

CHAPTER

COST

CURVES

8.1

LONG-RUN COST CURVES

APPLICATION 8.1 The Long Run Cost of Trucking

APPLICATION 8.2 The Costs of Higher Education

APPLICATION 8.3 Economies of Scale in Refining Alumina?

APPLICATION 8.4 Hospitals Are Businesses Too

8.2

SHORT-RUN COST CURVES

APPLICATION 8.5 Tracking Railroad Costs

8.3

SPECIAL TOPICS IN COST

APPLICATION 8.6 Economies of Scope for the Swoosh

APPLICATION 8.7 Experience Reduces Costs of Computer Chips

8.4

ESTIMATING COST FUNCTIONS

Appendix

SHEPHARD'S LEMMA AND DUALITY

How Can HiSense Get a Handle on Costs?

The Chinese economy in the 1990s underwent an unprecedented boom. As part of that boom, enterprises such as HiSense Group grew rapidly.¹ HiSense, one of China's largest television producers, increased its rate of production by 50 percent per year during the mid-1990s. Its goal was to transform itself from a sleepy domestic producer of television sets into a consumer electronics giant whose brand name was recognized throughout Asia. By 2004 HiSense was not only one of China's major producers of color TVs, but also one of its leading producers of personal computers.

¹This example is based on "Latest Merger Boom Is Happening in China and Bears Watching," *The Wall Street Journal* (July 30, 1997), pp. A1 and A9.

Of vital concern to HiSense and the thousands of other Chinese enterprises that were plotting similar growth strategies in the late 1990s and early 2000s was how production costs would change as volume of output increased. There is little doubt that HiSense's total production costs would go up as it produced more television sets. But *how fast* would they go up? HiSense's executives hoped that as it produced more television sets, the cost of *each television set* would go down, that is, its unit costs would fall as its annual rate of output went up.

HiSense's executives also needed to know how input prices would affect its production costs. For example, HiSense competes with other large Chinese television manufacturers to buy up smaller factories. This competition bids up the price of capital. HiSense had to reckon with the impact of this price increase on its total production costs.

This chapter picks up where Chapter 7 left off: with the comparative statics of the cost-minimization problem. The cost-minimization-problem—both in the long run and the short run—gives rise to total, average, and marginal cost curves. This chapter studies these curves.

CHAPTER PREVIEW In this chapter, you will

- Study cost curves, which show the relationships between costs and the volume of output. Cost curves include both long-run and short-run curves.
- Study long-run average and marginal cost curves and the relationships between them.
- Learn about economies and diseconomies of scale—situations in which average cost decreases or increases, respectively, as output goes up—including the concept of minimum efficient scale.
- Analyze the short-run total cost curve, which shows the minimized total cost of producing a given level of output when the quantity of at least one input is fixed.
- Learn about economies of scope (efficiencies that arise when a firm produces more than one product) and economies of experience (cost advantages that arise from accumulated experience).
- Learn how economists estimate cost functions, including the constant elasticity cost function and the translog cost function.



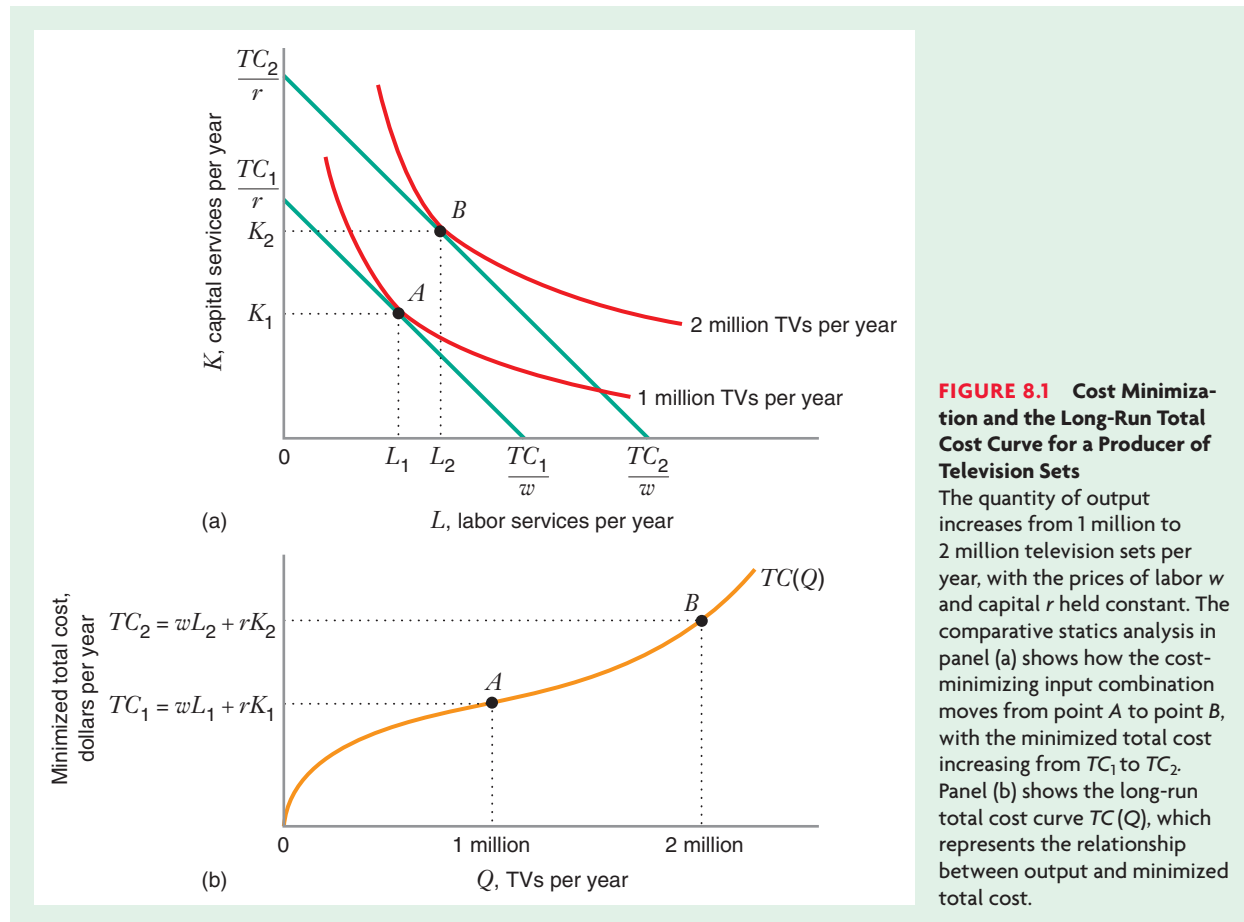
LONG-RUN TOTAL COST CURVE

In Chapter 7, we studied the firm's long-run cost-minimization problem and saw how the cost-minimizing combination of labor and capital depended on the quantity of output Q and the prices of labor and capital, w and r . Figure 8.1(a) shows how the optimal input combination for a television manufacturer changes as we vary output, holding input prices fixed. For example, when the firm produces 1 million televisions per year, the cost-minimizing input combination occurs at point A , with L_1 units of labor and K_1 units of capital. At this input combination, the firm is on an isocost line corresponding to TC_1 dollars of total cost, where $TC_1 = wL_1 + rK_1$. TC_1 is thus the minimized total cost when the firm produces 1 million units of output. When the firm increases output from 1 million to 2 million televisions per year, its isocost line shifts to the northeast, and its cost-minimizing input combination moves to point B , with L_2 units of labor and K_2 units of capital. Thus, its minimized total cost goes up (i.e., $TC_2 > TC_1$). It cannot be otherwise, because if the firm could decrease total cost by producing more output, it couldn't have been using a cost-minimizing combination of inputs in the first place.

Figure 8.1(b) shows the **long-run total cost curve**, denoted by $TC(Q)$. The long-run total cost curve shows how minimized total cost varies with output, holding input prices fixed, and selecting inputs to minimize cost. Because the cost-minimizing input

8.1 LONG-RUN COST CURVES

long-run total cost curve A curve that shows how total cost varies with output, holding input prices fixed, and choosing all inputs to minimize cost.



combination moves us to higher isocost lines, the long-run total cost curve must be increasing in Q . We also know that when $Q = 0$, long-run total cost is 0. This is because, in the long run, the firm is free to vary all its inputs, and if it produces a zero quantity, the cost-minimizing input combination is zero labor and zero capital. Thus, comparative statics analysis of the cost-minimization problem implies that the *long-run total cost curve must be increasing in Q and must equal 0 when $Q = 0$.*

LEARNING-BY-DOING EXERCISE 8.1



Finding the Long-Run Total Cost Curve from a Production Function

Let's return again to the production function $Q = 50\sqrt{LK}$ that we introduced in Learning-By-Doing Exercise 7.2.

Problem

- How does minimized total cost depend on the output Q and the input prices w and r for this production function?
- What is the graph of the long-run total cost curve when $w = 25$ and $r = 100$?

Solution

(a) In Learning-By-Doing Exercise 7.4 we saw that the following equations describe the cost-minimizing quantities of labor and capital: $L = (Q/50)\sqrt{r/w}$ and $K = (Q/50)\sqrt{w/r}$. To find the minimized total cost, we calculate the total cost the firm incurs when it uses this cost-minimizing input combination:

$$TC(Q) = wL + rK = w \frac{Q}{50} \sqrt{\frac{r}{w}} + r \frac{Q}{50} \sqrt{\frac{w}{r}} = \frac{Q}{50} \sqrt{wr} + \frac{Q}{50} \sqrt{wr} = \frac{\sqrt{wr}}{25} Q$$

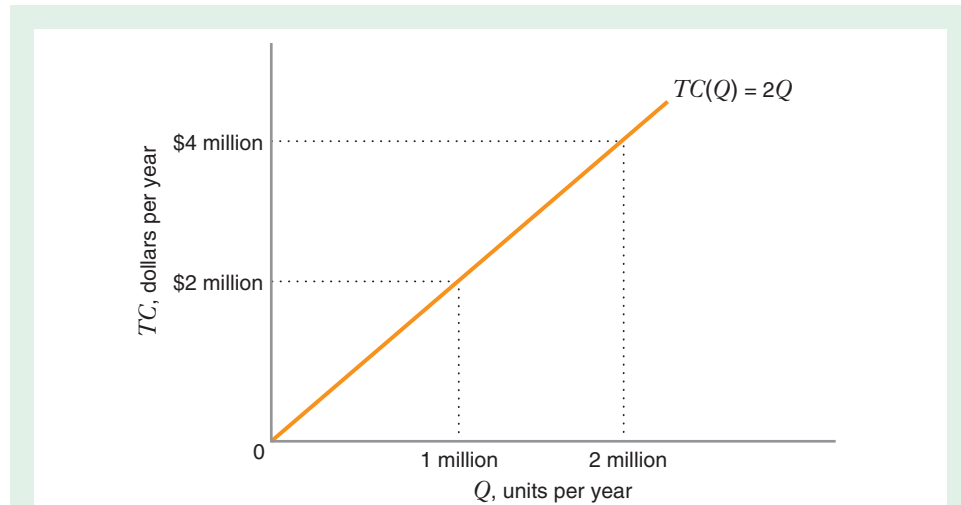


FIGURE 8.2 Long-Run Total Cost Curve

The graph of the long-run total cost curve $TC(Q) = 2Q$ is a straight line.

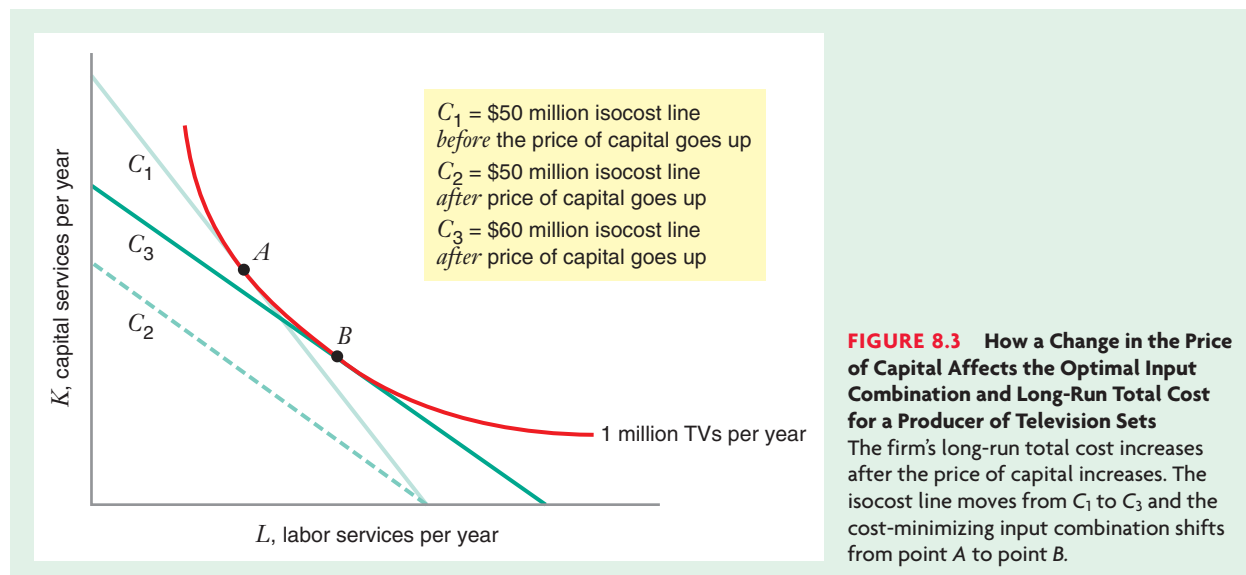
(b) If we substitute $w = 25$ and $r = 100$ into this equation for the total cost curve, we get $TC(Q) = 2Q$. Figure 8.2 shows that the graph of this long-run total cost curve is a straight line.

Similar Problems: 8.3, 8.7, and 8.10

HOW DOES THE LONG-RUN TOTAL COST CURVE SHIFT WHEN INPUT PRICES CHANGE?

What Happens When Just One Input Price Changes?

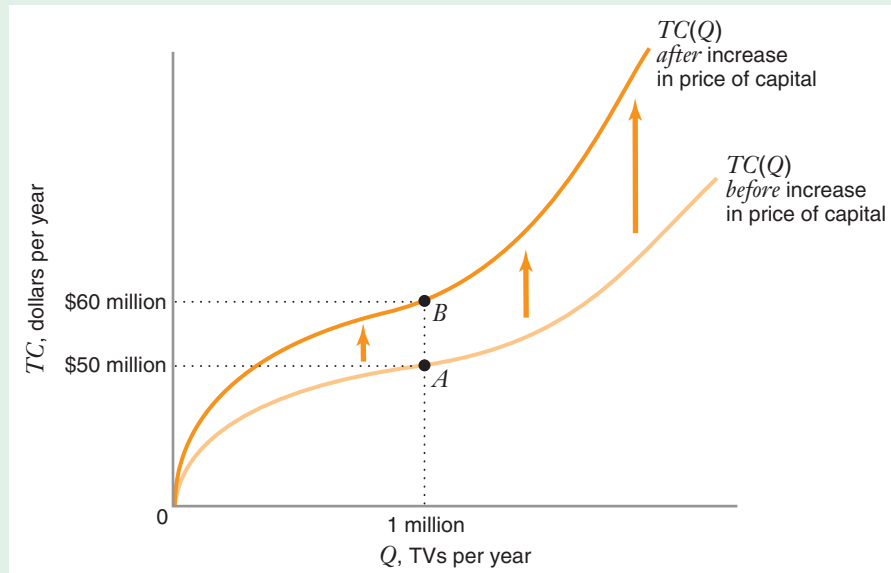
In the chapter introduction, we discussed how HiSense faced the prospect of higher prices for certain inputs, such as capital. To illustrate how an increase in an input price affects a firm's total cost curve, let's return to the cost-minimization problem for our hypothetical television producer. Figure 8.3 shows what happens when the price of capital increases, holding output and the price of labor constant. Suppose that at the initial situation, the optimal input combination for an annual output of 1 million television sets occurs at point A on isocost line C_1 , where the minimized total cost is \$50 million per year. After the increase in the price of capital, the optimal input combination is at point B on isocost line C_3 , corresponding to a total cost that is *greater* than \$50 million. To see why, note that the \$50 million isocost line *at the new input prices* (C_2) intersects the horizontal axis in the same place as the \$50 million isocost line *at the old input prices*. However, C_2 is flatter than C_1 because the price of capital has gone up. Thus, the firm could not operate on isocost line C_2 because it would be unable to produce the desired quantity of 1 million television sets. Instead, the firm must operate on an isocost line that is farther to the northeast (C_3) and thus corresponds to a higher level of cost (\$60 million perhaps). Thus, holding output fixed, the minimized total cost goes up when the price of an input goes up.²



²An analogous argument would show that minimized total cost goes down when the price of capital goes down.

