7 CHAPTER

Costs

An economist is a person who, when invited to give a talk at a banquet, tells the audience there's no such thing as a free lunch.

A semiconductor manufacturer can produce a chip using many pieces of equipment and relatively few workers' labor or many workers and relatively few machines. How does the firm make its choice?

The firm uses a two-step procedure in determining how to produce a certain amount of output efficiently. It first determines which production processes are *technologically efficient* so that it can produce the desired level of output with the least amount of inputs. As we saw in Chapter 6, the firm uses engineering and other information to determine its production function, which summarizes the many technologically efficient production processes available.

The firm's second step is to pick from these technologically efficient production processes the one that is also **economically efficient**, minimizing the cost of producing a specified amount of output. To determine which process minimizes its cost of production, the firm uses information about the production function and the cost of inputs.

By reducing its cost of producing a given level of output, a firm can increase its profit. Any profit-maximizing competitive, monopolistic, or oligopolistic firm minimizes its cost of production.

In this chapter, we examine five main topics

1. **Measuring costs:** Economists count both explicit costs and implicit (opportunity) costs.
2. **Short-run costs:** To minimize its costs in the short run, a firm adjusts its variable factors (such as labor), but it cannot adjust its fixed factors (such as capital).
3. **Long-run costs:** In the long run, a firm adjusts all its inputs because usually all inputs are variable.
4. **Lower costs in the long run:** Long-run cost is as low as or lower than short-run cost because the firm has more flexibility in the long run, technological progress occurs, and workers and managers learn from experience.
5. **Cost of producing multiple goods:** If the firm produces several goods simultaneously, the cost of each may depend on the quantity of all the goods produced.

Businesspeople and economists need to understand the relationship between costs of inputs and production to determine the least costly way to produce. Economists have an additional reason for wanting to know about costs. As we'll see in later chapters, the relationship between output and costs plays an important role in determining the nature of a market—how many firms are in the market and how high price is relative to cost.

7.1 MEASURING COSTS

How much would it cost you to stand at the wrong end of a shooting gallery?

—S. J. Perelman

To show how a firm's cost varies with its output, we first have to measure costs. Businesspeople and economists often measure costs differently.

Economic Cost

Economists include all relevant costs. To run a firm profitably, a manager acts like an economist and considers all relevant costs. However, this same manager may direct the firm's accountant or bookkeeper to measure cost in ways that are consistent with tax laws and other laws to make the firm's financial statement look good to stockholders or to minimize the firm's taxes this year.

Economists consider both explicit costs and implicit costs. *Explicit costs* are a firm's direct, out-of-pocket payments for inputs to its production process during a given time period such as a year. These costs include production workers' wages, managers' salaries, and payments for materials. However, firms use inputs that may not have an explicit price. These *implicit costs* include the value of the working time of the firm's owner and the value of other resources used but not purchased in a given period.

The **economic cost** or **opportunity cost** is the value of the best alternative use of a resource. The economic or opportunity cost includes both explicit and implicit costs. If a firm purchases and uses an input immediately, that input's opportunity cost is the amount the firm pays for it. If the firm uses an input from its inventory, its opportunity cost is not necessarily the price it paid for the input years ago. Rather, the opportunity cost is what it could buy or sell that input for today.

The classic example of an implicit opportunity cost is captured in the phrase "There's no such thing as a free lunch." Suppose that your parents offer to take you to lunch tomorrow. You know that they'll pay for the meal, but you also know that this lunch is not free for you. Your opportunity cost for the lunch is the best alternative use of your time. Presumably, the best alternative use of your time is studying this textbook, but other possible alternatives include what you could earn at a job or the value you place on watching TV. Often such opportunity costs are a substantial portion of total costs.

If you start your own firm, you should be very concerned about opportunity costs. Suppose that your explicit cost is $40,000, including the rent for your work space, the cost of materials, and the wage payments to your employees. Because you do not pay yourself a salary—instead, you keep any profit at the end of the year—the explicit cost does not include the value of your time. According to an economist, your firm's full economic cost is the sum of the explicit cost plus the opportunity value of your time. If the highest wage you could have earned working for some other firm is $25,000, your full economic cost is $65,000.

In deciding whether to continue running your firm or to work for someone else, you must consider both explicit and opportunity costs. If your annual revenue is $60,000, after you pay your explicit cost of $40,000, you keep $20,000 at the end of the year. The opportunity cost of your time, $25,000, exceeds $20,000, so you can earn more working for someone else. (What are you giving up to study opportunity costs?)

Application

OPPORTUNITY COST OF WAITING TIME

Canadian taxes pay for public health care. However, taxes do not cover the full cost of medical care. To contain costs, health care is rationed in part by having patients wait for treatment. People who are forced to wait are less likely to request treatment, some diseases clear up on their own during the wait, and some patients die while waiting. Are these additional waiting-time opportunity costs large?

Most Canadian patients remain on a waiting list less than two months. In one province, two-thirds of all inpatients and three-quarters of all outpatients reported waiting less than eight weeks for elective surgery. But 20% of inpatients and 14% of outpatients waited more than 12 weeks.

