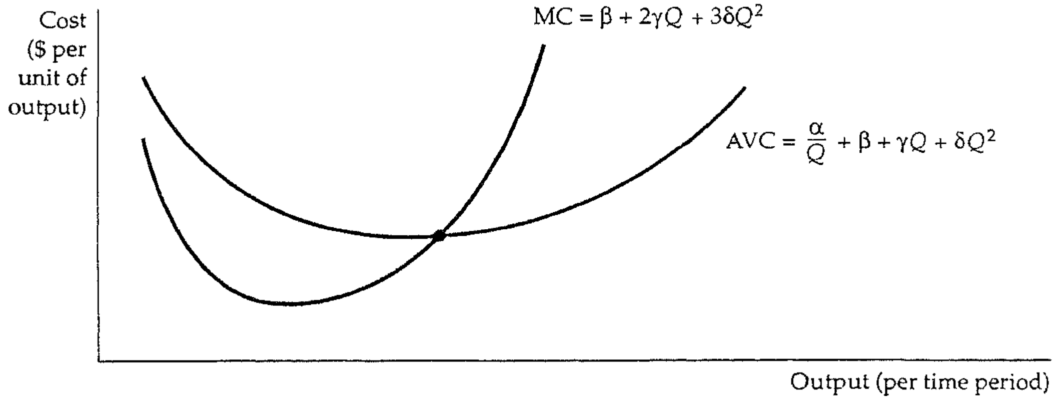

FIGURE 7.15 Cubic Cost Function. A cubic cost function implies that the average and the marginal cost curves are U-shaped.

fleet opportunity costs. Third, allocating maintenance and other plant costs to a particular product is difficult when the firm is a conglomerate that produces more than one product line.

Cost Functions and the Measurement of Scale Economies

Recall that the cost-output elasticity E_c is less than one when there are economies of scale and greater than one when there are diseconomies of scale. An alternative index, the *scale economies index* (SCI), is defined as follows:

$$\text{SCI} = 1 - E_c \tag{7.13}$$

When $E_c = 1$, SCI = 0, and there are no economies or diseconomies of scale. When £c is greater than one, SCI is negative, and there are diseconomies of scale. Finally, when E_c is less than 1, SCI is positive, and there are economies of scale.

EXAMPLE7.6COSTFUNCTIONFORELECTRICPOWER

In 1955, consumers bought 369 billion kilowatt-hours (kwh) of electricity; in 1970 they bought 1083 billion. Because there were fewer electric utilities in 1970, the output per firm had increased substantially. Was this increase due to economies of scale or other reasons? If it was the result of economies of scale, it would be economically inefficient for regulators to "break up" electric utility monopolies.

An interesting study of scale economies was based on the years 1955 and 1970 for investor-owned utilities with more than $1 million in revenues.[21] The cost of electric power was estimated by using a cost funcion that is somewhat more sophisticated than the quadratic and cubic functions discussed earlier.[22] Table 7.4 shows the resulting estimates of the Scale Economies Index (SCI). The results are based on a classification of all utilities into 5 size categories, with the median output (measured in kilowatt-hours) in each category listed.

The positive values of SCI tell us that all sizes of firms had some economies of scale in 1955. However, the magnitude of the economies of scale diminishes as firm size increases. The average cost curve associated with the 1955 study

TABLE 7.4 Scale Economies in the Electric Power Industry

Output (million kwh)	43	338	1109	2226	5819
Value of SCI, 1955	.41	.26	.16	.10	.04

[21] This example is based on Laurits Christensen and William H. Greene, Ecconomies of Scale in U.S. Electric Power Generation," Journal of Political Economy 84 (1976): 655-676.

[22] The translog cost function that was used provides a more general functional relationship than any of those we have discussed.