FIGURE 8.4 How a Change in the Price of Capital Affects the Long-Run Total Cost Curve for a Producer of Television Sets

An increase in the price of capital causes the long-run total cost curve $TC(Q)$ to rotate upward. Points A and B correspond to the cost-minimizing input combinations in Figure 8.3.



This analysis then implies that an increase in the price of capital results in a new total cost curve that lies above the original total cost curve at every $Q > 0$ (at $Q = 0$, long-run total cost is still zero). Thus, as Figure 8.4 shows, an increase in an input price rotates the long-run total cost curve upward.³

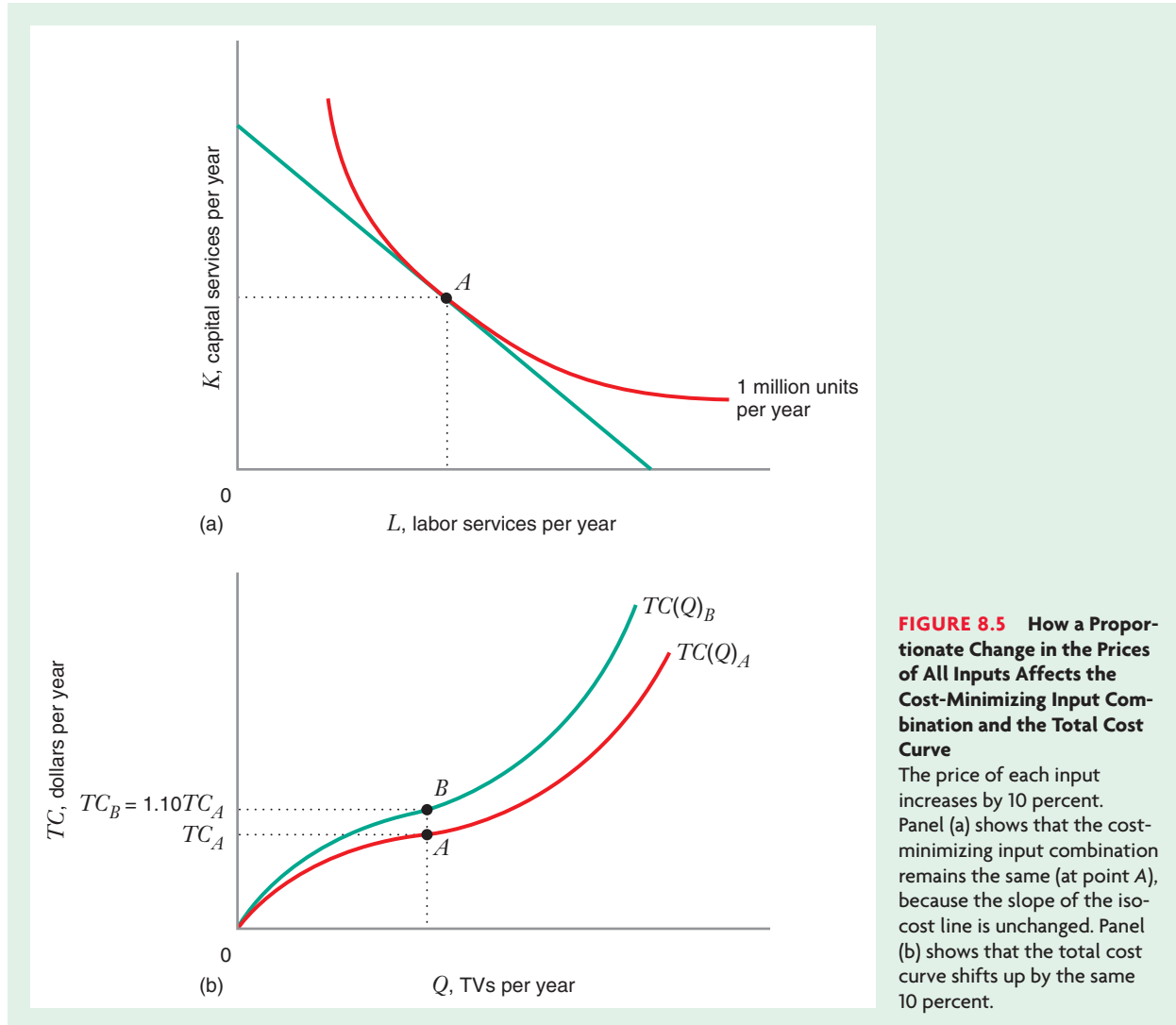
What Happens When All Input Prices Change Proportionately?

What if the price of capital and the price of labor both go up by the same percentage amount, say, 10 percent? The answer is that *a given percentage increase in both input prices leaves the cost-minimizing input combination unchanged, while the total cost curve shifts up by exactly the same percentage.*

As shown in Figure 8.5(a), at the initial prices of labor w and capital r , the cost-minimizing input combination is at point A. After both input prices increase by 10 percent, to $1.10w$ and $1.10r$, the ideal combination is still at point A. The reason is that the slope of the isocost line is unchanged by the price increase ($-w/r = -1.10w/1.10r$), so the point of tangency between the isocost line and the isoquant is also unchanged.

Figure 8.5(b) shows that the 10 percent increase in input prices shifts the total cost curve up by 10 percent. Before the price increase, total cost $TC_A = wL + rK$; after the price increase, total cost $TC_B = 1.10wL + 1.10rK$. Thus, $TC_B = 1.10TC_A$ (i.e., the total cost increases by 10 percent for any combination of L and K).

³There is one case in which an increase in an input price would not affect the long-run total cost curve. If the firm is initially at a corner point solution using a zero quantity of the input, an increase in the price of the input will leave the firm's cost-minimizing input combination—and thus its minimized total cost—unchanged. In this case, the increase in the input price may not shift the long-run total cost curve.



APPLICATION 8.1

The Long Run Cost of Trucking⁴

The intercity trucking business is a good setting in which to study the behavior of long-run total costs because when input prices or output changes, trucking firms can adjust their input mixes without too much difficulty. Drivers can be hired or laid off relatively easily, and trucks can be bought or sold as

circumstances dictate. There are also considerable data on output, expenditures on inputs, and input quantities, so we can use statistical techniques to estimate how total cost varies with input prices and output. Utilizing such data, Ann Friedlaender and Richard Spady estimated long-run total cost curves for trucking firms that carry general merchandise.

⁴This example draws from A. F. Friedlaender and R. H. Spady, *Freight Transport Regulation: Equity, Efficiency, and Competition in the Rail and Trucking Industries* (Cambridge, MA: MIT Press, 1981).

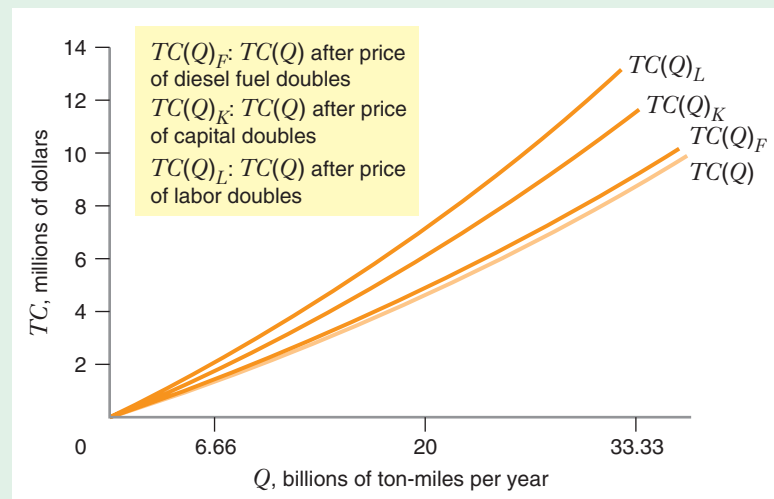
Trucking firms use three major inputs: labor, capital (e.g., trucks), and diesel fuel. Their output is transportation services, usually measured as ton-miles per year. One ton-mile is one ton of freight carried one mile. A trucking company that hauls 50,000 tons of freight 100,000 miles during a given year would thus have a total output of $50,000 \times 100,000$, or 5,000,000,000 ton-miles per year.

Figure 8.6 illustrates an example of the cost curve estimated by Friedlaender and Spady. Note that total cost increases with the quantity of output, as the

theory we just discussed implies. Total cost also increases with the price of each input (holding the prices of the other two inputs constant). Thus, doubling the price of labor causes the total cost curve $TC(Q)$ to shift upward to $TC(Q)_L$; doubling the cost of capital also shifts the total cost curve up [to $TC(Q)_K$], but not by as much. The smallest shift up [to $TC(Q)_F$] occurs when the price of fuel doubles. Friedlaender and Spady's analysis shows that the total cost of a trucking firm is most sensitive to changes in the price of labor and least sensitive to changes in the price of diesel fuel.

FIGURE 8.6 How Changes in Input Prices Affect the Long-Run Total Cost Curve for a Trucking Firm

Total cost is more sensitive to the price of labor than to the price of capital (trucks) or diesel fuel. Holding the prices of other inputs constant, doubling the price of labor shifts the cost curve up to $TC(Q)_L$; doubling the price of capital shifts it less, up to $TC(Q)_K$; and doubling the price of fuel shifts it least, up to $TC(Q)_F$.



LONG-RUN AVERAGE AND MARGINAL COST CURVES

What Are Long-Run Average and Marginal Costs?

Two other types of cost play an important role in microeconomics: long-run average cost and long-run marginal cost. **Long-run average cost** is the firm's cost per unit of output. It equals long-run total cost divided by Q : $AC(Q) = [TC(Q)]/Q$.

Long-run marginal cost is the rate at which long-run total cost changes with respect to a change in output: $MC(Q) = (\Delta TC)/(\Delta Q)$. Thus, $MC(Q)$ equals the slope of $TC(Q)$.

Although long-run average and marginal cost are both derived from the firm's long-run total cost curve, the two costs are generally different, as illustrated in Figure 8.7. At any particular output level, the long-run average cost is equal to the slope of a ray from the origin to the point on the long-run total cost curve corresponding to that output, whereas the long-run marginal cost is equal to the slope of the long-run total cost curve itself at that point. Thus, at point A on the total cost curve $TC(Q)$ in Figure 8.7(a), where the firm's output level is 50 units per year, the average cost is equal to the slope of ray OA , or $\$1,500/50$ units = $\$30$ per unit. By contrast, the marginal cost at point A is the slope of the line BAC (the line tangent to the total cost curve at A); the slope of this line is 10, so the marginal cost when output is 50 units per year is $\$10$ per unit.

long-run average cost The firm's total cost per unit of output. It equals long-run total cost divided by total quantity.

long-run marginal cost The rate at which long-run total cost changes with respect to change in output.

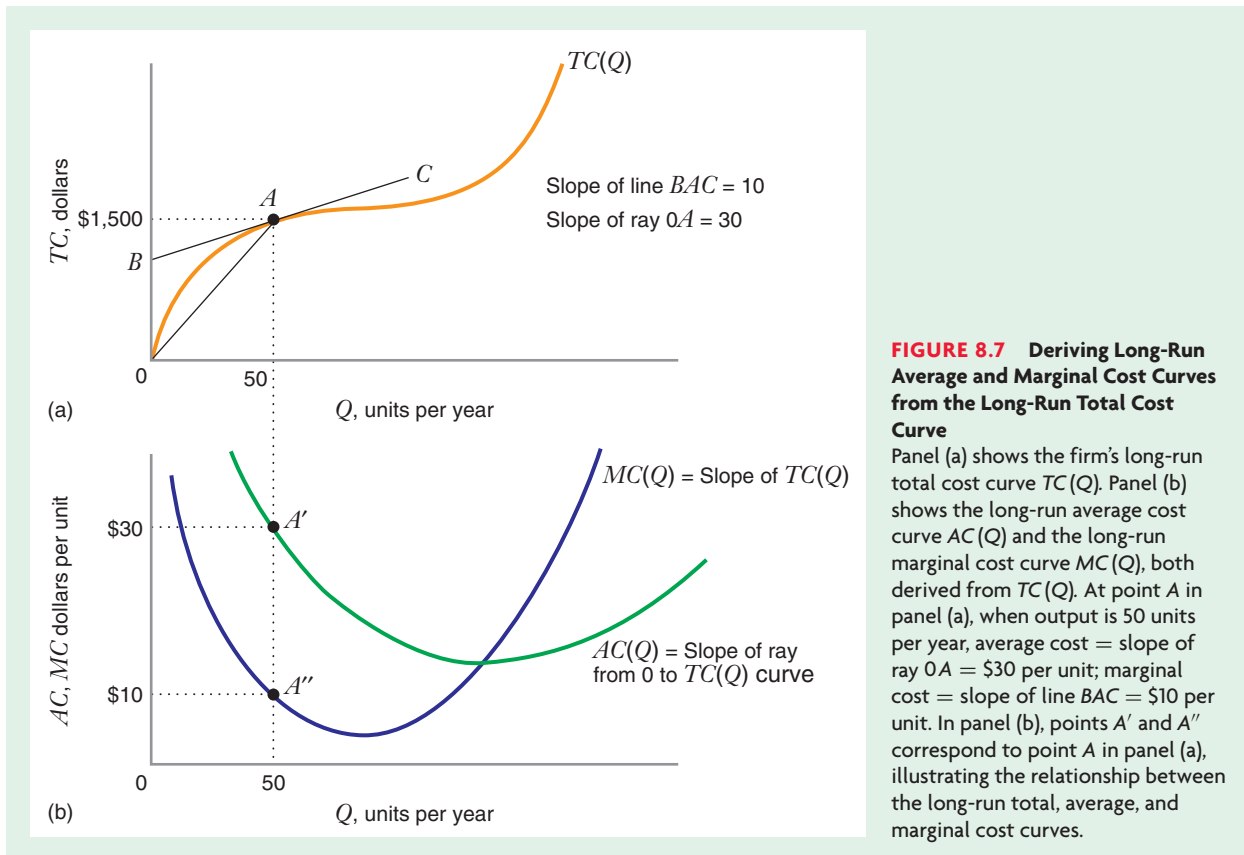


Figure 8.7(b) shows the long-run average cost curve $AC(Q)$ and the long-run marginal cost curve $MC(Q)$ corresponding to the long-run total cost curve $TC(Q)$ in Figure 8.7(a). The average cost curve shows how the slope of rays such as OA changes as we move along $TC(Q)$, whereas the marginal cost curve shows how the slope of tangent lines such as BAC changes as we move along $TC(Q)$. Thus, in Figure 8.7(b), when the firm's output equals 50 units per year, the average cost is \$30 per unit (point A') and the marginal cost is \$10 per unit (point A''), corresponding to the slope of ray OA and line BAC , respectively, at point A in Figure 8.7(a).

LEARNING-BY-DOING EXERCISE 8.2



Deriving Long-Run Average and Marginal Cost Curves from a Long-Run Total Cost Curve

In Learning-By-Doing Exercise 8.1 we derived the equation for the long-run total cost curve for the production function $Q = 50\sqrt{LK}$ when the price of labor L is $w = 25$ and the price of capital K is $r = 100$: $TC(Q) = 2Q$.

Problem What are the long-run average and marginal cost curves associated with this long-run total cost curve?

Solution Long-run average cost is $AC(Q) = [TC(Q)]/Q = 2Q/Q = 2$. Note that average cost does not depend on Q . Its graph would be a horizontal line, as Figure 8.8 shows.

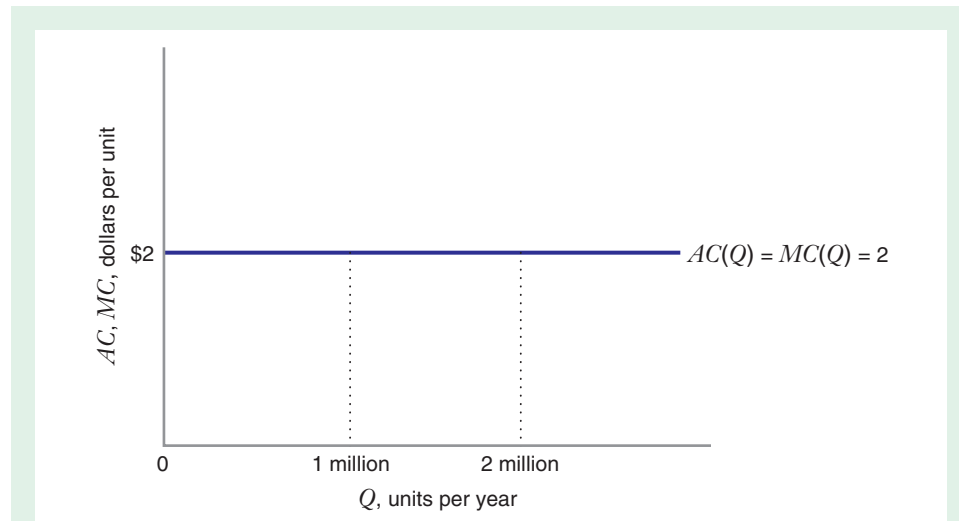


FIGURE 8.8 Long-Run Average and Marginal Cost Curves for the Production Function

$$Q = 50 \sqrt{LK}$$

The long-run average and marginal cost curves are identical horizontal lines at \$2 per unit when $w = 25$ and $r = 100$.