Many patients suffer or cannot work while waiting for treatment. Doctors estimate that 41% of all patients and 88% of cardiology patients have difficulty carrying on their work or daily duties as a result of their medical conditions.

As a proxy for the opportunity cost of waiting time for the 41% of people who experience difficulty while waiting, Globerman (1991) used average earnings. (Such a cost measure is an underestimate because it ignores pain and suffering.) He calculated that waiting-time cost was approximately $132 million in British Columbia, which amounted to 0.2% of that province's gross domestic product and about 8% of its total health cost. Bishai and Lang (2000) estimate the value people place on a one-month reduction in waiting time for cataract surgery at $128 per patient in Canada, $160 in Denmark, and $243 in Barcelona.

Capital Costs

Determining the opportunity cost of capital, such as land or equipment, requires special considerations. Capital is a **durable good**: a product that is usable for years. Two problems may arise in measuring the cost of capital. The first is how to allocate the initial purchase cost over time. The second is what to do if the value of the capital changes over time.

Allocating Capital Costs over Time. Capital may be rented or purchased. For example, a firm may rent a truck for $200 a month or buy it outright for $18,000.

If the firm rents the truck, the rental payment is the relevant opportunity cost. By using the rental rate, we avoid the two measurement problems. The truck is rented period by period, so the firm does not have to worry about how to allocate the purchase cost of a truck over time. Moreover, the rental rate adjusts if the cost of a new truck changes over time.

Suppose, however, that the firm buys the truck. The firm's bookkeeper may *expense* the cost by recording the full $18,000 when it's made or may *amortize* the cost by spreading the $18,000 over the life of the truck according to an arbitrary rule set by the relevant government authority, such as the Internal Revenue Service (IRS). If the IRS approves of several approaches to amortizing expenses, a bookkeeper or an accountant may use whichever arbitrary rule minimizes the firm's taxes.

An economist amortizes the cost of the truck on the basis of its opportunity cost at each moment of time, which is the amount that the firm could charge others to rent the truck. That is, regardless of whether the firm buys or rents the truck, an economist views the opportunity cost of this capital good as a rent per time period: the amount the firm will receive if it rents its truck to others at the going rental rate.[1] If the value of an older truck is less than that of a newer one, the rental rate for the truck falls over time.

Actual and Historical Costs. Not only may the rental rate for a piece of capital fall over time as the capital ages, but it may also change because of shifts in supply and demand in the market for capital goods or for other reasons. A piece of capital may be worth much more or much less today than it was when it was purchased.

To maximize its profit, a firm must properly measure the cost of a piece of capital—its current opportunity cost of the capital good—and not what the firm paid for it—its historical cost. Suppose that a firm paid $30,000 for a piece of land that it can resell for only $20,000. Also suppose that it uses the land itself and the current value of the land to the firm is only $19,000. Should the firm use the land or sell it? As any child can tell the firm, there's no point in crying over spilt milk. The firm should ignore how much it paid for the land in making its decision. As the value of the land to the firm, $19,000, is less than the opportunity cost of the land, $20,000, the firm can make more by selling the land.

The firm's current opportunity cost of capital may be less than what it paid if the firm cannot resell the capital. A firm that bought a specialized piece of equipment with no alternative use cannot resell the equipment. Because the equipment has no alternative use, the historical cost of buying that capital is a **sunk cost**: an expenditure that cannot be recovered. Because this equipment has no alternative use, the current or opportunity cost of the capital is zero. In short, when determining the rental value of capital, economists use the opportunity value and ignore the historical price.

SWARTHMORE COLLEGE'S COST OF CAPITAL

Many nonprofit institutions such as universities and governmental agencies are notorious for ignoring the implicit cost of their capital. When setting tuition and making other plans, Swarthmore College in Pennsylvania estimates

[1]If trucks cannot be rented, an economist calculates an implicit rental rate for trucks taking account of both explicit and opportunity costs. If the firm could sell the truck for $5,000, the opportunity cost of keeping it is the interest that could be earned on $5,000 (Chapter 16). In addition, the firm incurs direct maintenance costs and the opportunity cost due to *depreciation*: the drop in value from wear and tear.

its annual cost at $40,000 per student, based on the cost of salaries, academic and general institutional support, food, maintenance and additions to the physical plant, and other annual expenses such as student aid. This cost calculation is a gross underestimate, however, because it ignores the opportunity cost of the campus—the amount the college could earn by renting out its land and buildings. Including that opportunity cost of its land and buildings raises its true economic cost to about $50,000 annually per student.

7.2 SHORT-RUN COSTS

To make profit-maximizing decisions, a firm needs to know how its cost varies with output. A firm's cost rises as it increases its output. A firm cannot vary some of its inputs, such as capital, in the short run (Chapter 6). As a result, it is usually more costly for a firm to increase output in the short run than in the long run, when all inputs can be varied. In this section, we look at the cost of increasing output in the short run.