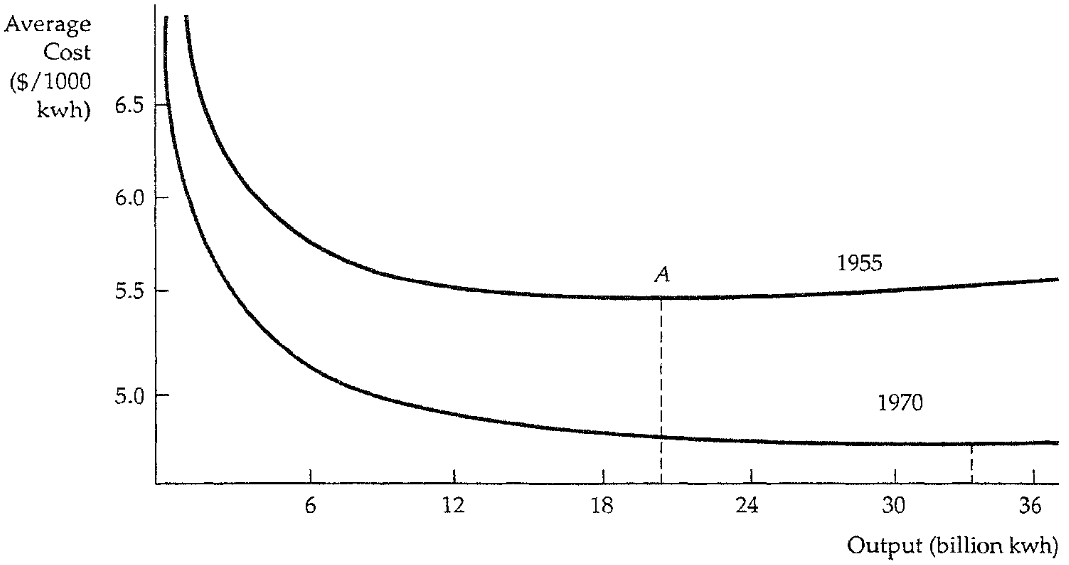

FIGURE 7.16 Average Cost of Production in the Electric Power Industry. The average cost of electric power in 1955 achieved a minimum at approximately 20 billion kilowatt-hours. By 1970 the average cost of production had fallen sharply and achieved a minimum at an output of greater than 32 billion kilowatt-hours.

is drawn in Figure 7.16 and labeled 1955. The point of minimum average cost occurs at point A at an output of approximately 20 billion kilowatts. Because there were no firms of this size in 1955, no firm had exhausted the opportunity for returns to scale in production. Note, however, that the average cost curve is relatively flat from an output of 9 billion kilowatts and higher, a range in which 7 of 124 firms produced.

When the same cost functions were estimated with 1970 data, the cost curve, labeled 1970 in Figure 7.16, was the result. The graph shows clearly that the average costs of production fell from 1955 to 1970. (The data are in real 1970 dollars.) But the flat part of the curve now begins at about 15 billion kwh. By 1970,24 of 80 firms were .producing in this range. Thus, many more firms were operating in the flat portion of the average cost curve in which economies of scale are not an important phenomenon. More important, most of the firms were producing in a portion of the 1970 cost curve that was flatter than their point of operation on the 1955 curve. (Five firms were at a point of diseconomies of scale: Consolidated Edison [SCI = -0.003], Detroit Edison [SCI = -0.004], Duke Power [SCI = -0.012], Commonwealth Edison [SCI = -0.014], and Southern [SCI = -0.028].) Thus, unexploited scale economies were much smaller in 1970 than in 1955.

This cost function analysis makes it clear that the decline in the cost of producing electric power cannot be explained by the ability of larger firms to take advantage of economies of scale. Rather, improvements in technology unrelated to the scale of the firms' operation and the decline in the real cost of energy inputs, such as coal and oil, are important reasons for the lower costs. The tendency toward lower average cost caused by a movement to the right along an average cost curve is minimal compared with the effect of technological improvement.

EXAMPLE 7.7 A COST FUNCTION FOR THE SAVINGS AND LOAN INDUSTRY

Understanding returns to scale in the savings and loan industry is important for regulators who must decide how savings and loans should be restructured in light of the failure of numerous institutions. In this regard, the empirical estimation of a long-run cost function can be useful.[23]

Data were collected for 86 savings and loan associations for 1975 and 1976 in a region that includes Idaho, Montana,Oregon, Utah, Washington, and Wyoming. Output is difficult to measure in this case because a savings and loan association provides a service to its customers, rather than a physical product. The output Q measure reported here (and used in other studies) is the total assets of each savings and loan association. In general, the larger the asset base of an association, the higher its profitability. Long-run average cost LAC is measured by average operating expense. Output and total operating costs are measured in hundreds of millions of dollars. Average operating costs are measured as a percentage of total assets.