Long-run marginal cost is the slope of the long-run total cost curve. With $TC(Q) = 2Q$, the slope of the long-run total cost curve is 2, and thus $MC(Q) = 2$. Long-run marginal cost also does not depend on Q . Its graph is the same horizontal line.

This exercise illustrates a general point. Whenever the long-run total cost is a straight line (as in Figure 8.2), long-run average and long-run marginal cost curves will be the same and will be a horizontal line.

Similar Problem: 8.4

Relationship between Long-Run Average and Marginal Cost Curves

As with other average and marginal concepts (e.g., average product versus marginal product, discussed in Chapter 6), there is a systematic relationship between the long-run average and long-run marginal cost curves:

- If average cost is decreasing as quantity is increasing, then average cost is greater than marginal cost: $AC(Q) > MC(Q)$.
- If average cost is increasing as quantity is increasing, then average cost is less than marginal cost: $AC(Q) < MC(Q)$.
- If average cost is neither increasing nor decreasing as quantity is increasing, then average cost is equal to marginal cost: $AC(Q) = MC(Q)$.

Figure 8.9 illustrates this relationship.

As we discussed in Chapter 6, the relationship between marginal cost and average cost is the same as the relationship between the marginal of anything and the average of anything. For example, suppose that your microeconomics teacher has just finished grading your most recent quiz. Your average score on all of the quizzes up to that point was 92 percent, and your teacher tells you that based on your most most recent quiz your average has risen to 93 percent. What can you infer about the score on your most recent quiz? Since your average has increased, the “marginal score” (your grade on the

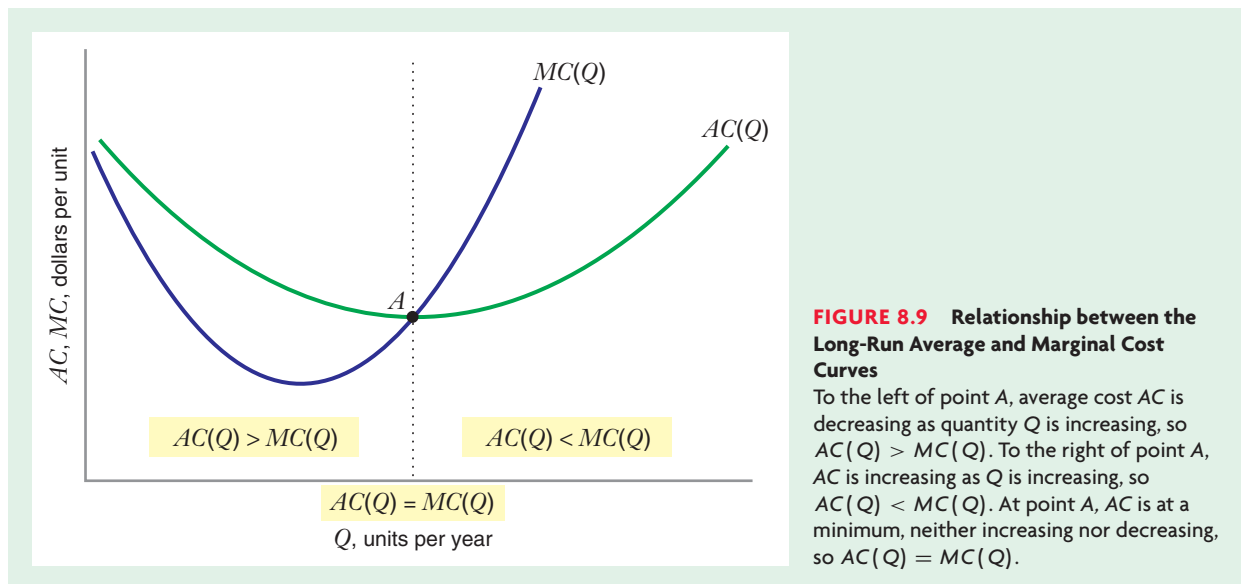


FIGURE 8.9 Relationship between the Long-Run Average and Marginal Cost Curves

To the left of point A, average cost AC is decreasing as quantity Q is increasing, so $AC(Q) > MC(Q)$. To the right of point A, AC is increasing as Q is increasing, so $AC(Q) < MC(Q)$. At point A, AC is at a minimum, neither increasing nor decreasing, so $AC(Q) = MC(Q)$.

APPLICATION 8.2

The Costs of Higher Education

How big is your college or university? Is it a large school, such as Ohio State, or a smaller one, such as Northwestern? At which school is the cost per student likely to be lower? Does university size affect the long-run average and marginal cost of “producing” education?

Rajindar and Manjulika Koshal have studied how school size affects the average and marginal cost of education.⁵ They collected data on the average cost per student from 195 U.S. universities from 1990 to 1991 and estimated an average cost curve for these universities.⁶ To control for differences in cost that stem from differences among universities in terms of their commitment to graduate programs, the Koshals estimated average cost curves for four groups of universities, primarily distinguished by the number of Ph.D.s awarded per year and the amount of government funding for Ph.D. students these universities received. For simplicity, we discuss the cost curves for the category that includes the 66 universities nationwide with the largest graduate programs (e.g., schools like Harvard, Northwestern, and the University of California at Berkeley).

Figure 8.10 shows the estimated average and marginal cost curves for this category of schools. It shows that the average cost per student declines until enrollment reaches about 30,000 full-time undergraduate students (about the size of Indiana University, for example). Because few universities are this large, the Koshals’ research suggests that for most universities in the United States with large graduate programs, the marginal cost of an additional undergraduate student is less than the average cost per student, and thus an increase in the size of the undergraduate student body would reduce the cost per student.

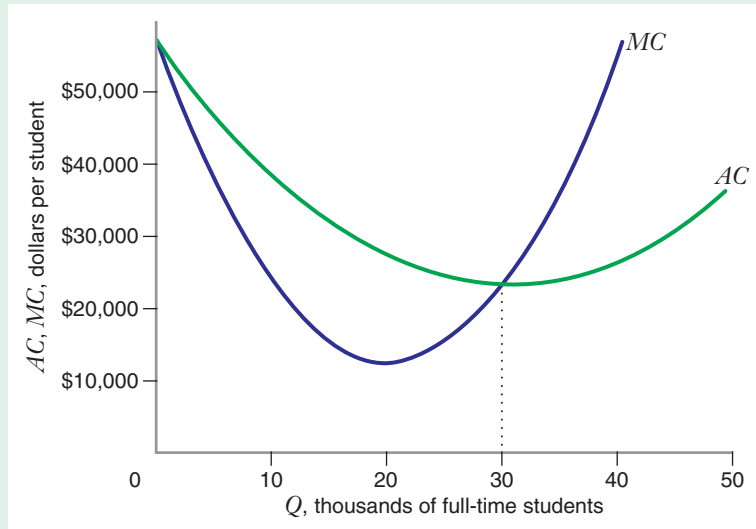
This finding seems to make sense. Think about your university. It already has a library and buildings for classrooms. It already has a president and a staff to run the school. These costs will probably not go up much if more students are added. Adding students is, of course, not costless. For example, more classes might have to be added. But it is not *that* difficult to find people who are able and willing to teach university classes (e.g., graduate students). Until the point is reached at which more dormitories or additional classrooms are needed, the extra costs of more students are not likely to be that large. Thus, for

⁵R. Koshal and M. Koshal, “Quality and Economies of Scale in Higher Education,” *Applied Economics* 27 (1995): 773–778.

⁶To control for variations in cost that might be due to differences in academic quality, their analysis also allowed average cost to depend on the student–faculty ratio and the academic reputation of the school, as measured by factors such as average SAT scores of entering freshmen. In Figure 8.10, these variables are assumed to be equal to their national averages.

FIGURE 8.10 The Long-Run Average and Marginal Cost Curves for Undergraduate Education at U.S. Universities

The marginal cost of an additional student is less than the average cost per student until enrollment reaches about 30,000 students. Until that point, average cost per student falls with the number of students. Beyond that point, the marginal cost of an additional student exceeds the average cost per student, and average cost increases with the number of students.



the typical university, while the *average* cost per student might be fairly high, the *marginal* cost of matriculating an

additional student is often fairly low. If so, average cost will decrease as the number of students increases.

most recent quiz) must be above your average. If your average had fallen to 91 percent, it would have been because your most recent quiz score was below your average. If your average had remained the same, the reason would have been that the score on your most recent quiz was equal to your average.

Economies and Diseconomies of Scale

The change in long-run average cost as output increases is the basis for two important concepts: economies of scale and diseconomies of scale. A firm enjoys **economies of scale** in a situation where average cost goes down when output goes up. By contrast, a firm suffers from **diseconomies of scale** in the opposite situation, where average cost goes up when output goes up. The extent of economies of scale can affect the structure of an industry. Economies of scale can also explain why some firms are more profitable than others in the same industry. Claims of economies of scale are often used to justify mergers between two firms producing the same product.⁷

Figure 8.11 illustrates economies and diseconomies of scale by showing a long-run average cost curve that many economists believe typifies many real-world production processes. For this average cost curve, there is an initial range of economies of scale (0 to Q'), followed by a range over which average cost is flat (Q' to Q''), and then a range of diseconomies of scale ($Q > Q''$).

Economies of scale have various causes. They may result from the physical properties of processing units that give rise to increasing returns to scale in inputs (e.g., as in the case of oil pipelines, discussed in Application 6.6 of Chapter 6). Economies of scale can also arise due to specialization of labor. As the number of workers increases

economies of scale A characteristic of production in which average cost decreases as output goes up.

diseconomies of scale A characteristic of production in which average cost increases as output goes up.

⁷See Chapter 4 of F. M. Scherer and D. Ross, *Industrial Market Structure and Economic Performance* (Boston: Houghton Mifflin, 1990) for a detailed discussion of the implications of economies of scale for market structure and firm performance.

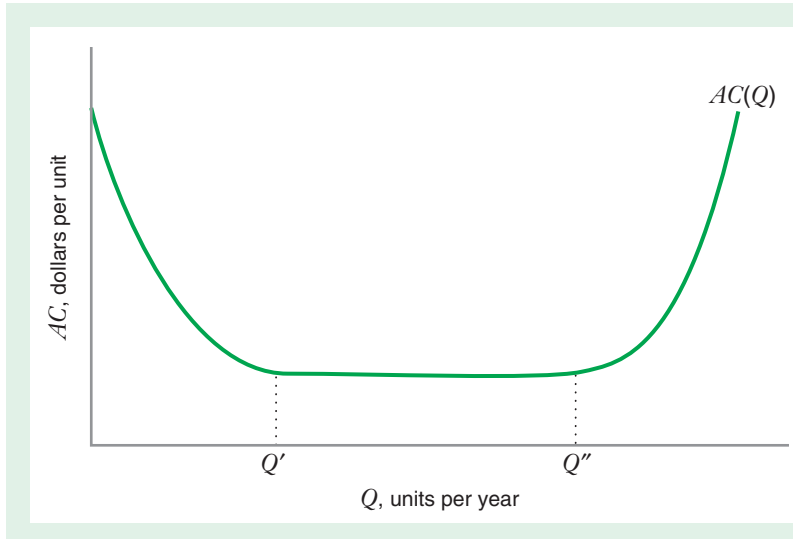


FIGURE 8.11 Economies and Diseconomies of Scale for a Typical Real-World Average Cost Curve

There are economies of scale for outputs less than Q' . Average costs are flat between Q' and Q'' , and there are diseconomies of scale thereafter. The output level Q' is called the minimum efficient scale.

with the output of the firm, workers can specialize on tasks, which often increases their productivity. Specialization can also eliminate time-consuming changeovers of workers and equipment. This, too, would increase worker productivity and lower unit costs.

Economies of scale may also result from the need to employ **indivisible inputs**. An indivisible input is an input that is available only in a certain minimum size; its quantity cannot be scaled down as the firm's output goes to zero. An example of an indivisible input is a high-speed packaging line for breakfast cereal. Even the smallest such lines have huge capacity, 14 million pounds of cereal per year. A firm that might only want to produce 5 million pounds of cereal a year would still have to purchase the services of this indivisible piece of equipment.

Indivisible inputs lead to decreasing average costs (at least over a certain range of output) because when a firm purchases the services of an indivisible input, it can "spread" the cost of the indivisible input over more units of output as output goes up. For example, a firm that purchases the services of a minimum-scale packaging line to produce 5 million pounds of cereal per year will incur the same total cost on this input when it increases production to 10 million pounds of cereal per year.⁸ This will drive the firm's average costs down.

The region of diseconomies of scale (e.g., the region where output is greater than Q'' in Figure 8.11) is usually thought to occur because of **managerial diseconomies**. Managerial diseconomies arise when a given percentage increase in output forces the firm to increase its spending on the services of managers by more than this percentage. To see why managerial diseconomies of scale can arise, imagine an enterprise whose success depends on the talents or insight of one key individual (e.g., the entrepreneur who started the business). As the enterprise grows, that key individual's contribution to the business cannot be replicated by any other single manager. The firm may have to employ so many additional managers that total costs increase at a faster rate than output, which then pushes average costs up.

The smallest quantity at which the long-run average cost curve attains its minimum point is called the **minimum efficient scale**, or MES (in Figure 8.11, the MES occurs at output Q'). The size of MES relative to the size of the market often indicates

indivisible input An input that is available only in a certain minimum size. Its quantity cannot be scaled down as the firm's output goes to zero.

managerial diseconomies A situation in which a given percentage increase in output forces the firm to increase its spending on the services of managers by more than this percentage.

minimum efficient scale The smallest quantity at which the long-run average cost curve attains its minimum point.

⁸Of course, it may spend more on other inputs, such as raw materials, that are not indivisible.

TABLE 8.1 MES as a Percentage of Industry Output for Selected U.S. Food and Beverage Industries*

Industry	MES as % of Output	Industry	MES as % of Output
Beet sugar	1.87	Breakfast cereal	9.47
Cane sugar	12.01	Mineral water	0.08
Flour	0.68	Roasted coffee	5.82
Bread	0.12	Pet food	3.02
Canned vegetables	0.17	Baby food	2.59
Frozen food	0.92	Beer	1.37
Margarine	1.75		

*Source: Table 4.2 in J. Sutton, *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration* (Cambridge, MA: MIT Press, 1991).

the significance of economies of scale in particular industries. The larger MES is in comparison to overall market sales, the greater the magnitude of economies of scale. Table 8.1 shows MES as a percentage of total industry output for a selected group of U.S. food and beverage industries.⁹ The industries with the largest MES-market size

APPLICATION 8.3

Economies of Scale in Refining Alumina?¹⁰

Manufacturing aluminum involves several steps, one of which is alumina refining. Alumina is a chemical compound consisting of aluminum and oxygen atoms (Al_2O_3). Alumina is created when bauxite ore—the basic raw material used to produce aluminum—is transformed using a technology known as the Bayer process.

There are substantial economies of scale in the refining of alumina. Table 8.2 shows estimated long-run average costs as a function of the capacity of an alumina refinery. As plant capacity doubles from 150,000 tons per year to 300,000 tons per year, long-run average cost declines by about 12 percent. Stuckey reports that average costs in alumina refining may continue to fall up to capacities of 500,000. If so, then the minimum efficient scale of an alumina refinery would occur at an output of 500,000 tons per year.

TABLE 8.2 Plant Capacity and Long-Run Average Cost in Alumina Refining*

Plant Capacity (tons)	Index of Average Cost (equals 100 at 300,000 tons)
55,000	139
90,000	124
150,000	114
300,000	100

*Source: Table 1-1 in Stuckey, *Vertical Integration and Joint Ventures in the Aluminum Industry* (Cambridge, MA: Harvard University Press, 1983).

If firms understand this, we would expect most alumina plants to have capacities of at least 500,000 tons per year. In fact, this is true. In 1979, the average capacity of the 10 alumina refineries in North America was 800,000 tons per year, and only two were under 500,000 tons per year. No alumina refinery's capacity exceeded 1.3 million tons per year. This suggests that diseconomies of scale set in at about this level of output.

⁹In this table, MES is measured as the capacity of the median plant in an industry. The median plant is the plant whose capacity lies exactly in the middle of the range of capacities of plants in an industry. That is, 50 percent of all plants in a particular industry have capacities that are smaller than the median plant in that industry, and 50 percent have capacities that are larger. Estimates of MES based on the capacity of the median plant correlate highly with “engineering estimates” of MES that are obtained by asking well-informed manufacturing and engineering personnel to provide educated estimates of minimum efficient scale plant sizes. Data on median plant size in U.S. industries are available from the U.S. Census of Manufacturing.

¹⁰The information in this example draws from J. Stuckey, *Vertical Integration and Joint Ventures in the Aluminum Industry* (Cambridge, MA: Harvard University Press, 1983), especially pp. 12–14.