Short-Run Cost Measures

We start by using a numerical example to illustrate the basic cost concepts. We then examine the graphic relationship between these concepts.

Table 7.1 Variation of Short-Run Cost with Output

Output, q	Fixed Cost, F	Variable Cost, VC	Total Cost, C	Marginal Cost, MC	Average Fixed Cost, $AFC = F/q$	Average Variable Cost, $AVC = VC/q$	Average Cost, $AC = C/q$
0	48	0	48				
1	48	25	73	25	48	25	73
2	48	46	94	21	24	23	47
3	48	66	114	20	16	22	38
4	48	82	130	16	12	20.5	32.5
5	48	100	148	18	9.6	20	29.6
6	48	120	168	**20**	8	**20**	28
7	48	141	189	21	6.9	20.1	27
8	48	168	216	**27**	6	21	**27**
9	48	198	246	30	5.3	22	27.3
10	48	230	278	32	4.8	23	27.8
11	48	272	320	42	4.4	24.7	29.1
12	48	321	369	49	4.0	26.8	30.8

Cost Levels. To produce a given level of output in the short run, a firm incurs costs for both its fixed and variable inputs. A firm's **fixed cost** (F) is its production expense that does not vary with output. The fixed cost includes the cost of inputs that the firm cannot practically adjust in the short run, such as land, a plant, large machines, and other capital goods. The fixed cost for a capital good a firm owns and uses is the opportunity cost of not renting it to someone else. The fixed cost is \$48 per day for the firm in Table 7.1.

A firm's **variable cost** (VC) is the production expense that changes with the quantity of output produced. The variable cost is the cost of the variable inputs—the inputs the firm can adjust to alter its output level, such as labor and materials. Table 7.1 shows that the firm's variable cost changes with output. Variable cost goes from \$25 a day when 1 unit is produced to \$46 a day when 2 units are produced.

A firm's **cost** (or **total cost**, C) is the sum of a firm's variable cost and fixed cost:

$$C = VC + F.$$

The firm's total cost of producing 2 units of output per day is \$94 per day, which is the sum of the fixed cost, \$48, and the variable cost, \$46. Because variable cost changes with the level of output, total cost also varies with the level of output, as the table illustrates.

To decide how much to produce, a firm uses several measures of how its cost varies with the level of output. Table 7.1 shows four such measures that we derive using the fixed cost, the variable cost, and the total cost.

Marginal Cost. A firm's **marginal cost** (MC) is the amount by which a firm's cost changes if the firm produces one more unit of output. The marginal cost is[2]

$$MC = \frac{\Delta C}{\Delta q},$$

where ΔC is the change in cost when output changes by Δq. Table 7.1 shows that, if the firm increases its output from 2 to 3 units, $\Delta q = 1$, its total cost rises from \$94 to \$114, $\Delta C = \$20$, so its marginal cost is $\$20 = \Delta C/\Delta q$.

Because only variable cost changes with output, we can also define marginal cost as the change in variable cost from a one-unit increase in output:

$$MC = \frac{\Delta VC}{\Delta q}.$$

As the firm increases output from 2 to 3 units, its variable cost increases by $\Delta VC = \$20 = \$66 - \$46$, so its marginal cost is $MC = \Delta VC/\Delta q = \$20$. A firm uses marginal cost in deciding whether it pays to change its output level.

Average Costs. Firms use three average cost measures. The **average fixed cost** (AFC) is the fixed cost divided by the units of output produced: $AFC = F/q$. The average

---

[2]If we use calculus, the marginal cost is $MC = dC(q)/dq$, where $C(q)$ is the cost function that shows how cost varies with output. The calculus definition says how cost changes for an infinitesimal change in output. To illustrate the idea, however, we use larger changes in the table.

fixed cost falls as output rises because the fixed cost is spread over more units. The average fixed cost falls from \$48 for 1 unit of output to \$4 for 12 units of output in Table 7.1.

The **average variable cost** (*AVC*) is the variable cost divided by the units of output produced: $AVC = VC/q$. Because the variable cost increases with output, the average variable cost may either increase or decrease as output rises. The average variable cost is \$25 at 1 unit, falls until it reaches a minimum of \$20 at 6 units, and then rises. As we show in Chapter 8, a firm uses the average variable cost to determine whether to shut down operations when demand is low.

The **average cost** (*AC*)—or *average total cost*—is the total cost divided by the units of output produced: $AC = C/q$. The average cost is the sum of the average fixed cost and the average variable cost:[3]

$$AC = AFC + AVC.$$

In Table 7.1, as output increases, average cost falls until output is 8 units and then rises. The firm makes a profit if its average cost is below its price, which is the firm's average revenue.

LOWERING TRANSACTION COSTS FOR USED GOODS AT EBAY AND ABEBOOKS

In the last century, department stores and supermarkets largely replaced smaller specialty stores, as consumers found it more efficient to go to one store than to many. Consumers incur a transaction or search cost to shop, primarily the opportunity cost of their time. This transaction cost consists of a fixed cost of traveling to and from the store and a variable cost that rises with the number of different types of items the consumer tries to find on the shelves. By going to a supermarket that carries meat, fruits and vegetables, and other items, consumers can avoid some of the fixed transaction costs of traveling to a separate butcher shop, produce mart, and so forth.