A quadratic long-run average cost function was estimated for the year 1975, yielding the following relationship:

$$\text{LAC} = 2.38 - 0.6153Q + 0.0536Q^2$$

The estimated long-run average cost function is U-shaped and reaches its point of minimum average cost when the total assets of the savings and loan reach $574 million.[24] (At this point the average operating expenses of the savings and loan are 0.61 percent of its total assets.) Because almost all savings and loans in the region being studied had substantially less than $574 million in assets, the cost function analysis suggests that an expansion of savings and loans through either growth or mergers would be valuable.

How appropriate such a policy is cannot be fully evaluated here, however. To do so, we would need to take into account the possible social costs associated with the lessening of competition from growth or mergers, and we would

[23] This example builds on J. Holton Wilson, "A Note on Scale Economies in the Savings and Loan *Industry," Business Economics* (Jan.1981): 45-49.

[24] This can be seen by graphing the curve, or by differentiating the average cost function with respect to Q, setting it equal to 0, and solving for Q.

need to assure ourselves that this particular cost function analysis accurately estimated the point of minimum average cost.

Summary

1. Managers, investors, and economists must take into account the opportunity cost associated with the use of the firm's resources-the cost associated with the opportunities foregone when the firm uses its resources in its next best alternative.
2. In the short run, one or more of the inputs of the firm are fixed. Total cost can be divided into fixed Cost and variable cost. A firm's *marginal cost* is the additional variable cost associated with each additional unit of output. The *average variable cost* is the total variable cost divided by the number of units of output.
3. When there is a single variable input, as in the short run, the presence of diminishing returns determines the shape of the cost curves. In particular, there is an inverse relationship between the marginal product of the variable input and the marginal cost of production. The average variable cost and average total costcurves are U-shaped. The short-run marginal cost curve increases beyond a certain point, and cuts both average cost curves from below at their minimum points.
4. In the long run, all inputs to the production process are variable. As a result, the choice of inputs depends both on the relative costs of the factors of production and on the extent to which the firm can substitute among inputs in its production process. The cost-minimizing input choice is made by finding the point of tangency between the isoquant representing the level of desired output and an isocost line.
5. The firm's expansion path describes how its cost-minimizing input choices vary as the scale or output of its operation increases. As a result, the expansion path provides useful information relevant for long-run planning decisions.
6. The long-run average cost curve is the envelope of the firm's short-run average cost curves, and it reflects the presence or absence of returns to scale. When there are constant returns to scale and many plant sizes are possible, the long-run cost curve is horizontal, and the envelope consists of the points of minimum short-run average cost. However, when there are increasing returns to scale initially and then decreasing returns to scale, the long-run average cost curve is U-shaped, and the envelope does not include all points of minimum short-run average cost.
7. A firm enjoys economies of scale when it can double its output at less than twice the cost. Correspondingly, there are diseconomies of scale when a doubling of output requires more than twice the cost. Scale economies and diseconomies apply even when input proportions are variable; returns to scale applies only when input proportions are fixed.
8. When a firm produces two (or more) outputs, it is important to note whether there are economies of scope in production. Economies of scope arise when the firm can produce any combination of the two outputs more cheaply than couJd two independent firms that each produced a single product. The degree of economies of scope is measured by the per-

centage in reduction in cost when one firm produces two products relative to the cost of producing them individually.

9. A firm's average cost of production can fall over time if the firm 'Teams" how to produce more effectively. The *learning curve* describes how much the input needed to produce a given output falls as the cumulative output of the firm increases.

10. Cost functions relate the cost of production to the level of output of the firm. The functions can be measured in both the short run and the long run by using either data for firms in an industry at a given time or data for an industry over time. A number of functional relationships including linear, quadratic, and cubic can be used to represent cost functions.

Questions for Review

1. A firm pays its accountant an annual retainer of $10,000. Is this an explicit or an implicit cost?

2. The owner of a small retail store does her own accounting work. How would you measure the opportunity cost of her work?