TABLE 8.3 Relationship between Economies of Scale and Returns to Scale

	Production Function		
	$Q = L^2$	$Q = \sqrt{L}$	$Q = L$
Labor requirements function	$L = \sqrt{Q}$	$L = Q^2$	$L = Q$
Long-run total cost	$TC = w\sqrt{Q}$	$TC = wQ^2$	$TC = wQ$
Long-run average cost	$AC = w/\sqrt{Q}$	$AC = wQ$	$AC = w$
How does long-run average cost vary with Q ?	Decreasing	Increasing	Constant
Economies/diseconomies of scale?	Economies of scale	Diseconomies of scale	Neither
Returns to scale?	Increasing	Decreasing	Constant

ratios are breakfast cereal and cane sugar refining. These industries have significant economies of scale. The industries with the lowest MES-market size ratios are mineral water and bread. Economies of scale in manufacturing in these industries appear to be weak.

Economies of Scale and Returns to Scale

Economies of scale and returns to scale are closely related, because the returns to scale of the production function determine how long-run average cost varies with output. Table 8.3 illustrates these relationships with respect to three production functions where output Q is a function of a single input, quantity of labor L . The table shows each production function and the corresponding labor requirements function (which specifies the quantity of labor needed to produce a given quantity of output, as discussed in Chapter 6), as well as the expressions for total cost and long-run average cost given a price of labor w .

The relationships illustrated in Table 8.3 between economies of scale and returns to scale can be summarized as follows:

- If average cost decreases as output increases, we have *economies of scale* and *increasing returns to scale* (e.g., production function $Q = L^2$ in Table 8.3).
- If average cost increases as output increases, we have *diseconomies of scale* and *decreasing returns to scale* (e.g., production function $Q = \sqrt{L}$ in Table 8.3).
- If average cost stays the same as output increases, we have *neither economies nor diseconomies of scale* and *constant returns to scale* (e.g., production function $Q = L$ in Table 8.3).

Measuring the Extent of Economies of Scale: The Output Elasticity of Total Cost

In Chapter 2 you learned that elasticities of demand, such as the price elasticity of demand or income elasticity of demand, tell us how sensitive demand is to the various factors that drive demand, such as price or income. We can also use elasticities to tell us how sensitive total cost is to the factors that influence it. An important cost elasticity is the **output elasticity of total cost**, denoted by $\epsilon_{TC,Q}$. It is defined as the percentage change in total cost per 1 percent change in output:

$$\epsilon_{TC,Q} = \frac{\frac{\Delta TC}{TC}}{\frac{\Delta Q}{Q}} = \frac{\frac{\Delta TC}{\Delta Q}}{\frac{TC}{Q}}$$

output elasticity of total cost The percentage change in total cost per 1 percent change in output.

APPLICATION 8.4

Hospitals Are Businesses Too

The business of health care has been in the news a lot during the 1990s and early 2000s. One of the most interesting trends was the consolidation of hospitals through mergers. In the Chicago area, for example, Northwestern Memorial Hospital merged with several suburban hospitals, such as Evanston Hospital, to form a large multihospital system covering the North Side of Chicago and the North Shore Suburbs.

Proponents of hospital mergers argue that mergers enable hospitals to achieve cost savings through economies of scale in “backoffice” operations—activities such as laundry, housekeeping, cafeterias, printing and duplicating services, and data processing that do not generate revenue for a hospital directly, but that no hospital can function without. Opponents argue that such cost savings are illusory and that hospital mergers mainly reduce competition in local hospital markets. The U.S. antitrust authorities have blocked several hospital mergers on this basis.

David Dranove has studied the extent to which backoffice activities within a hospital are subject to economies of scale.¹¹ Figure 8.12 summarizes some of his

findings. The figure shows the long-run average cost curves for three different activities: cafeterias, printing and duplicating, and data processing. Output is measured as the annual number of patients who are discharged by the hospital. (For each activity, average cost is normalized to equal an index of 1.0, at an output of 10,000 patients per year.) These figures show that economies of scale vary from activity to activity. Cafeterias are characterized by significant economies of scale. For printing and duplicating, the average cost curve is essentially flat. And for data processing, diseconomies of scale arise at a fairly low level of output. Overall, averaging the 14 backoffice activities that he studied, Dranove found that there are economies of scale in these activities, but they are largely exhausted at an output of about 7500 patient discharges per year. This would correspond to a hospital with 200 beds, which is medium-sized by today’s standards.

Dranove’s analysis shows that a merger of two large hospitals would be unlikely to achieve additional economies of scale in backoffice operations. This suggests that claims that hospital mergers generally reduce costs per patient should be viewed with skepticism, unless both merging hospitals are small.

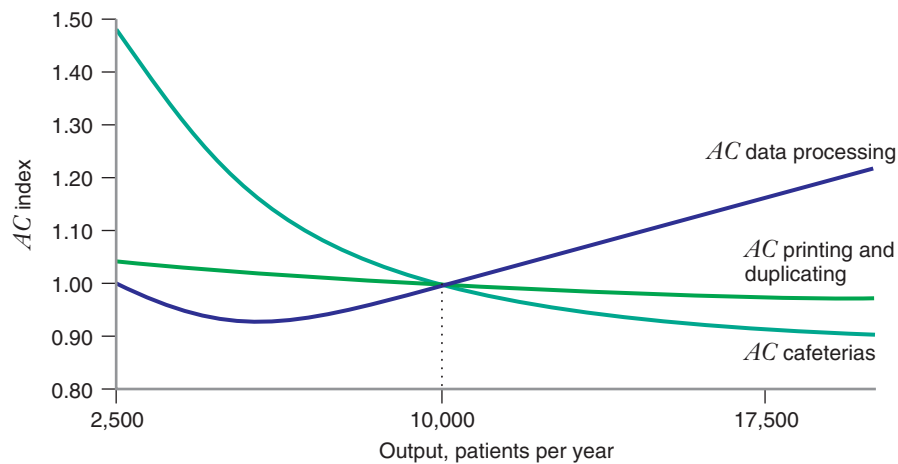


FIGURE 8.12 Average Cost Curves for Three “Backoffice” Activities in a Hospital

Cafeterias exhibit significant economies of scale. Data processing exhibits diseconomies of scale beyond an output of about 5000 patients per year. And the average cost curve for printing and duplicating is essentially flat (i.e., there are no significant economies or diseconomies of scale in this activity).

¹¹“Economies of Scale in Non-Revenue Producing Cost Centers: Implications for Hospital Mergers,” *Journal of Health Economics* 17 (1998): 69–83.

TABLE 8.4 Relationship between Output Elasticity of Total Cost and Economies of Scale

Value of $\epsilon_{TC,Q}$	MC Versus AC	How AC Varies as Q Increases	Economies/ Diseconomies of Scale
$\epsilon_{TC,Q} < 1$	$MC < AC$	Decreases	Economies of scale
$\epsilon_{TC,Q} > 1$	$MC > AC$	Increases	Diseconomies of scale
$\epsilon_{TC,Q} = 1$	$MC = AC$	Constant	Neither

TABLE 8.5 Estimates of the Output Elasticities for Selected Manufacturing Industries in India

Industry	Output Elasticity of Total Cost
Iron and steel	0.553
Cotton textiles	1.211
Cement	1.162
Electricity and gas	0.3823

Since $\Delta TC/\Delta Q =$ marginal cost (MC) and $TC/Q =$ average cost (AC),

$$\epsilon_{TC,Q} = \frac{MC}{AC}$$

Thus, the output elasticity of total cost is equal to the ratio of marginal to average cost.

As we have noted (see page 268), the relationship between long-run average and marginal cost corresponds with the way average cost AC varies with output quantity Q . This means that output elasticity of total cost tells us the extent of economies of scale, as shown in Table 8.4.

Output elasticity of total cost is often used to characterize the extent of economies of scale in different industries. Table 8.5, for example, shows the results of a study that estimated the output elasticity of total cost for several manufacturing industries in India.¹² Iron and steel industries and electricity and gas industries have output elasticities significantly less than 1, indicating the presence of significant economies of scale. By contrast, textile and cement firms' output elasticities are a little higher than 1, indicating slight diseconomies of scale.¹³

SHORT-RUN TOTAL COST CURVE

The long-run total cost curve shows how the firm's minimized total cost varies with output when the firm is free to adjust all its inputs. The **short-run total cost curve** $STC(Q)$ tells us the minimized total cost of producing Q units of output when at least one input is fixed at a particular level. In the following discussion we assume that the

¹²R. Jha, M. N. Murty, S. Paul, and B. Bhaskara Rao, "An Analysis of Technological Change, Factor Substitution, and Economies of Scale in Manufacturing Industries in India," *Applied Economics* 25 (October 1993): 1337–1343. The estimated output elasticities are reported in Table 5.

¹³The estimated output elasticities for textiles and cement are not *statistically* different from 1. Thus, these industries might be characterized by constant returns to scale.

8.2 SHORT-RUN COST CURVES

short-run total cost curve A curve that shows the minimized total cost of producing a given quantity of output when at least one input is fixed.

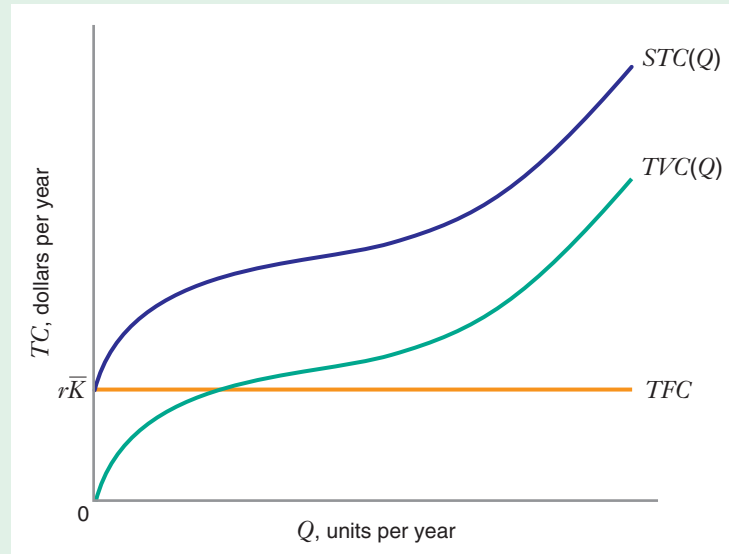


FIGURE 8.13 Short-Run Total Cost Curve

The short-run total cost curve $STC(Q)$ is the sum of the total variable cost curve $TVC(Q)$ and the total fixed cost curve TFC . Total fixed cost is equal to the cost $r\bar{K}$ of the fixed capital services.

total variable cost curve A curve that shows the sum of expenditures on variable inputs, such as labor and materials, at the short-run cost-minimizing input combination.

total fixed cost curve A curve that shows the cost of fixed inputs and does not vary with output.

amount of capital used by the firm is fixed at \bar{K} . The short-run total cost curve is the sum of two components: the **total variable cost curve** $TVC(Q)$ and the **total fixed cost curve** TFC —that is, $STC(Q) = TVC(Q) + TFC$. The total variable cost curve $TVC(Q)$ is the sum of expenditures on variable inputs, such as labor and materials, at the short-run cost-minimizing input combination. Total fixed cost is equal to the cost of the fixed capital services (i.e., $TFC = r\bar{K}$) and thus does not vary with output. Figure 8.13 shows a graph of the short-run total cost curve, the total variable cost curve, and the total fixed cost curve. Because total fixed cost is independent of output, its graph is a horizontal line with the value $r\bar{K}$. Thus, $STC(Q) = TVC(Q) + r\bar{K}$, which means that the vertical distance between $STC(Q)$ and $TVC(Q)$ is equal to $r\bar{K}$ at every quantity Q .

LEARNING-BY-DOING EXERCISE 8.3



Deriving a Short-Run Total Cost Curve

Let us return to the production function in Learning-By-Doing Exercises 7.2, 7.4, 7.5, and 8.1, $Q = 50\sqrt{LK}$.

Problem What is the short-run total cost curve for this production function when capital is fixed at a level \bar{K} and the input prices of labor and capital are $w = 25$ and $r = 100$, respectively?

Solution In Learning-By-Doing Exercise 7.5, we derived the short-run cost-minimizing quantity of labor when capital was fixed at \bar{K} : $L = Q^2/(2500\bar{K})$. We can obtain the short-run total cost curve directly from this solution: $STC(Q) = wL + r\bar{K} = Q^2/(100\bar{K}) + 100\bar{K}$. The total variable and total fixed cost curves follow: $TVC(Q) = Q^2/(100\bar{K})$ and $TFC = 100\bar{K}$.

Note that, holding Q constant, total variable cost is decreasing in the quantity of capital \bar{K} . The reason is that, for a given amount of output, a firm that uses more capital

can reduce the amount of labor it employs. Since TVC is the firm's labor expense, it follows that TVC should decrease in \bar{K} .

Similar Problems: 8.12 and 8.13

RELATIONSHIP BETWEEN THE LONG-RUN AND THE SHORT-RUN TOTAL COST CURVES

Consider again a firm that uses just two inputs, labor and capital. In the long run, the firm can freely vary the quantity of both inputs, but in the short run the quantity of capital is fixed. Thus, the firm is more constrained in the short run than in the long run, so it makes sense that it will be able to achieve lower total costs in the long run.

Figure 8.14 shows a graphical analysis of the long-run and short-run cost-minimization problems for a producer of television sets in this situation. Initially, the firm wants to produce 1 million television sets per year. In the long run, when it is free to vary both capital and labor, it minimizes total cost by operating at point A , using L_1 units of labor and K_1 units of capital.

Suppose the firm wants to increase its output to 2 million TVs per year and that, in the short run, its usage of capital must remain fixed at K_1 . In that case, the firm would operate at point B , using L_3 units of labor and the same K_1 units of capital. In the long run, however, the firm could move along the expansion path and operate at point C , using L_2 units of labor and the same K_2 units of capital. Since point B is on a higher isocost line than point C , the short-run total cost is higher than the long-run total cost when the firm is producing 2 million TVs per year.

When the firm is producing 1 million TVs per year, point A is cost minimizing in both the long run and the short run, if the short-run constraint is K_1 units of capital. Figure 8.15 shows the firm's corresponding long-run and short-run total cost curves

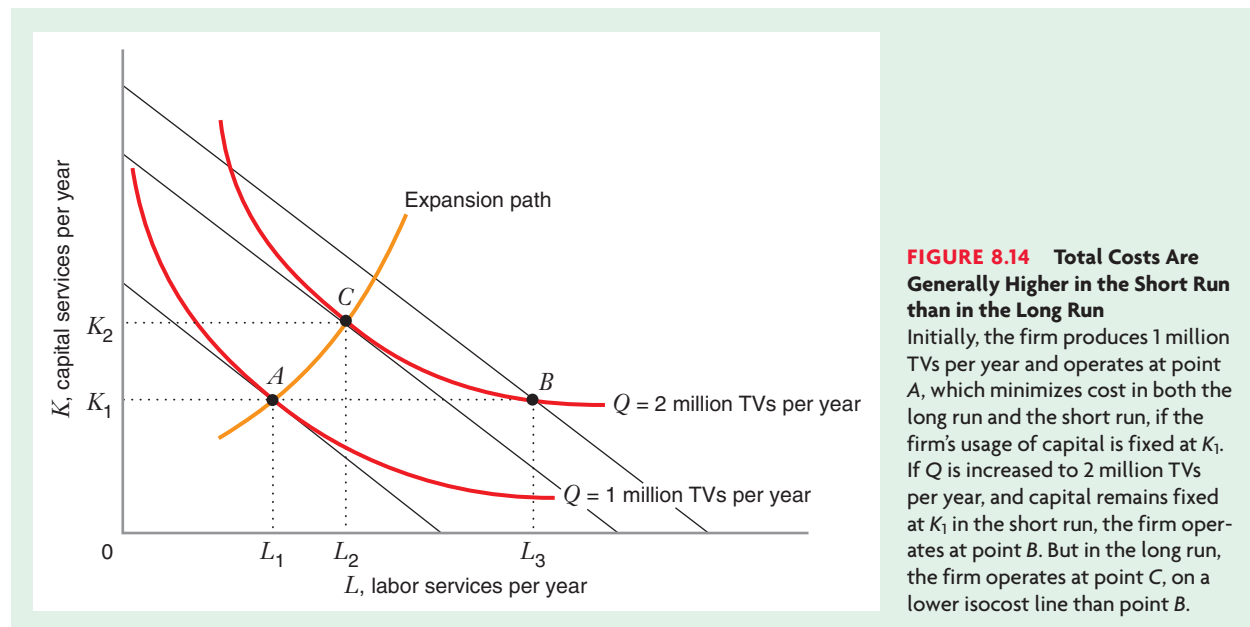
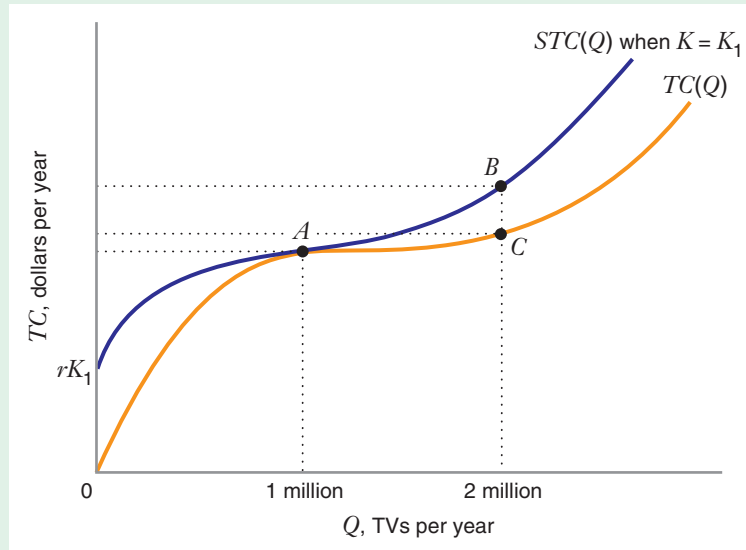


FIGURE 8.15 Relationship between Short-Run and Long-Run Total Cost Curves

When the quantity of capital is fixed at K_1 , $STC(Q)$ is always above $TC(Q)$, except at point A . Point A solves both the long-run and the short-run cost-minimization problem when the firm produces 1 million TVs per year.