Until recently, if you wanted a collectable lunch box or snow globe, you had to go to many yard sales, subscribe to collectors' newsletters, and otherwise hope for the best. For people like you, the Internet has provided nothing short of a garage-sale nirvana. The Internet lowers transactions costs for both buyers and sellers. At a negligible fixed cost (logging onto a Web site), you can buy or sell used books, strange collectibles, paintings, . . . virtually anything but vital organs at **www.eBay.com**. One reason why eBay's auction site competitors have been unsuccessful in staying in this market is that it lowers buyers' transaction costs to have only one site they need to check. (Of course, offsetting that benefit is a possibly higher price due to a lack of competition.)

The Internet has also invigorated the used-book business. As recently as the mid-1990s, it could be extremely difficult to find a specific out-of-print book.

[3]Because $C = VC + F$, if we divide both sides of the equation by q, we obtain

$$AC = C/q = F/q + VC/q = AFC + AVC.$$

Virtually the only way was to scour the local used-book stores or scan the 30-plus pages of A B Bookman, a weekly magazine that publishes lists of titles.

Keith Waters, a computer consultant, learned about this problem from his wife, a used-book dealer. He created Advanced Book Exchange, **www.abebooks.com**, which now links 10,000 used- and rare-book dealers, who pay a monthly subscription to join. These dealers can use the site to sell and buy. Final customers can also use it to buy. As of 2002, Abebooks listed more than 38 million books, with 50,000 to 100,000 added daily. It competes with **www.amazon.com** (which bought Abebooks' chief rival **biblofind.com**), **www.half.com** (owned by eBay), and other used book sellers on the Web.

Dan Adams, who runs Waverly Books out of his home in Santa Monica, California, lists 20,000 titles on Abebooks. He now sells almost exclusively over the Internet to customers around the world.

Even the Strand, a gigantic Manhattan book store with 16 miles of books and sales in excess of $20 million, introduced an Internet web site, **www.strandbooks.com**, which links with Abebooks, Amazon, and other virtual outlets. After only a year, its 2002 Internet sales were 15% of its total, its sales of used books rose more than 15%, and its sales of rare books grew by 34%.

How much does lowering transactions costs matter? By one estimate, sales of used books are 100 times what they were in 1995.

Short-Run Cost Curves

We illustrate the relationship between output and the various cost measures using curves in Figure 7.1. Panel a shows the variable cost, fixed cost, and total cost curves that correspond to Table 7.1. The fixed cost, which does not vary with output, is a horizontal line at $48. The variable cost curve is zero at zero units of output and rises with output. The total cost curve, which is the vertical sum of the variable cost curve and the fixed cost line, is $48 higher than the variable cost curve at every output level, so the variable cost and total cost curves are parallel.

Panel b shows the average fixed cost, average variable cost, average cost, and marginal cost curves. The average fixed cost curve falls as output increases. It approaches zero as output gets large because the fixed cost is spread over many units of output. The average cost curve is the vertical sum of the average fixed cost and average variable cost curves. For example, at 6 units of output, the average variable cost is 20 and the average fixed cost is 8, so the average cost is 28.

The relationships between the average and marginal curves to the total curves are similar to those between the total product, marginal product, and average product curves, which we discussed in Chapter 6. The average cost at a particular output level is the slope of a line from the origin to the corresponding point on the cost curve. The slope of that line is the rise—the cost at that output level—divided by the run—the output level—which is the definition of the average cost. In panel a, the slope of the line from the origin to point *A* is the average cost for 8 units of output. The height of the cost curve at *A* is 216, so the slope is 216/8 = 27, which is the height of the average cost curve at the corresponding point *a* in panel b.