3. Suppose a chair manufacturer finds that the marginal rate of technical substitution of capital for labor in his production process is substantially greater than the ratio of the rental rate on machinery to the wage rate for assembly-line labor. How should he alter his use of capital and labor to minimize the cost of production?

4. Why are isocost lines straight lines?

5. If the marginal cost of production is increasing, does this tell you whether the average variable cost is increasing or decreasing? Explain.

6. If the marginal cost of production is greater than the average variable cost, does this tell you whether the average variable cost is increasing or decreasing? Explain.

7. If the firm's average cost curves are U-shaped, why does its average variable cost curve achieve its minimum at a lower level of output than the average total cost curve?

8. If a firm enjoys increasing returns to scale up to a certain output level, and then constant returns to scale, what can you say about the shape of the firm's long-run average cost curve?

9. How does a change in the price of one input change the firm's long-run expansion path?

10. Distinguish between economies of scale and economies of scope. Why can one be present without the other?

Exercises

1. Assume a computer firm's marginal costs of production are constant at $1000 per computer. However, the fixed costs of production are equal to $10,000.

a. Calculate the firm's average variable cost and average total cost curves.

b. If the firm wanted to minimize the average total cost of production, would it choose to be very large or very small? Explain.

2. If a firm hires a currently unemployed worker, the opportunity cost of utilizing the worker's service is zero. Is this true? Discuss.

3. a. Suppose a firm must pay an annual franchise fee, which is a fixed sum, independent of whether it produces any output. How does this tax affect the firm's fixed, marginal, and average costs?

b. Now suppose the firm is charged a tax that is proportional to the number of items it produces. Again, how does this tax affect the firm's fixed, marginal, and average costs?

4. A chair manufacturer hires its assembly-line labor for $22 an hour and calculates that the rental cost of its machinery is $110 per hour. Suppose that a chair can be produced using 4 hours of labor or machinery in any combination. If the firm is currently using 3 hours of labor for each hour of machine time, is it minimizing its costs of production? If so, why? If not, how can it improve the situation?

5. Suppose the economy takes a downturn, and labor costs fall by 50 percent and are expected to stay at that level for a long time. Show graphically how this change in the relative price of labor and capital affects the firm's expansion path.

6. You are in charge of cost control in a large metropolitan transit district. A consultant you have hired comes to you with the following report:

> Our research has shown that the cost of running **a** bus for each trip down its line is $30 regardless of the number of passengers it carries. Each bus can carry 50 people. At rush hour, when the buses are full, the average cost per passenger is 60 cents. However, during off-peak hours, average ridership falls to 18 people, and average cost soars to $1.67 per passenger. As a result, we should encourage more rush-hour business when costs are cheaper and discourage off-peak business when costs are higher.

Do you follow the consultant's advice? Discuss.

7. An oil refinery consists of different pieces of processing equipment, each of which differs in its ability to break down heavy sulfurized crude oil into final products. The refinery process is such that the marginal cost of producing gasoline is constant up to a point as crude oil is put through a basic distilling unit. However, as the unit fills up, the firm finds that in the short run the amount of crude oil that can be processed is limited. The marginal cost of producing gasoline is also constant up to a capacity limit when crude oil is put through a more sophisticated hydrocracking unit. Graph the marginal cost of gasoline production when a basic distilling unit and a hydrocracker are used.

* **8.** A computer company's cost function, which relates its average cost of production AC to its cumulative output in thousands of computers *CQ* and its plant size in terms of thousands of computers produced per year *Q*, within the production range of 10/000 to 50,000 computers, is given by

$$\text{AC} = 10 - 0.1\text{CQ} + 0.3\text{Q}$$

a. Is there a learning curve effect?

b. Are there increasing or decreasing returns to scale?

c. During its existence, the firm has produced a total of 40,000 computers and is producing 10,000 computers this year. Next year it plans to increase its production to 12,000 computers. Will its average cost of production increase or decrease? Explain.

9. The total short-run cost function of a company is given by the equation C = 190 + 53Q, where C is the total cost and *Q* is the total quantity of output, both measured in tens of thousands.

a. What is the company's fixed cost?

b. If the company produced 100/000 units of goods, what is its average variable cost?

c. What is its marginal cost *per unit* produced?

d. What is its average fixed cost?

e. Suppose the company borrows money and expands its factory. Its fixed cost rises by $50,000, but its variable cost falls to $45,000 per 10,000 units. The cost of interest (*I*) also enters into the equation. Each one-point increase in the interest rate raises costs by $30,000. Write the new cost equation.