$TC(Q)$ and $STC(Q)$. We see that $STC(Q)$ always lies above $TC(Q)$ (i.e., short-run total cost is greater than long-run total cost) except at point A , where $STC(Q)$ and $TC(Q)$ are equal.

SHORT-RUN AVERAGE AND MARGINAL COST CURVES

Just as we can define long-run average and long-run marginal cost curves (see page 266) we can also define the curves for **short-run average cost** (SAC) and **short-run marginal cost** (SMC): $SAC(Q) = [STC(Q)]/Q$ and $SMC(Q) = (\Delta STC)/(\Delta Q)$. Thus, just as long-run marginal cost is equal to the slope of the long-run total cost curve, short-run marginal cost is equal to the slope of the short-run total cost curve. (Note that in Figure 8.15 at point A , when output equals 1 million units per year, the slopes of the long-run total cost and short-run total cost curves are equal. It therefore follows that at this level of output, not only does $STC = TC$, but $SMC = MC$.)

In addition, just as we can break short-run total cost into two pieces (total variable cost and total fixed cost), we can break short-run average cost into two pieces: **average variable cost** (AVC) and **average fixed cost** (AFC): $SAC = AVC + AFC$. Average fixed cost is total fixed cost per unit of output ($AFC = TFC/Q$). Average variable cost is total variable cost per unit of output ($AVC = TVC/Q$).

Figure 8.16 illustrates typical graphs of the short-run marginal, short-run average cost, average variable cost, and average fixed cost curves. We obtain the short-run average cost curve by “vertically summing” the average variable cost curve and the average fixed cost curve.¹⁴ The average fixed cost curve decreases everywhere and approaches the horizontal axis as Q becomes very large. This reflects the fact that as output increases, fixed capital costs are “spread out” over an increasingly large volume of output, driving fixed costs per unit downward toward zero. Because AFC becomes

short-run average cost

The firm's total cost per unit of output when it has one or more fixed inputs.

short-run marginal cost

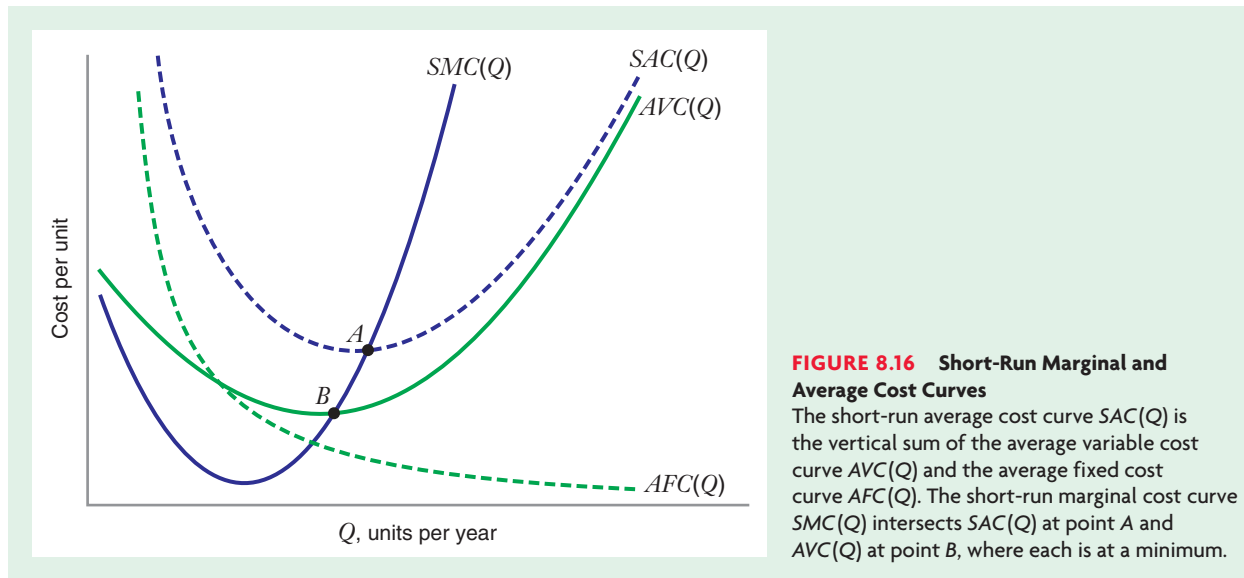
The slope of the short-run total cost curve.

average variable cost

Total variable cost per unit of output.

average fixed cost Total fixed cost per unit of output.

¹⁴Vertically summing means that, for any Q , we find the height of the SAC curve by adding together the heights of the AVC and AFC curves at that quantity.



smaller and smaller as Q increases, the $AVC(Q)$ and $SAC(Q)$ curves get closer and closer together. The short-run marginal cost curve $SMC(Q)$ intersects the short-run average cost curve and the average variable cost curve at the minimum point of each curve. This property mirrors the relationship between the long-run marginal and long-run average cost curves (see page 268), again reflecting the relationship between the average and marginal measures of anything.

RELATIONSHIPS BETWEEN THE LONG-RUN AND THE SHORT-RUN AVERAGE AND MARGINAL COST CURVES

The Long-Run Average Cost Curve as an Envelope Curve

The long-run average cost curve forms a boundary (or *envelope*) around the set of short-run average cost curves corresponding to different levels of output and fixed input. Figure 8.17 illustrates this for a producer of television sets. The firm's long-run average cost curve $AC(Q)$ is U-shaped, as are its short-run average cost curves $SAC_1(Q)$, $SAC_2(Q)$, and $SAC_3(Q)$, which correspond to different levels of fixed capital K_1 , K_2 , and K_3 (where $K_1 < K_2 < K_3$). (Moving to an increased level of fixed capital might mean increasing the firm's plant size or its degree of automation.)

The short-run average cost curve corresponding to any level of fixed capital lies above the long-run curve except at the level of output for which the fixed capital is optimal (points A, B, and D in the figure). Thus, the firm would minimize its costs when producing 1 million TVs if its level of fixed capital were K_1 , but if it expanded its output to 2 million or 3 million TVs, it would minimize costs if its level of fixed capital were K_2 or K_3 , respectively. (In practice, if K represents plant size, the firm's high short-run average cost of \$110 to produce 2 million TVs using fixed capital K_1 might reflect reductions in the marginal product of labor resulting from crowding too many workers into a small plant. To achieve the minimal average cost of \$35, the firm would have to increase its plant size to K_2 .)

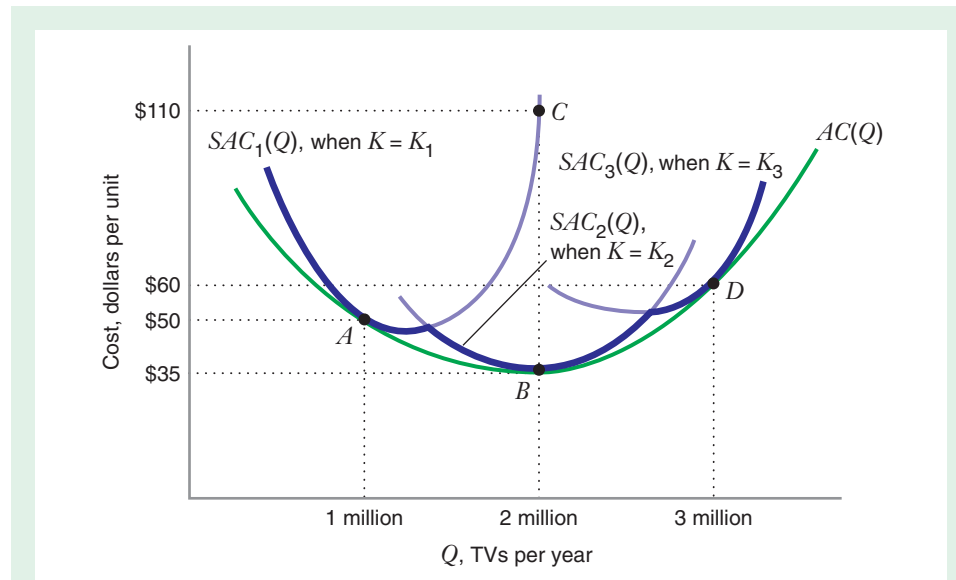


FIGURE 8.17 The Long-Run Average Cost Curve as an Envelope Curve

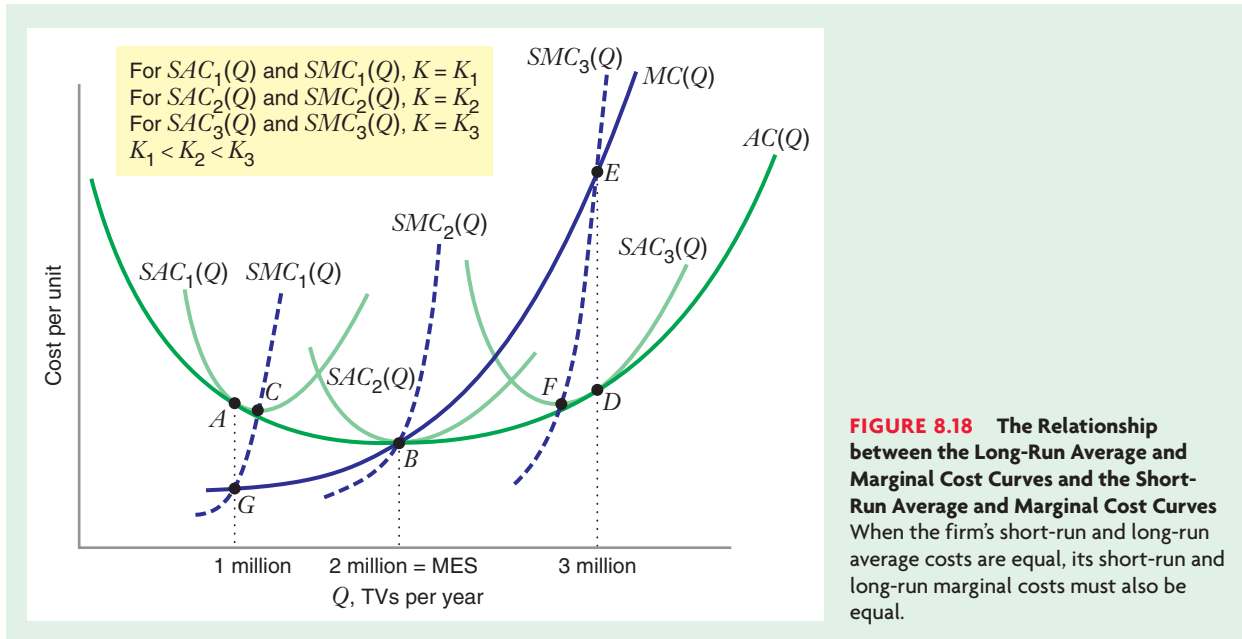
The short-run average cost curves $SAC_1(Q)$, $SAC_2(Q)$, and $SAC_3(Q)$, lie above the long-run average cost curve $AC(Q)$ except at points A, B, and D. This shows that short-run average cost is always greater than long-run average cost except at the level of output for which a plant size (K_1 , K_2 , or K_3) is optimal. Point C shows where the firm would operate in the short run if it produced 2 million TV sets per year with capital remaining fixed at K_1 . If the figure included progressively more short-run curves, the dark scalloped lower boundary of the short-run curves would smooth out and ultimately coincide with the long-run curve.

Now observe the dark scalloped lower boundary of the short-run cost curves in Figure 8.17, and imagine that the figure included more and more short-run curves. The dark boundary would become progressively smoother (i.e., with increasingly many shallow scallops instead of a few deep scallops), and as the number of short-run curves grew larger the dark curve would more and more closely approximate the long-run curve. Thus, you can think of the long-run curve as the lower envelope of an infinite number of short-run curves. That's why the long-run average cost curve is sometimes referred to as the *envelope curve*.

When Are Long-Run and Short-Run Average and Marginal Costs Equal, and When Are They Not?

The curves shown in Figure 8.18 are the same as those in Figure 8.17, but with the addition of the long-run marginal cost curve $MC(Q)$ and the three short-run marginal cost curves $SMC_1(Q)$, $SMC_2(Q)$, and $SMC_3(Q)$. Figure 8.18 shows the special relationships between the short-run average and marginal cost curves and the long-run average and marginal cost curves. As we have seen, if the firm is required to produce 1 million units, in the long run it would choose a plant size K_1 . Therefore, if the firm has a fixed plant of size K_1 , the combination of inputs it would use to produce 1 million units in the short run is the same as the combination it would choose in the long run. At an output of 1 million units not only are $SAC_1(Q)$ and $AC(Q)$ equal (at point A), but also $SMC_1(Q)$ and $MC(Q)$ are equal (at point G).

Similar relationships hold at all levels of output. For example, if the firm has a fixed plant of size K_3 , it can produce 3 million units as efficiently in the short run as it



can in the long run. Therefore $SAC_3(Q)$ and $AC(Q)$ are equal (at point D), and $SMC_3(Q)$ and $MC(Q)$ are also equal (at point E).

Figure 8.18 also illustrates another feature of short-run average cost curves that you may find surprising. A short-run average cost curve does not generally reach its minimum at the output where short-run and long-run average costs are equal. For example, at point A , $SAC_1(Q)$ and $AC(Q)$ are equal, and they are both downward sloping. $SAC_1(Q)$ must be falling because $SMC_1(Q)$ lies below $SAC_1(Q)$. The minimum of $SAC_1(Q)$ occurs at point C , where $SMC_1(Q)$ equals $SAC_1(Q)$. Similarly, at point D , $SAC_3(Q)$ and $AC(Q)$ are equal and have the same upward slope. $SAC_3(Q)$ must be rising because $SMC_3(Q)$ lies above $SAC_3(Q)$. The minimum of $SAC_3(Q)$ occurs at point F , where $SMC_3(Q)$ equals $SAC_3(Q)$.

The figure also illustrates that it is possible for a short-run average cost curve to reach its minimum at the output where short-run and long-run average costs are equal. For example, at point B , $SAC_2(Q)$ and $AC(Q)$ are equal, and they both achieve a minimum. $SAC_2(Q)$ must have a slope of zero because $SMC_2(Q)$ passes through $SAC_2(Q)$ at B .

LEARNING-BY-DOING EXERCISE 8.4

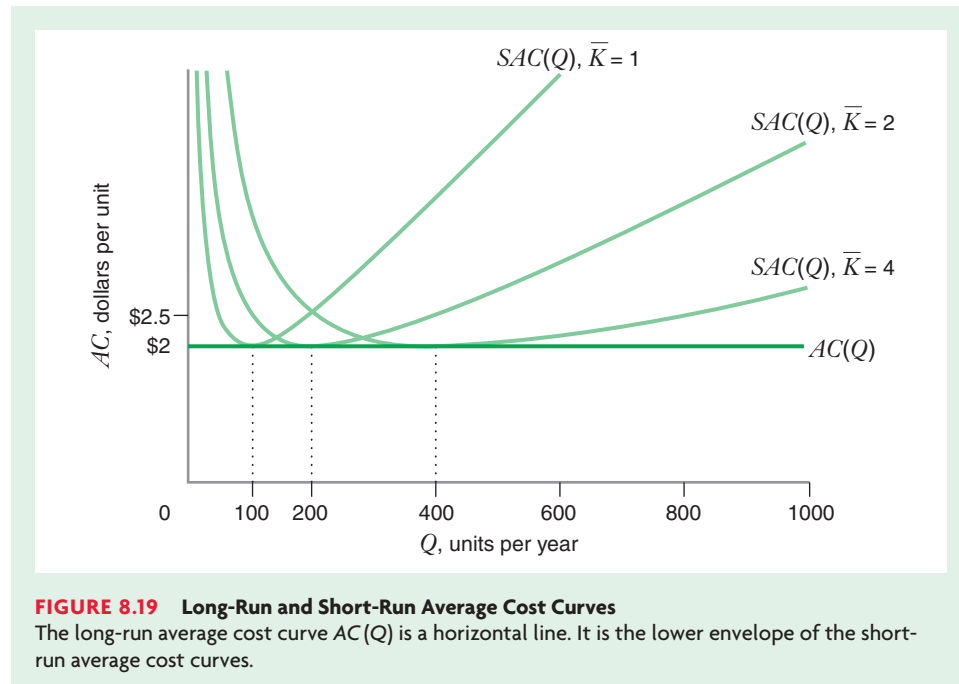


The Relationship between Short-Run and Long-Run Average Cost Curves

Let us return to the production function in Learning-By-Doing Exercises 8.1, 8.2, and 8.3: $Q = 50\sqrt{LK}$.