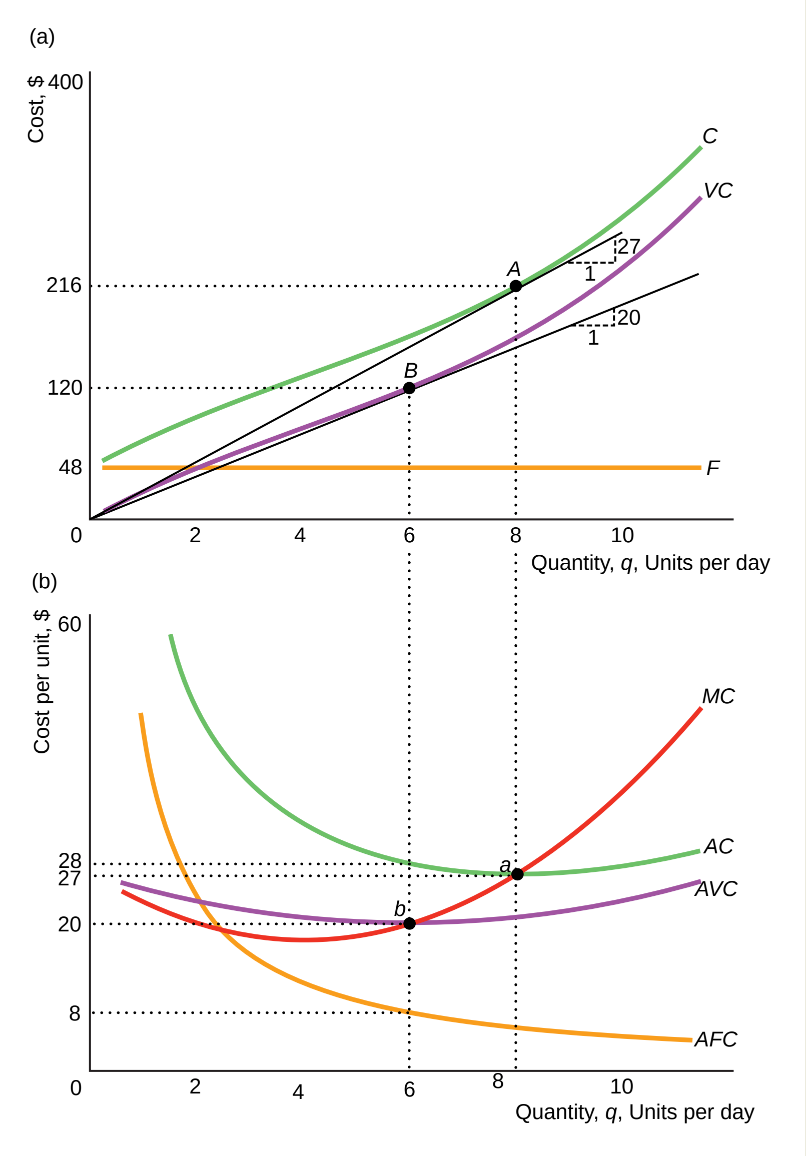


Figure 7.1 Short-Run Cost Curves. (a) Because the total cost differs from the variable cost by the fixed cost, *F*, of $48, the total cost curve, *C*, is parallel to the variable cost curve, *VC*. (b) The marginal cost curve, *MC*, cuts the average variable cost, *AVC*, and average cost, *AC*, curves at their minimums. The height of the *AC* curve at point *a* equals the slope of the line from the origin to the cost curve at *A*. The height of the *AVC* at *b* equals the slope of the line from the origin to the variable cost curve at *B*. The height of the marginal cost is the slope of either the *C* or *VC* curve at that quantity.

Similarly, the average variable cost is the slope of a line from the origin to a point on the variable cost curve. The slope of the dashed line from the origin to *B* in panel a is 20—the height of the variable cost curve, 120, divided by the number of units of output, 6—which is the height of the average variable cost at 6 units of output, point *b* in panel b.

The marginal cost is the slope of either the cost curve or the variable cost curve at a given output level. As the cost and variable cost curves are parallel, they have the same slope at any given output. The difference between cost and variable cost is fixed cost, which does not affect marginal cost.

The dashed line from the origin is tangent to the cost curve at *A* in panel a. Thus the slope of the dashed line equals both the average cost and the marginal cost at 8 units of output. This equality occurs at the corresponding point *a* in panel b, where the marginal cost curve intersects the average cost. (See Appendix 7A for a mathematical proof.)

Where the marginal cost curve is below the average cost, the average cost curve declines with output. Because the average cost of 47 for 2 units is greater than the marginal cost of the third unit, 20, the average cost for 3 units falls to 38. Where the marginal cost is above the average cost, the average cost curve rises with output. At 8 units, the marginal cost equals the average cost, so the average is unchanging, which is the minimum point, *a*, of the average cost curve.

We can show the same results using the graph. Because the dashed line from the origin is tangent to the variable cost curve at *B* in panel a, the marginal cost equals the average variable cost at the corresponding point *b* in panel b. Again, where marginal cost is above average variable cost, the average variable cost curve rises with output; and where marginal cost is below average variable cost, the average variable cost curve falls with output. Because the average cost curve is everywhere above the average variable cost curve and the marginal cost curve is rising where it crosses both average curves, the minimum of the average variable cost curve, *b*, is at a lower output level than the minimum of the average cost curve, *a*.

Production Functions and the Shape of Cost Curves

The production function determines the shape of a firm's cost curves. The production function shows the amount of inputs needed to produce a given level of output. The firm calculates its cost by multiplying the quantity of each input by its price and summing.

If a firm produces output using capital and labor and its capital is fixed in the short run, the firm's variable cost is its cost of labor. Its labor cost is the wage per hour, w, times the number of hours of labor, L, employed by the firm: $VC = wL$.

In the short run, when the firm's capital is fixed, the only way the firm can increase its output is to use more labor. If the firm increases its labor enough, it reaches the point of *diminishing marginal return to labor*, at which each extra worker increases output by a smaller amount. We can use this information about the relationship between labor and output—the production function—to determine the shape of the variable cost curve and its related curves.