***10.** Suppose the long-run total cost function for an industry is given by the cubic equation $\text{TC} = a + bQ + cQ_2 + dQ_3$. Show (using calculus) that this total cost function is consistent with a U-shaped average cost curve for at least some values of the parameters a, *b,* c, *d.*

***11.** A computer company produces hardware and software using the same plant and labor. The total cost of producing computer processing units H and software programs S is given by

$$\text{TC} = aH + bS - cHS$$

where *a, b,* and *c* are positive. Is this total cost function consistent with the presence of economies or diseconomies of scale? With economies or diseconomies of scope?

APPENDIX TO CHAPTER 7

Production and Cost Theory-A Mathematical Treatment

This appendix presents a mathematical treatment of the basics of production and cost theory. As in the appendix to Chapter 4, we use the method of Lagrange multipliers to solve the firm's cost-miniinizing problem.

Cost Minimization

The theory of the firm relies on the assumption that firms choose inputs to the production process that minimize the cost of producing output. If there are two inputs, capital K and labor L, the production function $F(K, L)$ describes . the maximum output that can be produced for every possible combination of inputs. We assume that each of the factors in the production process has positive but decreasing marginal products. Writing the marginal product of capital as $MP_K(K, L) = \partial F(K, L)/\partial K$, we assume that $MP_K(K, L) > 0$ and $\partial MP_K(K, L)/\partial K < 0$. Similarly, if the marginal product of labor is given by $MP_l(K, L) = \partial F(K, L)/\partial L$, we assume that $MP_L(K, L) > 0$ and $\partial MP_L(K, L)/\partial L < 0$.

A competitive firm takes the prices of both labor w and capital ı rıs given. Then the cost-minimization problem can be written as

$$\text{Minimize } C = wL + rK \tag{A7.1}$$

subject to the constraint that a fixed output Q_0 be produced:

$$F(K,L)=Q_0 \tag{A7.2}$$

C represents the cost of producing the fixed level of output Q_0, and w and r are the wage rate and price of capital, respectively.

To determine the firm's demand for the inputs capital and labor, we choose the values of K and L that minimize (A7.1) subject to (A7.2). We solve this constrained optimization problem using the method discussed in Appendix 4. The Lagrangian is

$$\Phi = wL + rK - \lambda[F(K, L) - Q_0] \tag{A7.3}$$

where X is the Lagrange multiplier. By differentiating with respect to K, L, and then X, and equating the derivatives to zero, we obtain the necessary conditions for a maximum:[1]

$$\begin{aligned} \partial\Phi/\partial K &= r - \lambda \mathrm{MP}_K(K, L) = 0 \\ \partial\Phi/\partial L &= w - \lambda \mathrm{MP}_L(K, L) = 0 \\ \partial\Phi/\partial\lambda &= F(K, L) = Q_0 \end{aligned} \tag{A7.4}$$

By combining the first two conditions we obtain

$$\mathrm{MP}_K(K, L)/r = \mathrm{MP}_L(K, L)/w \tag{A7.5}$$

Equation (A7.5) tells us that if the firm is minimizing costs it will choose its factor inputs to equate the ratio of the marginal product of each factor divided by its price. To see that this makes sense, suppose MP_K/r were greater than MP_L/w. Then the firm could reduce its cost while still producing the same output by using more capital and less labor.

Finally, we can combine the first two conditions of (A7.4) in a different way to evaluate the Lagrange multiplier:

$$\lambda = r/\mathrm{MP}_K(K, L) = w/\mathrm{MP}_L(K, L) \tag{A7.6}$$

Suppose output increases by one unit. Because the marginal product of capital measures the extra output associated with an additional input of capital, $1/\mathrm{MP}_K(K, L)$ measures the extra capital needed to produce one unit of output. Therefore, $r/\mathrm{MP}_K(K, L)$ measures the additional input cost of producing an additional unit of output by increasing capital. Likewise, $w/\mathrm{MP}_L(K, L)$ measures the additional cost of producing a unit of output using additional labor as an input. In both cases, the Lagrange multiplier is equal to the marginal cost of production, because it tells us how much the cost increases if the amount is increased by one unit.