Problem What is the short-run average cost curve for this production function for a fixed level of capital \bar{K} and input prices $w = 25$ and $r = 100$? Sketch a graph of the short-run average cost curve for levels of capital $\bar{K} = 1$, $\bar{K} = 2$, and $\bar{K} = 4$.

Solution We derived the short-run total cost curve for this production function in Learning-By-Doing Exercise 8.3: $STC(Q) = Q^2/(100\bar{K}) + 100\bar{K}$. Thus, the short-run



average cost curve is $SAC(Q) = Q/(100\bar{K}) + 100\bar{K}/Q$. Figure 8.19 shows graphs of the short-run average cost curve for $\bar{K} = 1$, $\bar{K} = 2$, and $\bar{K} = 4$. It also shows the long-run average cost curve for this production function (derived in Learning-By-Doing Exercise 8.2). The short-run average cost curves are U-shaped, while the long-run average cost curve (a horizontal line) is the lower envelope of the short-run average cost curves.

Similar Problems: 8.18 and 8.19

APPLICATION 8.5

Tracking Railroad Costs¹⁵

The 1990s and early 2000s were an interesting time for U.S. railroads. On the positive side, the railroad industry was healthier than it had been in years, and the bankruptcies that had plagued the industry in the 1960s and 1970s were over. Some railroads, such as the Burlington Northern, had become so optimistic about the future that they had begun ambitious investments in new track. On the negative side, however, U.S. railroads had developed a generally poor reputation for service, particularly

speed of delivery. On some routes, shipping freight by train in the late 1990s took longer than it had 30 years earlier. These problems re-emerged in 2003 as the U.S. economy began to climb out of recession. Said one shipper, "I've been in the grain business 25 years and this is the worst delay I've ever seen." Part of the problem, according to industry observers, arose because the railroad industry downsized too much. During the 1980s and 1990s, U.S. railroads sold or abandoned 55,000 miles of track. According to one expert, the railroads "have too much freight trying to go over too little track."

¹⁵The first part of this example box draws from "A Long Haul: America's Railroads Struggle to Capture Their Former Glory," *The Wall Street Journal* (December 5, 1997), pp. A1 and A6, and "Railroad Logjams Threaten Boom in the Farm Belt: Delays in Grain Shipments Reduce Potential Profits, May Affect Overall Economy," *The Wall Street Journal* (December 1, 2003), pp. A1 and A6.

These concerns over the quality of rail service and how they relate to the amount of track a railroad employs might make you wonder how a railroad's production costs depend on these factors. For example, would a railroad's total variable costs go down as it adds track? If so, at what rate? Would faster service cause an increase or a decrease in a railroad's cost of operation?

One way to study these questions would be to estimate the short-run and long-run cost curves for a railroad. In the 1980s, Ronald Braeutigam, Andrew Daughety, and Mark Turnquist (hereafter BDT) undertook such a study.¹⁶ With the cooperation of the management of a large American railroad firm, BDT obtained data on costs of shipment, input prices (price of fuel, price of labor service), volume of output, and speed of service for this railroad.¹⁷ Using statistical techniques, they estimated a short-run total variable cost curve for the railroad. In the study, total variable cost is the sum of the railroad's monthly costs for labor, fuel, maintenance, rail cars, locomotives, and supplies.

Table 8.6 shows the impact on total variable costs of a hypothetical 10 percent increase in (1) traffic volume (carloads of freight per month); (2) the quantity of the railroad's track (in miles); (3) speed of service (miles

per day of loaded cars); and (4) the prices of fuel, labor, and equipment.¹⁸ You should think of track miles as a fixed input, analogous to capital in our previous discussion. A railroad cannot instantly vary the quantity or quality of its track to adjust to month-to-month variations in shipment volumes in the system and thus must regard track as a fixed input.

Table 8.6 contains several interesting findings. First, total variable cost increases with total output and with the prices of the railroad's inputs. This is consistent with the predictions of the theory you have been learning in this chapter and Chapter 7. Second, total variable costs go down as the volume of the fixed input is increased (as discussed in Learning-By-Doing Exercise 8.3). Holding volume of output and speed of service fixed, an increase in track mileage (or an increase in the quality of track, holding mileage fixed) would be expected to decrease the amount the railroad spends on variable inputs, such as labor and fuel. For example, with more track (holding output and speed fixed), the railroad would reduce the congestion of trains on its mainlines and in its train yards. As a result, it would probably need fewer dispatchers (i.e., less labor) to control the movement of trains. Third, improvements in average speed may also reduce costs. Although this impact is not large, it does suggest that improvements in service not only can benefit the railroad's consumers, but might also benefit the railroad itself through lower variable costs. For this railroad, higher speeds might reduce the use of labor (e.g., fewer train crews would be needed to haul a given amount of freight) and increase the fuel efficiency of the railroad's locomotives.

Having estimated the total variable cost function, BDT go on to estimate the long-run total and average cost curves for this railroad. They do so by finding the track mileage that, for each quantity Q , minimizes the sum of total variable costs and total fixed cost, where total fixed cost is the monthly opportunity cost to the firm's owners of a given amount of track mileage. Figure 8.20 shows the long-run average cost function estimated by BDT using this approach. It also shows two short-run average cost curves, each corresponding

TABLE 8.6 What Affects Total Variable Costs for a Railroad?*

A 10 Percent Increase in . . .	Changes Total Variable Cost by . . .
Volume of output	+3.98%
Track mileage	−2.71%
Speed of service	−0.66%
Price of fuel	+1.90%
Price of labor	+5.25%
Price of equipment	+2.85%

*Source: Adapted from Table 1 of R. R. Braeutigam, A. F. Daughety, and M. A. Turnquist, "A Firm-Specific Analysis of Economies of Density in the U.S. Railroad Industry," *Journal of Industrial Economics*, 33 (September 1984): 3–20. The percentage changes in the various factors are changes away from the average values of these factors over the period studied by BDT.

¹⁶R. R. Braeutigam, A. F. Daughety, and M. A. Turnquist, "A Firm-Specific Analysis of Economies of Density in the U. S. Railroad Industry," *Journal of Industrial Economics* 33 (September 1984); 3–20.

¹⁷The identity of the firm remained anonymous to ensure the confidentiality of its data.

¹⁸In this study, the railroad's track mileage was adjusted to reflect changes in the quality of its track over time.

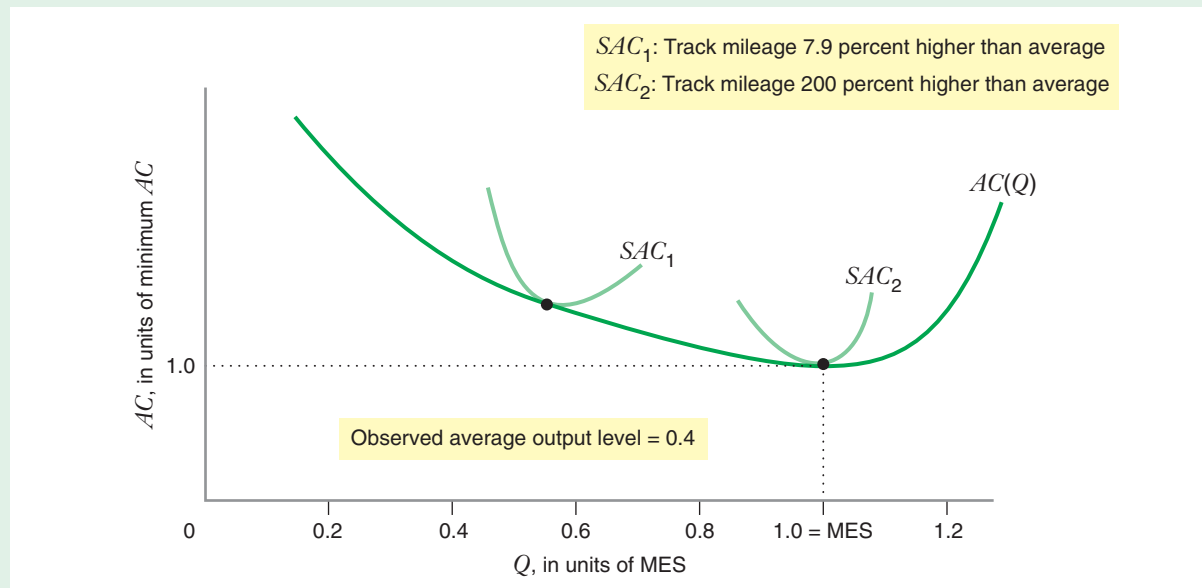


FIGURE 8.20 Long-Run and Short-Run Average Cost Curves for a Railroad

The two short-run average cost curves SAC_1 and SAC_2 correspond to a different amount of track (expressed in relation to the average amount of track observed in the data). The cost curves show that with a cost-minimizing adjustment in amount of track, this railroad could decrease its unit costs over a wide range of output above its current output level. As we have seen with other such U-shaped cost curves, the long-run curve $AC(Q)$ is the lower envelope of the short-run curves.

to a different level of track mileage. (Track mileage is stated in relation to the average track mileage observed in BDT's data.) The units of output in Figure 8.20 are expressed as a percentage of MES; the average level of output produced by the railroad at the time of the

study was about 40 percent of MES. This study thus suggests that increases in traffic volume, accompanied by cost-minimizing adjustments in track mileage, would reduce this railroad's average production costs over a wide range of output.

8.3 SPECIAL TOPICS IN COST

economies of scope A production characteristic in which the total cost of producing given quantities of two goods in the same firm is less than the total cost of producing those quantities in two single-product firms.

ECONOMIES OF SCOPE

This chapter has concentrated on cost curves for firms that produce just one product or service. In reality, though, many firms produce more than one product. For a firm that produces two products, total costs would depend on the quantity Q_1 of the first product the firm makes and the quantity Q_2 of the second product it makes. We will use the expression $TC(Q_1, Q_2)$ to denote how the firm's costs vary with Q_1 and Q_2 .

In some situations, efficiencies arise when a firm produces more than one product. That is, a two-product firm may be able to manufacture and market its products at a lower total cost than two single-product firms. These efficiencies are called **economies of scope**. Mathematically, economies of scope are present when:

$$TC(Q_1, Q_2) < TC(Q_1, 0) + TC(0, Q_2) \quad (8.1)$$

The zeros in the expressions on the right-hand side of equation (8.1) indicate that the single-product firms produce positive amounts of one good but none of the other. These expressions are sometimes called the **stand-alone costs** of producing goods 1 and 2.

Intuitively, the existence of economies of scope tells us that “variety” is more efficient than “specialization,” which we can see mathematically by representing equation (8.1) as follows: $TC(Q_1, Q_2) - TC(Q_1, 0) < TC(0, Q_2) - TC(0, 0)$. This is equivalent to equation (8.1) because $TC(0, 0) = 0$, i.e., the total cost of producing zero quantities of both products is zero. The left-hand side of this equation is the *additional cost* of producing Q_2 units of product 2 *when the firm is already producing Q_1 units of product 1*. The right-hand side of this equation is the *additional cost* of producing Q_2 *when the firm does not produce Q_1* . Economies of scope exist if it is less costly for a firm to add a product to its product line given that it already produces another product. Economies of scope would exist, for example, if it were less costly for Coca-Cola to add a cherry-flavored soft drink to its product line than it would be for a new company starting from scratch.

Why would economies of scope arise? An important reason is a firm’s ability to use a common input to make and sell more than one product. For example, BSkyB, the British satellite television company, can use the same satellite to broadcast a news channel, several movie channels, several sports channels, and several general entertainment channels.¹⁹ Companies specializing in the broadcast of a single channel would each need to have a satellite orbiting the Earth. BSkyB’s channels save hundreds of millions of dollars as compared to stand-alone channels by sharing a common satellite. Another example is Eurotunnel, the 31-mile tunnel that runs underneath the English Channel between Calais, France, and Dover, Great Britain. The Eurotunnel accommodates both highway and rail traffic. Two separate tunnels, one for highway traffic and one for rail traffic, would have been more expensive to construct and operate than a single tunnel that accommodates both forms of traffic.

stand-alone cost The cost of producing a good in a single-product firm.

ECONOMIES OF EXPERIENCE: THE EXPERIENCE CURVE

Learning-by-Doing and the Experience Curve

Economies of scale refer to the cost advantages that flow from producing a larger output at a given point in time. **Economies of experience** refer to cost advantages that result from accumulated experience over an extended period of time, or from *learning-by-doing*, as it is sometimes called. This is the reason we gave that title to the exercises in this book—they are designed to help you *learn* microeconomics *by doing* microeconomics problems.

Economies of experience arise for several reasons. Workers often improve their performance of specific tasks by performing them over and over again. Engineers often perfect product designs as they accumulate know-how about the manufacturing process. Firms often become more adept at handling and processing materials as they deepen their production experience. The benefits of learning are usually greater labor productivity (more output per unit of labor input), fewer defects, and higher material yields (more output per unit of raw material input).

economies of experience Cost advantages that result from accumulated experience, or as it is sometimes called, *learning-by-doing*.

¹⁹BSkyB is a subsidiary of Rupert Murdoch’s News Corporation.

APPLICATION 8.6

*Economies of Scope for the Swoosh*²⁰

An important source of economies of scope is marketing. A company with a well-established brand name in one product line can sometimes introduce additional products at a lower cost than a stand-alone company would be able to. This is because when consumers are unsure about a product's quality they often make inferences about its quality from the product's brand name. This can give a firm with an established brand reputation an advantage over a stand-alone firm in introducing new products. Because of its brand reputation, an established firm would not have to spend as much on advertising as the stand-alone firm to persuade consumers to try its product. This is an example of an economy of scope based on the ability of all products in a firm's product line to "share" the benefits of its established brand reputation.

A company with an extraordinary brand reputation is Nike. Nike's "swoosh," the symbol that appears on its athletic shoes and sports apparel, is one of the most recognizable marketing symbols of the modern age. Nike's swoosh is so recognizable that Nike can run television commercials that never mention its name and be confident that consumers will know whose products are being advertised.

In the late 1990s, Nike turned its attention to the sports equipment market, introducing products such as hockey sticks and golf balls. Nike's goal was to become the dominant firm in the \$40 billion per year sports equipment market by 2005. This was a bold ambition. The sports equipment market is highly fragmented, and no single company has ever dominated the entire range of product categories. In addition, while no one can deny Nike's past success in the athletic shoe and sports apparel markets, producing a high-quality hockey stick or an innovative golf ball has little in common with making sneakers or jogging clothes. It therefore seems

unlikely that Nike could attain economies of scope in manufacturing or product design.

Nike hoped to achieve economies of scope in marketing. These economies of scope would be based on its incredibly strong brand reputation, its close ties to sports equipment retailers, and its special relationships with professional athletes such as Tiger Woods and Derek Jeter. Nike's plan was to develop sports equipment that it can claim is innovative and then use its established brand reputation and its ties with the retail trade to convince consumers that its products are technically superior to existing products. If this plan works, Nike will be able to introduce its new products at far lower costs than a stand-alone company would incur to introduce otherwise identical products.

Economies of scope in marketing can be powerful, but they also have their limits. A strong brand reputation can induce consumers to try a product once, but if it does not perform as expected or if its quality is inferior, it may be difficult to penetrate the market or get repeat business. Nike's initial forays into the sports equipment market illustrate this risk. In July 1997, Nike "rolled out" a new line of roller skates at the annual sports equipment trade show in Chicago. But when a group of skaters equipped with Nike skates rolled into the parking lot, the wheels on the skates began to disintegrate! Quality problems have also arisen with a line of ice skates that Nike introduced several years ago. Jeremy Roenick, a star with the Phoenix Coyote's NHL hockey team, turned down a six-figure endorsement deal with Nike because he felt the skates were poorly designed and did not fit properly. Rumor has it that other hockey players who do have equipment deals with Nike use the products of competitors. According to one NHL equipment manager, "They're still wearing the stuff they've been wearing for years. They just slap the swoosh on it."

experience curve A relationship between average variable cost and cumulative production volume. It is used to describe the economies of experience.