Shape of the Variable Cost Curve. If input prices are constant, the production function determines the shape of the variable cost curve. We illustrate this relationship for the

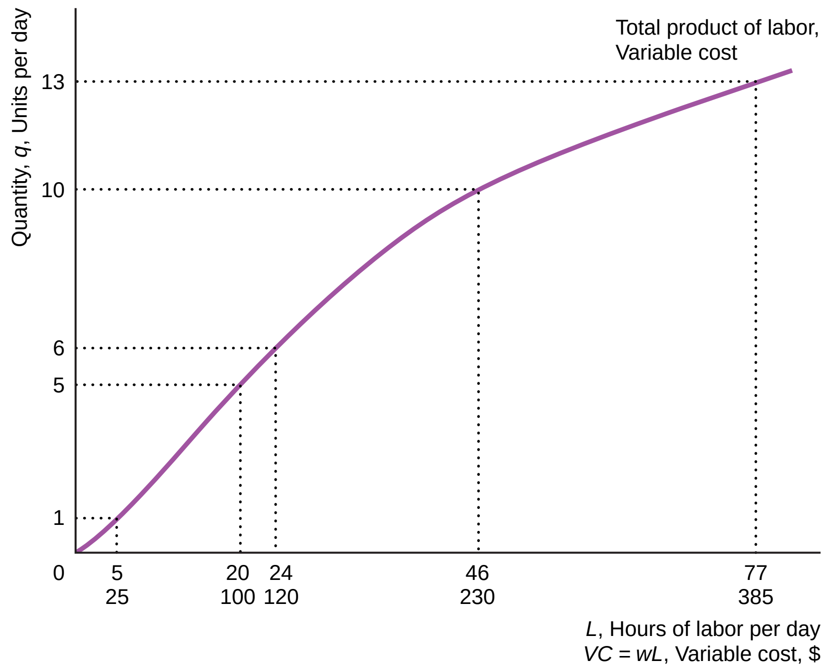


Figure 7.2 Variable Cost and Total Product of Labor. The firm's short-run variable cost curve and its total product of labor curve have the same shape. The total product of labor curve uses the horizontal axis measuring hours of work. The variable cost curve uses the horizontal axis measuring labor cost, which is the only variable cost.

firm in Figure 7.2. The firm faces a constant input price for labor, the wage, of $5 per hour.

The total product of labor curve in Figure 7.2 shows the firm's short-run production function relationship between output and labor when capital is held fixed. For example, it takes 24 hours of labor to produce 6 units of output. Nearly doubling labor to 46 hours causes output to increase by only two-thirds to 10 units of output. As labor increases, the total product of labor curve increases less than in proportion. This flattening of the total product of labor curve at higher levels of labor reflects the diminishing marginal return to labor.

This curve shows both the production relation of output to labor and the variable cost relation of output to cost. Because each hour of work costs the firm $5, we can relabel the horizontal axis in Figure 7.2 to show the firm's variable cost, which is its cost of labor. To produce 6 units of output takes 24 hours of labor, so the firm's variable cost is $120. By using the variable cost labels on the horizontal axis, the total product of labor curve becomes the variable cost curve, where each worker costs the firm $120 per day in wages. The variable cost curve in Figure 7.2 is the same as the one in panel a of Figure 7.1, in which the output and cost axes are reversed. For example, the variable cost of producing 6 units is \$120 in both figures.

Diminishing marginal returns in the production function cause the variable cost to rise more than in proportion as output increases. Because the production function determines the shape of the variable cost curve, it also determines the shape of the marginal, average variable, and average cost curves. We now examine the shape of each of these cost curves in detail because in making decisions, firms rely more on these per-unit cost measures than on total variable cost.

Shape of the Marginal Cost Curve. The marginal cost is the change in variable cost as output increases by one unit: $MC = \Delta VC/\Delta q$. In the short run, capital is fixed, so the only way the firm can produce more output is to use extra labor. The extra labor required to produce one more unit of output is $\Delta L/\Delta q$. The extra labor costs the firm w per unit, so the firm's cost rises by $w(\Delta L/\Delta q)$. As a result, the firm's marginal cost is

$$MC = \frac{\Delta VC}{\Delta q} = w\frac{\Delta L}{\Delta q}.$$

The marginal cost equals the wage times the extra labor necessary to produce one more unit of output. To increase output by one unit from 5 to 6 units takes four extra workers in Figure 7.2. If the wage is \$5 per hour, the marginal cost is \$20.

How do we know how much extra labor we need to produce one more unit of output? That information comes from the production function. The marginal product of labor—the amount of extra output produced by another unit of labor, holding other inputs fixed—is $MP_L = \Delta q/\Delta L$. Thus the extra labor we need to produce one more unit of output, $\Delta L/\Delta q$, is $1/MP_L$, so the firm's marginal cost is

$$MC = \frac{w}{MP_L}. \tag{7.1}$$

Equation 7.1 says that the marginal cost equals the wage divided by the marginal product of labor. If the firm is producing 5 units of output, it takes four extra hours of labor to produce one more unit of output in Figure 7.2, so the marginal product of an hour of labor is $\frac{1}{4}$. Given a wage of \$5 an hour, the marginal cost of the sixth unit is \$5 divided by $\frac{1}{4}$, or \$20, as panel b of Figure 7.1 shows.