Marginal Rate of Technical Substitution

Recall that an isoquant is a curve that represents the set of all input combinations that give the firm the same level of output, say, Q^*. Thus, the condition that $F(K, L) = Q^*$ represents a production isoquant. As input combinations are changed along an isoquant, the change in output, given by the total derivative of $F(K, L)$, equals zero (i.e., $dQ = 0$). Thus,

$$\mathrm{MP}_K(K, L)dK + \mathrm{MP}_L(K, L)d\mathrm{L} = dQ = 0 \tag{A7.7}$$

It follows by rearrangement that

$$-dK/dL = \mathrm{MRTS}_{LK} = \mathrm{MP}_L(K, L)/\mathrm{MP}_K(K, L) \tag{A7.8}$$

[1] These conditions are necessary for a solution involving positive amounts of both inputs.

where MRTSLK is the firm's marginal rate of technical substitution between labor and capital.

Now, rewrite the condition given by (A7.5) to get

$$MP_L(K, L)/MP_K(K, L) = w/r \tag{A7.9}$$

Because the left-hand side of (A7.8) represents the negative of the slope of the isoquant, it follows that at the point of tangency of the isoquant and the isocost line, the firm's marginal rate of technical substitution (which trades off inputs while keeping output constant) is equal to the ratio of the input prices (which represents the slope of the firm's isocost line).

We can look at this result another way by rewriting (A7.9) again:

$$MP_L/w = MP_K/r \tag{A7.10}$$

Equation (A7.10) tells us that the marginal products of all production inputs must be equal when these marginal products are adjusted by the unit cost of each input. If *the* cost-adjusted marginal products were not equal, the firm could change its inputs to produce the same output at a lower cost.

Duality in Production and Cost Theory

As in consumer theory, the firm's input decision has a dual nature. The optimum choice of K and L can be analyzed not only as the problem of choosing the lowest isocost line tangent to the production isoquant, but also as the problem of choosing the highest production isoquant tangent to a given isocost line. To see this, consider the following dual producer problem:

$$\text{Maximize}\, F(K, L)$$

subject to the cost constraint that

$$wL + rK = C_0 \tag{A7.11}$$

The corresponding Lagrangian is given by

$$\Phi = F(K, L) - \mu(wL + rK - C_0) \tag{A7.12}$$

where μ is the Lagrange multiplier. The necessary conditions for output maximization are:

$$\begin{aligned} MP_K(K, L) - \mu r &= 0 \\ MP_L(K, L) - \mu w &= 0 \\ wL + rK - C_0 &= 0 \end{aligned} \tag{A7.13}$$

By solving the first two equations, we see that

$$MP_K(K, L)/r = MP_L(K, L)/w \tag{A7.14}$$

which is identical to the condition that was necessary for cost minimization.

The Cobb-Douglas Cost and Production Functions

Given a specific production function $F(K,L)$, conditions (A7.13) and (A7.14) can be used to derive the *cost function* $C(Q)$. To see this, let's work through the example of a Cobb-Douglas production function. This production function is

$$F(K, L) = AK^{\alpha}L^{\beta}$$

or, by taking the logs of both sides of the production function equation:

$$\log [F(K, L)] = \log A + \alpha \log K + \beta \log L$$

We assume that a < 1 and (B < 1, so that the firm has decreasing marginal products of labor and capital.[2] If a + (3 = 1, the firm has *constant returns to scale,* because doubling *K* and L doubles F. If a -h (3 > 1, the firm has *increasing returns to scale,* and if *a* + (3 < 1, it has *decreasing returns to scale.*

To find the amounts of capital and labor that the firm should utilize to minimize the cost of producing an output Q_0, we first write the Lagrangian:

$$\Phi = wL + rK - \lambda(AK^{\alpha}L^{\beta} - Q_0) \qquad \text{(A7.15)}$$

Differentiating with respect to L, K, and λ, and setting those derivatives equal to 0, we obtain

$$\partial\Phi/\partial L = w - \lambda(\beta AK^{\alpha}L^{\beta-1}) = 0 \qquad \text{(A7.16)}$$

$$\partial\Phi/\partial K = r - \lambda(\alpha AK^{\alpha-1}L^{\beta}) = 0 \qquad \text{(A7.17)}$$

$$\partial\Phi/\partial\lambda = AK^{\alpha}L^{\beta} - Q_0 = 0 \qquad \text{(A7.18)}$$