Economies of experience are described by the **experience curve**, a relationship between average variable cost and cumulative production volume.²¹ A firm's cumulative production volume at any given time is the total amount of output that it has

²⁰This example is based on "Just Doing It: Nike Plans to Swoosh Into Sports Equipment But It's a Tough Game," *The Wall Street Journal* (January 6, 1998), pp. A1 and A10.

²¹The experience curve is also known as the *learning curve*.

produced over the history of the product until that time. For example, if Boeing's output of a type of jet aircraft was 30 in 2001, 45 in 2002, 50 in 2003, 70 in 2004, and 60 in 2005, its cumulative output as of the beginning of 2006 would be $30 + 45 + 50 + 70 + 60$, or 255 aircraft. A typical relationship between average variable cost and cumulative output is $AVC(N) = AN^B$, where AVC is the average variable cost of production and N denotes cumulative production volume. In this formulation, A and B are constants, where $A > 0$ and B is a negative number between -1 and 0 . The constant A represents the average variable cost of the first unit produced, and B represents the **experience elasticity**: the percentage change in average variable cost for every 1 percent increase in cumulative volume.

The magnitude of cost reductions that are achieved through experience is often expressed in terms of the **slope of the experience curve**,²² which tells us how much average variable costs go down as a percentage of an initial level when cumulative output doubles.²³ For example, if doubling a firm's cumulative output of semiconductors results in average variable cost falling from \$10 per megabyte to \$8.50 per megabyte, we would say that the slope of the experience curve for semiconductors is 85 percent, since average variable costs fell to 85 percent of their initial level. In terms of an equation,

$$\text{slope of experience curve} = \frac{AVC(2N)}{AVC(N)}$$

The slope and the experience elasticity are systematically related. If the experience elasticity is equal to B , the slope equals 2^B . Figure 8.21 shows experience curves with three different slopes: 90 percent, 80 percent, and 70 percent. The smaller the slope, the "steeper" the experience curve (i.e., the more rapidly average variable costs fall as the firm accumulates experience). Note, though, that all three curves eventually flatten out. For example, beyond a volume of $N = 40$, increments in cumulative

experience elasticity The percentage change in average variable cost for every 1 percent increase in cumulative volume.

slope of the experience curve How much average variable costs go down, as a percentage of an initial level, when cumulative output doubles.

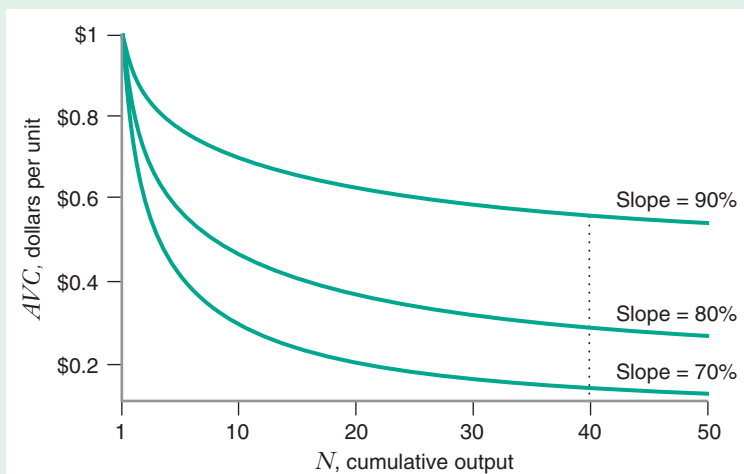


FIGURE 8.21 Experience Curves with Different Slopes

The smaller the slope, the "steeper" the experience curve, and the more rapidly average variable costs fall as cumulative output goes up. No matter what the slope, though, once cumulative experience becomes sufficiently large (e.g., $N = 40$), additional increments to experience do not lower average variable costs by much.

²²The slope of the experience curve is also known as the *progress ratio*.

²³Note that the term "slope" as used here is *not* the usual notion of the slope of a straight line.

APPLICATION 8.7

Experience Reduces Costs of Computer Chips²⁴

An interesting example of economies of experience occurs in the production of semiconductors, the memory chips that are used in personal computers, cellular telephones, and electronic games. It is widely believed that the “yield” of semiconductor chips—the ratio of usable chips to total chips on a silicon wafer—goes up as a firm gains production experience.²⁵ Silicon is an expensive raw material, and the cost of a chip is primarily determined by how much silicon it uses. The rate at which yields go up with experience is thus important for a semiconductor manufacturer to know.

Harald Gruber estimated the experience curve for a particular type of semiconductor: erasable programmable read-only memory (EPROM) chips. EPROM chips are used to store program code for cellular phones, pagers, modems, video games, printers, and hard disk drives. An EPROM chip differs from the more common DRAM in that it is *nonvolatile*, which means that, unlike a DRAM chip, it retains its stored data when the power is turned off.

Gruber recognized that other factors, such as economies of scale and memory capacity, could influence the average cost of producing an EPROM chip. After controlling for these factors, Gruber found evidence of economies of experience in the production of EPROM chips. His estimate of the slope of the EPROM experience curve was 78 percent. Thus, by doubling its cumulative volume of chips, an EPROM producer would expect its average variable costs to fall to 78 percent of their initial level.

This is an interesting finding. The market for EPROM chips is smaller than markets for other semiconductors, such as DRAMs. Moreover, new generations of EPROM chips are introduced frequently, typically about once every 18 months. By contrast, new generations of DRAM chips were introduced about every 3 years during the 1980s and 1990s. This suggests that it is unlikely that an EPROM manufacturer will operate on the “flat” portion of the experience curve for long. By the time a firm starts to “move down” the experience curve, a new generation of chip will have come along. This, then, implies that a firm that can achieve a head start in bringing a new generation of EPROM chips to market may achieve a significant cost advantage over slower competitors.

experience have a small impact on average variable costs, no matter what the slope of the experience curve is. At this point, most of the economies of experience are exhausted.

Experience curve slopes have been estimated for many different products. The median slope appears to be about 80 percent, implying that for the typical firm, each doubling of cumulative output reduces average variable costs to 80 percent of what they were before. Slopes vary from firm to firm and industry to industry, however, so that the slope enjoyed by any one firm for any given production process generally falls between 70 and 90 percent and may be as low as 60 percent or as high as 100 percent (i.e., no economies of experience).

Economies of Experience versus Economies of Scale

Economies of experience differ from economies of scale. Economies of scale refer to the ability to perform activities at a lower unit cost when those activities are performed on a larger scale at a given point in time. Economies of experience refer to reductions in unit costs due to accumulating experience over time. Economies of scale may be substantial even when economies of experience are minimal. This is likely to be the

²⁴This example draws from H. Gruber, “The Learning Curve in the Production of Semiconductor Memory Chips,” *Applied Economics*, 24 (August 1992): 885–894.

²⁵A *wafer* is a slice of polycrystalline silicon. A chip producer will etch hundreds of circuits onto a single wafer.

case in mature, capital-intensive production processes, such as aluminum can manufacturing. Likewise, economies of experience may be substantial even when economies of scale are minimal, as in such complex labor-intensive activities as the production of handmade watches.

Firms that do not correctly distinguish between economies of scale and experience might draw incorrect inferences about the benefits of size in a market. For example, if a firm has low average costs because of economies of scale, reductions in the current volume of production will increase unit costs. If the low average costs are the result of cumulative experience, the firm may be able to cut back current production volumes without raising its average costs.

Suppose you wanted to estimate how the total costs for a television producer varied with the quantity of its output or the magnitude of its input prices. To do this, you might want to estimate what economists call a **total cost function**. A total cost function is a mathematical relationship that shows how total costs vary with the factors that influence total costs. These factors are sometimes called **cost drivers**. We've spent much of this chapter analyzing two key cost drivers: input prices and scale (volume of output). Our discussion in the previous section suggests two other factors that could also be cost drivers: scope (variety of goods produced by the firm) and cumulative experience.

When estimating cost functions, economists first gather data from a cross-section of firms or plants at a particular point in time. A cross-section of television producers would consist of a sample of manufacturers or manufacturing facilities in a particular year, such as 2005. For each observation in your cross-section, you would need information about total costs and cost drivers. The set of *cost drivers* that you include in your analysis is usually specific to what you are studying. In television manufacturing, scale, cumulative experience, labor wages, materials prices, and costs of capital would probably be important drivers for explaining the behavior of average costs in the long run.

Having gathered data on total costs and cost drivers, you would then use statistical techniques to construct an estimated total cost function. The most common technique used by economists is multiple regression. The basic idea behind this technique is to find the function that best fits our available data.

CONSTANT ELASTICITY COST FUNCTION

An important issue when you use multiple regression to estimate a cost function is choosing the functional form that relates the dependent variable of interest—in this case, total cost—to the independent variables of interest, such as output and input prices. One commonly used functional form is the **constant elasticity cost function**, which specifies a multiplicative relationship between total cost, output, and input prices. For a production process that involves two inputs, capital and labor, the constant elasticity long-run total cost function is $TC = aQ^b w^c r^d$, where a , b , c , and d are positive constants. It is common to convert this into a linear relationship using logarithms: $\log TC = \log a + b \log Q + c \log w + d \log r$. With the function in this form, the positive constants a , b , c , and d can be estimated using multiple regression.

A useful feature of the constant elasticity specification is that the constant b is the output elasticity of total cost, discussed earlier. Analogously, the constants c and d are the elasticities of long-run total cost with respect to the prices of labor and capital,

8.4 ESTIMATING COST FUNCTIONS

total cost function A mathematical relationship that shows how total costs vary with the factors that influence total costs, including the quantity of output and the prices of inputs.

cost driver A factor that influences or “drives” total or average costs.

constant elasticity cost function A cost function that specifies constant elasticities of total cost with respect to output and input prices.

respectively. These elasticities must be positive since, as we saw earlier, an increase in an input price will increase long-run total cost. We also learned earlier that a given percentage increase in w and r would have to increase long-run total cost by the same percentage amount. This implies that the constants c and d must add up to 1 (i.e., $c + d = 1$) for the estimated long-run total cost function to be consistent with long-run cost minimization. This restriction can be readily incorporated into the multiple regression analysis.

TRANSLOG COST FUNCTION

translog cost function A cost function that postulates a quadratic relationship between the log of total cost and the logs of input prices and output.

The constant elasticity cost function does not allow for the possibility of average costs that first decrease and then increase as Q increases (i.e., economies of scale, followed by diseconomies of scale). The **translog cost function**, which postulates a quadratic relationship between the log of total cost and the logs of input prices and output, does allow for this possibility. The equation of the translog cost function is

$$\begin{aligned} \log TC = & b_0 + b_1 \log Q + b_2 \log w + b_3 \log r + b_4(\log Q)^2 \\ & + b_5(\log w)^2 + b_6(\log r)^2 + b_7(\log w)(\log r) \\ & + b_8(\log w)(\log Q) + b_9(\log r)(\log Q) \end{aligned}$$

This formidable-looking expression turns out to have many useful properties. For one thing, it is often a good approximation of the cost functions that come from just about *any* production function. Thus, if (as is often the case) we don't know the exact functional form of the production function, the translog might be a good choice for the functional form of the cost function. In addition, the average cost function can be U-shaped. Thus, it allows for both economies of scale and diseconomies of scale. For instance, the short-run average cost curves in Figure 8.20 (Application 8.5) were estimated as translog functions. Note, too, that if $b_4 = b_5 = b_6 = b_7 = b_8 = b_9 = 0$, the translog cost function reduces to the constant elasticity cost function. Thus, the constant elasticity cost function is a special case of the translog cost function.

CHAPTER SUMMARY

- The long-run total cost curve shows how the minimized level of total cost varies with the quantity of output. **(LBD Exercise 8.1)**
- An increase in input prices rotates the long-run total cost curve upward through the point $Q = 0$.
- Long-run average cost is the firm's cost per unit of output. It equals total cost divided by output. **(LBD Exercise 8.2)**
- Long-run marginal cost is the rate of change of long-run total cost with respect to output. **(LBD Exercise 8.2)**
- Long-run marginal cost can be less than, greater than, or equal to long-run average cost, depending on whether long-run average cost decreases, increases, or remains constant, respectively, as output increases.
- Economies of scale describe a situation in which long-run average cost decreases as output increases. Economies of scale arise because of the physical properties of processing units, specialization of labor, and indivisibilities of inputs.
- Diseconomies of scale describe a situation in which long-run average cost increases as output increases. A key source of diseconomies of scale is managerial diseconomies.
- The minimum efficient scale (MES) is the smallest quantity at which the long-run average cost curve attains its minimum.
- With economies of scale, there are increasing returns to scale; with diseconomies of scale, there are decreasing

returns to scale; and with neither economies nor diseconomies of scale, there are constant returns to scale.

- The output elasticity of total cost measures the extent of economies of scale; it is the percentage change in total cost per 1 percent change in output.
- The short-run total cost curve tells us the minimized total cost as a function of output, input prices, and the level of the fixed input(s). (LBD Exercise 8.3)
- Short-run total cost is the sum of two components: total variable cost and total fixed cost.
- Short-run total cost is always greater than long-run total cost, except at the quantity of output for which the level of fixed input is cost minimizing.
- Short-run average cost is the sum of average variable cost and average fixed cost. Short-run marginal cost is the rate of change of short-run total cost with respect to output.

• The long-run average cost curve is the lower envelope of the short-run average cost curves. (LBD Exercise 8.4)

- Economies of scope exist when it is less costly to produce given quantities of two products with one firm than it is with two firms that each specialize in the production of a single product.
- Economies of experience exist when average variable cost decreases with cumulative production volume. The experience curve tells us how average variable costs are affected by changes in cumulative production volume. The magnitude of this effect is often expressed in terms of the slope of the experience curve.
- Cost drivers are factors such as output or the prices of inputs that influence the level of costs.
- Two common functional forms that are used for real-world estimation of cost functions are the constant elasticity cost function and the translog cost function.

REVIEW QUESTIONS

1. What is the relationship between the solution to the firm's long-run cost-minimization problem and the long-run total cost curve?
2. Explain why an increase in the price of an input typically causes an increase in the long-run total cost of producing any particular level of output.
3. If the price of labor increases by 20 percent, but all other input prices remain the same, would the long-run total cost at a particular output level go up by more than 20 percent, less than 20 percent, or exactly 20 percent? If the prices of all inputs went up by 20 percent, would long-run total cost go up by more than 20 percent, less than 20 percent, or exactly 20 percent?
4. How would an increase in the price of labor shift the long-run *average* cost curve?
5. a) If the *average* cost curve is increasing, must the marginal cost curve lie above the average cost curve? Why or why not?
b) If the *marginal* cost curve is increasing, must the marginal cost curve lie above the average cost curve? Why or why not?
6. Sketch the long-run marginal cost curve for the "flat-bottomed" long-run average cost curve shown in Figure 8.11.
7. Could the output elasticity of total cost ever be negative?
8. Explain why the short-run marginal cost curve must intersect the average variable cost curve at the minimum point of the average variable cost curve.
9. Suppose the graph of the average variable cost curve is flat. What shape would the short-run marginal cost curve be? What shape would the short-run average cost curve be?
10. Suppose that the minimum level of short-run average cost was the same for every possible plant size. What would that tell you about the shapes of the long-run average and long-run marginal cost curves?
11. What is the difference between economies of scope and economies of scale? Is it possible for a two-product firm to enjoy economies of scope but not economies of scale? Is it possible for a firm to have economies of scale but not economies of scope?
12. What is an experience curve? What is the difference between economies of experience and economies of scale?

PROBLEMS

8.1. The following incomplete table shows a firm's various costs of producing up to 6 units of output. Fill in as much of the table as possible. If you cannot determine the number in a box, explain why it is not possible to do so.

Q	TC	TVC	TFC	AC	MC	AVC
1	100					
2		160				
3			20			
4				95		
5					170	
6						120

8.2. The following incomplete table shows a firm's various costs of producing up to 6 units of output. Fill in as much of the table as possible. If you cannot determine the number in a box, explain why it is not possible to do so.

Q	TC	TVC	AFC	AC	MC	AVC
1				100		
2		50	30			
3					10	
4						30
5						
6	330				80	

8.3. A firm produces a product with labor and capital, and its production function is described by $Q = LK$. The marginal products associated with this production function are $MP_L = K$ and $MP_K = L$. Suppose that the price of labor equals 2 and the price of capital equals 1. Derive the equations for the long-run total cost curve and the long-run average cost curve.

8.4. A firm's long-run total cost curve is $TC(Q) = 1000Q - 30Q^2 + Q^3$. Derive the expression for the corresponding long-run average cost curve and then sketch it. At what quantity is minimum efficient scale?