Equation 7.1 shows that the marginal cost moves in the direction opposite that of the marginal product of labor. At low levels of labor, the marginal product of labor commonly rises with additional labor because extra workers help the original workers and they can collectively make better use of the firm's equipment (Chapter 6). As the marginal product of labor rises, the marginal cost falls.

Eventually, however, as the number of workers increases, workers must share the fixed amount of equipment and may get in each other's way, so the marginal cost curve slopes upward because of diminishing marginal returns to labor. Thus the marginal cost first falls and then rises, as panel b of Figure 7.1 illustrates.

Shape of the Average Cost Curves. Diminishing marginal returns to labor, by determining the shape of the variable cost curve, also determine the shape of the average variable cost curve. The average variable cost is the variable cost divided by output:

$AVC = VC/q$. For the firm we've been examining, whose only variable input is labor, variable cost is wL, so average variable cost is

$$AVC = \frac{VC}{q} = \frac{wL}{q}.$$

Because the average product of labor is q/L, average variable cost is the wage divided by the average product of labor:

$$AVC = \frac{w}{AP_L}. \tag{7.2}$$

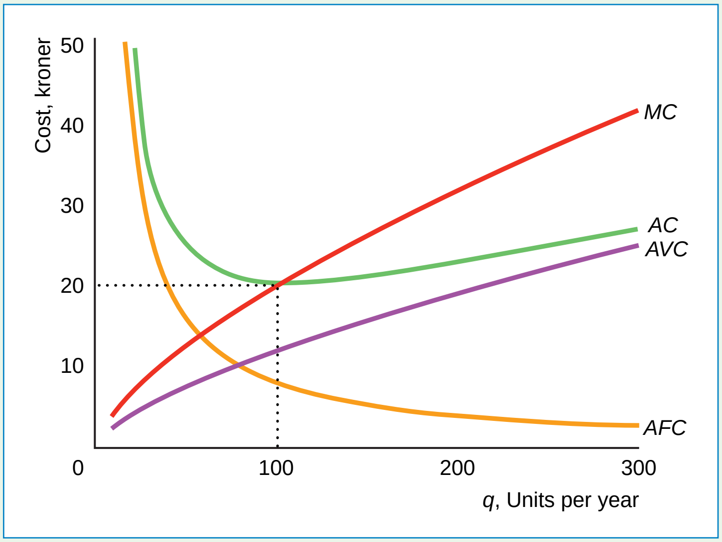
In Figure 7.2, at 6 units of output, the average product of labor is $\frac{1}{4}$ (= q/L = 6/24), so the average variable cost is \$20, which is the wage, \$5, divided by the average product of labor, $\frac{1}{4}$.

With a constant wage, the average variable cost moves in the opposite direction of the average product of labor in Equation 7.2. As we discussed in Chapter 6, the average product of labor tends to rise and then fall, so the average cost tends to fall and then rise, as in panel b of Figure 7.1.

The average cost curve is the vertical sum of the average variable cost curve and the average fixed cost curve, as in panel b. If the average variable cost curve is U-shaped, adding the strictly falling average fixed cost makes the average cost fall more steeply than the average variable cost curve at low output levels. At high output levels, the average cost and average variable cost curves differ by ever smaller amounts, as the average fixed cost, F/q, approaches zero. Thus the average cost curve is also U-shaped.

SHORT-RUN COST CURVES FOR A PRINTING FIRM

The short-run average cost curve for the Norwegian printing firm (in Chapter 6) is U-shaped, even though its average variable cost is strictly upward sloping. The graph (based on the estimates of Griliches and Ringstad, 1971) shows the



firm's various short-run cost curves, where the firm's capital is fixed at $\overline{K} = 100$. Appendix 7B derives the firm's short-run cost curves mathematically.

The firm's average fixed cost (AFC) falls as output increases. The firm's average variable cost curve is strictly increasing. The average cost (AC) curve is the vertical sum of the average variable cost (AVC) and average fixed cost curves. Because the average fixed cost curve falls with output and the average variable cost curve rises with output, the average cost curve is U-shaped. The firm's marginal cost (MC) lies above the rising average variable cost curve for all positive quantities of output and cuts the average cost curve at its minimum.

Effects of Taxes on Costs

Taxes applied to a firm shift some or all of the marginal and average cost curves. For example, suppose that the government collects a specific tax of \$10 per unit of output from the firm. This tax, which varies with output, affects the firm's variable cost but not its fixed cost. As a result, it affects the firm's average cost, average variable cost, and marginal cost curves but not its average fixed cost curve.

At every quantity, the average variable cost and the average cost rise by the full amount of the tax. The second column of Table 7.2 shows the firm's average variable cost before the tax, AVC^b. For example, if it sells 6 units of output, its average variable cost is \$20. After the tax, the firm must pay the government \$10 per unit, so the firm's after-tax average variable cost rises to \$30. More generally, the firm's after-tax average variable cost, AVC^a, is its average variable cost of production—the before-tax average variable cost—plus the tax per unit, \$10: $AVC^a = AVC^b + \$10$.