From equation (A7.16) we have

$$\lambda = w/A\beta K^{\alpha}L^{\beta-1} \qquad \text{(A7.19)}$$

Substituting this into equation (A7.17) gives us

$$r\beta AK^{\alpha}L^{\beta-1} = w\alpha AK^{\alpha-1}L^{\beta} \qquad \text{(A7.20)}$$

or

$$L = \beta rK/\alpha w \qquad \text{(A7.21)}$$

Now, use equation (A7.21) to eliminate L from equation (A7.18):

$$AK^{\alpha}\beta^{\beta}r^{\beta}K^{\beta}/\alpha^{\beta}w^{\beta} = Q_0 \qquad \text{(A7.22)}$$

Rewrite this as

$$K^{\alpha+\beta} = (\alpha w/\beta r)^{\beta}Q_0/A \qquad \text{(A7.23)}$$

or

$$K = [(\alpha w/\beta r)^{\beta/(\alpha+\beta)}](Q_0/A)^{1/(\alpha+\beta)} \qquad \text{(A7.24)}$$

[2] For example, the marginal product of labor is given by $MP_L = \partial[F(K,L)]/\partial L = \beta AK^{\alpha}L^{\beta-1}$, so that MPL falls as L increases.

We have now determined the cost-minimizing quantity of capital. To determine the cost-minimizing quantity of labor, just substitute equation (A7.24) into equation (A7.21):

$$L = [(\beta r/\alpha w)^{\alpha/(\alpha+\beta)}](Q_0/A)^{1/(\alpha+\beta)} \tag{A7.25}$$

Note that if the wage rate $\hat{w}$ rises relative to the price of capital r, the firm will use more capital and less labor. If, say because of technological change, A increases (so the firm can produce more output with the same inputs), both K and L will fall.

We have shown how cost-minimization subject to an output constraint can be used to determine the firm's optimal mix of capital and labor. Now we will determine the firm's cost function. The total cost of producing *any output Q* can be obtained by substituting equations (A7.24) for K and (A7.25) for L into the equation $C = wL + rK$. After some algebraic manipulation we find that

$$C = w^{\beta/(\alpha+\beta)}\, r^{\alpha/(\alpha+\beta)}\left[\left(\frac{\alpha}{\beta}\right)^{\beta/(\alpha+\beta)} + \left(\frac{\alpha}{\beta}\right)^{-\alpha/(\alpha+\beta)}\right]\left(\frac{Q}{A}\right)^{1/(\alpha+\beta)} \tag{A7.26}$$

This *cost function* tells both how the total cost of production increases as the level of output Q increases, and also how cost changes as input prices change. When $\alpha + \beta$ equals 1, cost will increase proportionately with output, which means that the production process exhibits constant returns to scale. Likewise if $\alpha + \beta$ is greater than 1, there are decreasing returns to scale, and if $a\alpha + \beta$ is less than 1, there are increasing returns to scale.

Now consider the dual problem of maximizing the output that can be produced with the expenditure of Co dollars. We leave it to you to work through this problem for the Cobb-Douglas production function, and show that equations (A7.24) and (A7.25) describe the cost-minimizing input choices. To get you started, note that the Lagrangian for this dual problem is $\Phi = AK^{\alpha}L^{\beta} - \mu(wL + rK - C_0)$.

Exercises

1. Of the following production functions, which exhibit increasing, constant, or decreasing returns to scale?

a.$F(K,L)=K_2L$

b. $F(K, L) = 10K+5L$

c. $F(K, L) = (KL)_{.5}$

2. The production function for a product is given by $Q = 100KL$. If the price of capital is $120 per day and the price of labor is $30 per day, what is the minimum cost of producing 1000 units of output?

3. Suppose a production function is given by $F(K, L) = KL_2$, and that the price of capital is $10 and the price of labor $15. What combination of labor and capital minimizes the cost of producing any given output?