8.5. A firm's long-run total cost curve is $TC(Q) = 40Q - 10Q^2 + Q^3$, and its long-run marginal cost curve is $MC(Q) = 40 - 20Q + 3Q^2$. Over what range of output does the production function exhibit economies of

scale, and over what range does it exhibit diseconomies of scale?

8.6. For each of the total cost functions, write the expressions for the total fixed cost, average variable cost, and marginal cost (if not given), and draw the average total cost and marginal cost curves.

a) $TC(Q) = 10Q$

b) $TC(Q) = 160 + 10Q$

c) $TC(Q) = 10Q^2$, where $MC(Q) = 20Q$

d) $TC(Q) = 10\sqrt{Q}$, where $MC(Q) = 5/\sqrt{Q}$

e) $TC(Q) = 160 + 10Q^2$, where $MC(Q) = 20Q$

8.7. Consider a production function of two inputs, labor and capital, given by $Q = (\sqrt{L} + \sqrt{K})^2$. The marginal products associated with this production function are as follows:

$$MP_L = [L^{\frac{1}{2}} + K^{\frac{1}{2}}]L^{-\frac{1}{2}}$$

$$MP_K = [L^{\frac{1}{2}} + K^{\frac{1}{2}}]K^{-\frac{1}{2}}$$

Let $w = 2$ and $r = 1$.

a) Suppose the firm is required to produce Q units of output. Show how the cost-minimizing quantity of labor depends on the quantity Q . Show how the cost-minimizing quantity of capital depends on the quantity Q .

b) Find the equation of the firm's long-run total cost curve.

c) Find the equation of the firm's long-run average cost curve.

d) Find the solution to the firm's short-run cost-minimization problem when capital is fixed at a quantity of 9 units (i.e., $\bar{K} = 9$).

e) Find the short-run total cost curve, and graph it along with the long-run total cost curve.

f) Find the associated short-run average cost curve.

8.8. Tricycles must be produced with 3 wheels and 1 frame for each tricycle. Let Q be the number of tricycles, W be the number of wheels, and F be the number of frames. The price of a wheel is P_W and the price of a frame is P_F .

a) What is the long-run total cost function for producing tricycles, $TC(Q, P_W, P_F)$?

b) What is the production function for tricycles, $Q(F, W)$?

8.9. A hat manufacturing firm has the following production function with capital and labor being the inputs: $Q = \min(4L, 7K)$ —i.e., it has a fixed-proportions

production function. If w is the cost of a unit of labor and r is the cost of a unit of capital, derive the firm's long-run total cost curve and average cost curve in terms of the input prices and Q .

8.10. A packaging firm relies on the production function $Q = KL + K$, with $MP_L = K$ and $MP_K = L + 1$. Assume that the firm's optimal input combination is interior (it uses positive amounts of both inputs). Derive its long-run total cost curve in terms of the input prices, w and r . Verify that if the input prices double, then total cost doubles as well.

8.11. A firm has the linear production function $Q = 3L + 5K$, with $MP_L = 3$ and $MP_K = 5$. Derive the expression for the long-run total cost that the firm incurs, as a function of Q and the factor prices, w and r .

8.12. When a firm uses K units of capital and L units of labor, it can produce Q units of output with the production function $Q = K\sqrt{L}$. Each unit of capital costs 20, and each unit of labor costs 25. The level of K is fixed at 5 units.

- Find the equation of the firm's short-run total cost curve.
- On a graph, draw the firm's short-run average cost.

8.13. When a firm uses K units of capital and L units of labor, it can produce Q units of output with the production function $Q = \sqrt{L} + \sqrt{K}$. Each unit of capital costs 2, and each unit of labor costs 1.

- The level of K is fixed at 16 units. Suppose $Q \leq 4$. What will the firm's short-run total cost be? (Hint: How much labor will the firm need?)
- The level of K is fixed at 16 units. Suppose $Q > 4$. Find the equation of the firm's short-run total cost curve.

8.14. Consider a production function of three inputs, labor, capital, and materials, given by $Q = LKM$. The marginal products associated with this production function are as follows: $MP_L = KM$, $MP_K = LM$, and $MP_M = LK$. Let $w = 5$, $r = 1$, and $m = 2$, where m is the price per unit of materials.

- Suppose that the firm is required to produce Q units of output. Show how the cost-minimizing quantity of labor depends on the quantity Q . Show how the cost-minimizing quantity of capital depends on the quantity Q . Show how the cost-minimizing quantity of materials depends on the quantity Q .
- Find the equation of the firm's long-run total cost curve.
- Find the equation of the firm's long-run average cost curve.
- Suppose that the firm is required to produce Q units of output, but that its capital is fixed at a quantity of 50 units (i.e., $\bar{K} = 50$). Show how the cost-minimizing quantity of labor depends on the quantity Q . Show how

the cost-minimizing quantity of materials depends on the quantity Q .

- Find the equation of the short-run total cost curve when capital is fixed at a quantity of 50 units (i.e., $\bar{K} = 50$) and graph it along with the long-run total cost curve.
- Find the equation of the associated short-run average cost curve.

8.15. The production function $Q = KL + M$ has marginal products $MP_K = L$, $MP_L = K$, and $MP_M = 1$. The input prices of K , L , and M are 4, 16, and 1, respectively. The firm is operating in the long run. What is the long-run total cost of producing 400 units of output?

8.16. The production function $Q = KL + M$ has marginal products $MP_K = L$, $MP_L = K$, and $MP_M = 1$. The input prices of K , L , and M are 4, 16, and 1, respectively. The firm is operating in the short run, with K fixed at 20 units. What is the short-run total cost of producing 400 units of output?

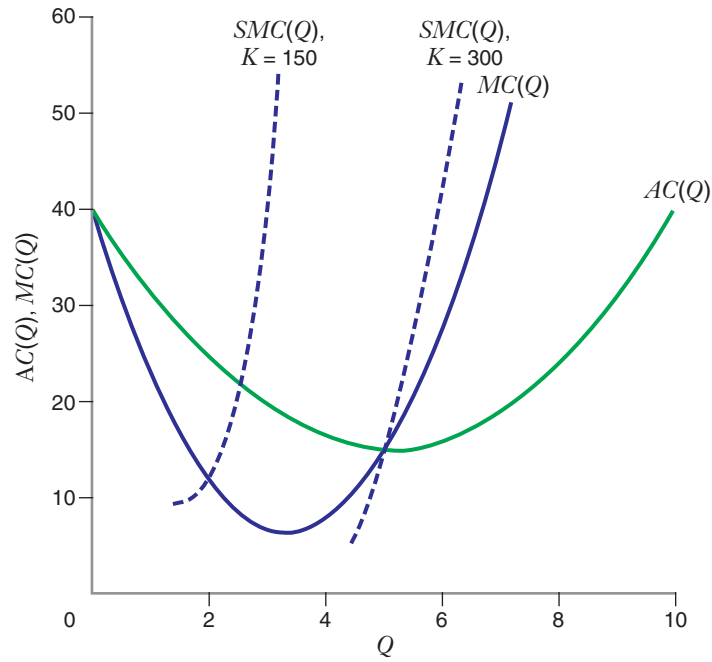
8.17. The production function $Q = KL + M$ has marginal products $MP_K = L$, $MP_L = K$, and $MP_M = 1$. The input prices of K , L , and M are 4, 16, and 1, respectively. The firm is operating in the short run, with K fixed at 20 units and M fixed at 40. What is the short-run total cost of producing 400 units of output?

8.18. A short-run total cost curve is given by the equation $STC(Q) = 1000 + 50Q^2$. Derive expressions for, and then sketch, the corresponding short-run average cost, average variable cost, and average fixed cost curves.

8.19. A producer of hard disk drives has a short-run total cost curve given by $STC(Q) = \bar{K} + Q^2/\bar{K}$. Within the same set of axes, sketch a graph of the short-run average cost curves for three different plant sizes: $\bar{K} = 10$, $\bar{K} = 20$, and $\bar{K} = 30$. Based on this graph, what is the shape of the long-run average cost curve?

8.20. Figure 8.18 shows that the short-run marginal cost curve may lie above the long-run marginal cost curve. Yet, in the long run, the quantities of all inputs are variable, whereas in the short run, the quantities of just some of the inputs are variable. Given that, why isn't short-run marginal cost less than long-run marginal cost for all output levels?

8.21. The following diagram shows the long-run average and marginal cost curves for a firm. It also shows the short-run marginal cost curve for two levels of fixed capital: $K = 150$ and $K = 300$. For each plant size, draw the corresponding short-run average cost curve and explain briefly why that curve should be where you drew it and how it is consistent with the other curves.



8.22. Suppose that the total cost of providing satellite television services is as follows:

$$TC(Q_1, Q_2) = \begin{cases} 0, & \text{if } Q_1 = 0 \text{ and } Q_2 = 0 \\ 1000 + 2Q_1 + 3Q_2, & \text{otherwise} \end{cases}$$

where Q_1 and Q_2 are the number of households that subscribe to a sports and movie channel, respectively. Does the provision of satellite television services exhibit economies of scope?

8.23. A railroad provides passenger and freight service. The table shows the long-run total annual costs $TC(F, P)$, where P measures the volume of passenger traffic and F the volume of freight traffic. For example, $TC(10, 300) = 1000$. Determine whether there are economies of scope for a railroad producing $F = 10$ and $P = 300$. Briefly explain.

Total Annual Costs for Freight and Passenger Service

		<u>P, Units of Passenger Service</u>	
		0	300
F, Units of Freight Service	0	Cost = 0	Cost = 400
	10	Cost = 500	Cost = 1000

8.24. A researcher has claimed to have estimated a long-run total cost function for the production of automobiles. His estimate is that $TC(Q, w, r) = 100w^{-\frac{1}{2}}r^{\frac{1}{2}}Q^3$, where w and r are the prices of labor and capital. Is this a valid cost function—that is, is it consistent with long-run cost minimization by the firm? Why or why not?

8.25. A firm owns two production plants that make widgets. The plants produce identical products and each plant (i) has a production function given by $Q_i = \sqrt{K_i}L_i$, for $i = 1, 2$. The plants differ, however, in the amount of capital equipment in place in the short run. In particular, plant 1 has $K_1 = 25$, whereas plant 2 has $K_2 = 100$. Input prices for K and L are $w = r = 1$.

- Suppose the production manager is told to minimize the short-run total cost of producing Q units of output. While total output Q is exogenous, the manager can choose how much to produce at plant 1 (Q_1) and at plant 2 (Q_2), as long as $Q_1 + Q_2 = Q$. What percentage of its output should be produced at each plant?
- When output is optimally allocated between the two plants, calculate the firm's short-run total, average, and marginal cost curves. What is the marginal cost of the 100th widget? Of the 125th widget? The 200th widget?
- How should the entrepreneur allocate widget production between the two plants in the long run? Find the firm's long-run total, average, and marginal cost curves.

APPENDIX: Shephard's Lemma and Duality

WHAT IS SHEPHARD'S LEMMA?

Let's compare our calculations in Learning-By-Doing Exercises 7.4 and 8.1. Both pertain to the production function $Q = 50\sqrt{KL}$. Our input demand functions were

$$K^*(Q, w, r) = \frac{Q}{50} \sqrt{\frac{w}{r}}$$

$$L^*(Q, w, r) = \frac{Q}{50} \sqrt{\frac{r}{w}}$$

Our long-run total cost function was

$$TC(Q, w, r) = \frac{\sqrt{wr}}{25} Q$$

How does the long-run total cost function vary with respect to the price of labor w , holding Q and r fixed? *The rate of change of long-run total cost with respect to the price of labor is equal to the labor demand function:*

$$\frac{\partial TC(Q, w, r)}{\partial w} = \frac{Q}{50} \sqrt{\frac{r}{w}} = L^*(Q, w, r) \quad (\text{A8.1})$$

Similarly, *the rate of change of long-run total cost with respect to the price of capital is equal to the capital demand function:*

$$\frac{\partial TC(Q, w, r)}{\partial r} = \frac{Q}{50} \sqrt{\frac{w}{r}} = K^*(Q, w, r) \quad (\text{A8.2})$$

The relationships summarized in equations (A8.1) and (A8.2) are no coincidence. They reflect a general relationship between the long-run total cost function and the input demand functions. This relationship is known as **Shephard's Lemma**, which states that *the rate of change of the long-run total cost function with respect to an input price is equal to the corresponding input demand function.*²⁶ Mathematically,

$$\frac{\partial TC(Q, w, r)}{\partial w} = L^*(Q, w, r)$$

$$\frac{\partial TC(Q, w, r)}{\partial r} = K^*(Q, w, r)$$

Shephard's Lemma The relationship between the long-run total cost function and the input demand functions: the rate of change of the long-run total cost function with respect to an input price is equal to the corresponding input demand function.

²⁶Shephard's Lemma also applies to the relationship between short-run total cost functions and the short-run input demand functions. For that reason, we will generally not specify whether we are in the short run or long run in the remainder of this section. However, to maintain a consistent notation, we will use the "long-run" notation used in this chapter and Chapter 7.

Shephard's Lemma makes intuitive sense: If a firm experiences an increase in its wage rate by \$1 per hour, then its total costs should go up (approximately) by the \$1 increase in wages multiplied by the amount of labor it is currently using; that is, the rate of increase in total costs should be approximately equal to its labor demand function. We say "approximately" because if the firm minimizes its total costs, the increase in w should cause the firm to decrease the quantity of labor and increase the quantity of capital it uses. Shephard's Lemma tells us that for small enough changes in w (i.e., Δw sufficiently close to 0), we can use the firm's current usage of labor as a good approximation for how much a firm's costs will rise.

DUALITY

What is the significance of Shephard's Lemma? It provides a key link between the production function and the cost function, a link that in the appendix to Chapter 7 we called *duality*. With respect to Shephard's Lemma, duality works like this:

- Shephard's Lemma tells us that if we know the total cost function, we can derive the input demand functions.
- In turn, as we saw in the appendix to Chapter 7, if we know the input demand functions, we can infer properties of the production function from which it was derived (and maybe even derive the equation of the production function).

Thus, if we know the total cost function, we can always "characterize" the production function from which it must have been derived. In this sense, the cost function is dual (i.e., linked) to the production function. For any production function, there is a unique total cost function that can be derived from it via the cost-minimization problem.

This is a valuable insight. Estimating a firm's production function by statistical methods is often difficult. For one thing, data on input prices and total costs are often more readily available than data on the quantities of inputs. An example of research that took advantage of Shephard's Lemma are the studies of economies of scale in electricity power generation discussed in Application 8.5. In these studies, the researchers estimated cost functions using statistical methods. They then applied Shephard's Lemma and the logic of duality to infer the nature of returns to scale in the production function.

PROOF OF SHEPHARD'S LEMMA

For a fixed Q , let L_0 and K_0 be the cost-minimizing input combination for any arbitrary combination of input prices (w_0, r_0) :

$$L_0 = L^*(Q, w_0, r_0)$$

$$K_0 = K^*(Q, w_0, r_0)$$

Now define a function of w and r , $g(w, r)$:

$$g(w, r) = TC(Q, w, r) - wL_0 - rK_0$$

Since L_0, K_0 is the cost-minimizing input combination when $w = w_0$ and $r = r_0$, it must be the case that

$$g(w_0, r_0) = 0 \tag{A8.3}$$

Moreover, since (L_0, K_0) is a feasible (but possibly nonoptimal) input combination to produce output Q at other input prices (w, r) besides (w_0, r_0) , it must be the case that:

$$g(w, r) \leq 0 \text{ for } (w, r) \neq (w_0, r_0) \quad (\text{A8.4})$$

Conditions (A8.3) and (A8.4) imply that the function $g(w, r)$ attains its maximum when $w = w_0$ and $r = r_0$. Hence, at these points, its partial derivatives with respect to w and r must be zero:²⁷

$$\frac{\partial g(w_0, r_0)}{\partial w} = 0 \Rightarrow \frac{\partial TC(Q, w_0, r_0)}{\partial w} = L_0 \quad (\text{A8.5})$$

$$\frac{\partial g(w_0, r_0)}{\partial r} = 0 \Rightarrow \frac{\partial TC(Q, w_0, r_0)}{\partial r} = K_0 \quad (\text{A8.6})$$

But since $L_0 = L^*(Q, w_0, r_0)$ and $K_0 = K^*(Q, w_0, r_0)$, (A8.5) and (A8.6) imply

$$\frac{\partial TC(Q, w_0, r_0)}{\partial w} = L^*(Q, w_0, r_0) \quad (\text{A8.7})$$

$$\frac{\partial TC(Q, w_0, r_0)}{\partial r} = K^*(Q, w_0, r_0) \quad (\text{A8.8})$$

Since (w_0, r_0) is an arbitrary combination of input prices, conditions (A8.7) and (A8.8) hold for any pair of input prices, and this is exactly what we wanted to show to prove Shephard's Lemma.

²⁷For more on the use of partial derivatives to find the optimum of a function depending on more than one variable, see the Mathematical Appendix in this book.