The average cost equals the average variable cost plus the average fixed cost. Because the tax increases average variable cost by \$10 and does not affect the average fixed cost, the tax increases average cost by \$10.

The tax also increases the firm's marginal cost. Suppose that the firm wants to increase output from 7 to 8 units. The firm's actual cost of producing the third unit—

Table 7.2 Effect of a Specific Tax of \$10 per Unit on Short-Run Costs

Q	AVC^b	$AVC^a = AVC^b + \$10$	$AC^b = C/q$	$AC^a = C/q + \$10$	MC^b	$MC^a = MC^b + \$10$
1	25	35	73	83	25	35
2	23	33	47	57	21	31
3	22	32	38	48	20	30
4	20.5	30.5	32.5	42.5	16	26
5	20	30	29.6	39.6	18	28
6	**20**	**30**	28	38	**20**	**30**
7	20.1	30.1	27	37	21	31
8	21	31	**27**	**37**	**27**	**37**
9	22	32	27.3	37.3	30	40
10	23	33	27.8	37.8	32	42
11	24.7	34.7	29.1	39.1	42	52
12	26.8	36.8	30.8	40.8	49	59

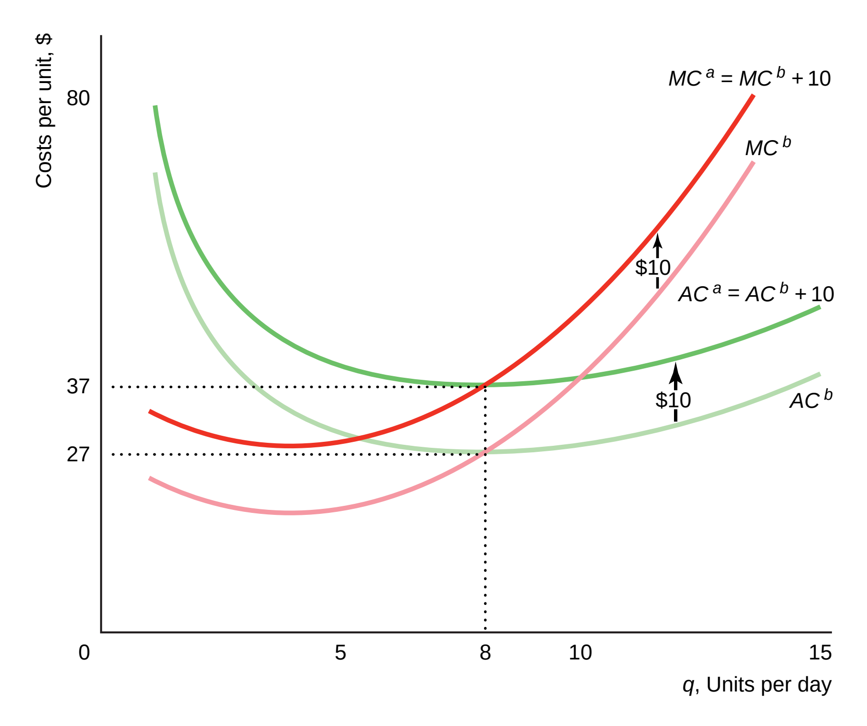


Figure 7.3 Effect of a Specific Tax on Cost Curves. A specific tax of $10 per unit shifts both the marginal cost and average cost curves upward by $10. Because of the parallel upward shift of the average cost curve, the minimum of both the before-tax average cost curve, AC^b, and the after-tax average cost curve, AC^a, occurs at the same output, 8 units.

its before-tax marginal cost, MC^b—is $27. To produce an extra unit of output, the cost to the firm is the marginal cost of producing the extra unit plus $10, so its after-tax marginal cost is $MC^a = MC^b + \$10$. In particular, its after-tax marginal cost of producing the eighth unit is $37.

A specific tax shifts the marginal cost and the average cost curves upward in Figure 7.3 by the amount of the tax, $10 per unit. The after-tax marginal cost intersects the after-tax average cost at its minimum. Because both the marginal and average cost curves shift upward by exactly the same amount, the after-tax average cost curve reaches its minimum at the same level of output, 8 units, as the before-tax average cost, as both panel a and Table 7.2 show. At 8 units, the minimum of the before-tax average cost curve is $27 and that of the after-tax average cost curve is $37. So even though a specific tax increases a firm's average cost, it does not affect the output at which average cost is minimized.

Similarly, we can analyze the effect of a franchise tax on costs. A *franchise tax*—also called a *business license fee*—is a lump sum that a firm pays for the right to operate a business. An \$800-per-year tax is levied "for the privilege of doing business in California." A three-year license to sell hot dogs in front of New York City's Metropolitan Museum of Art costs \$900,600. These taxes do not vary with output, so they affect firms' fixed costs only—not their variable costs.

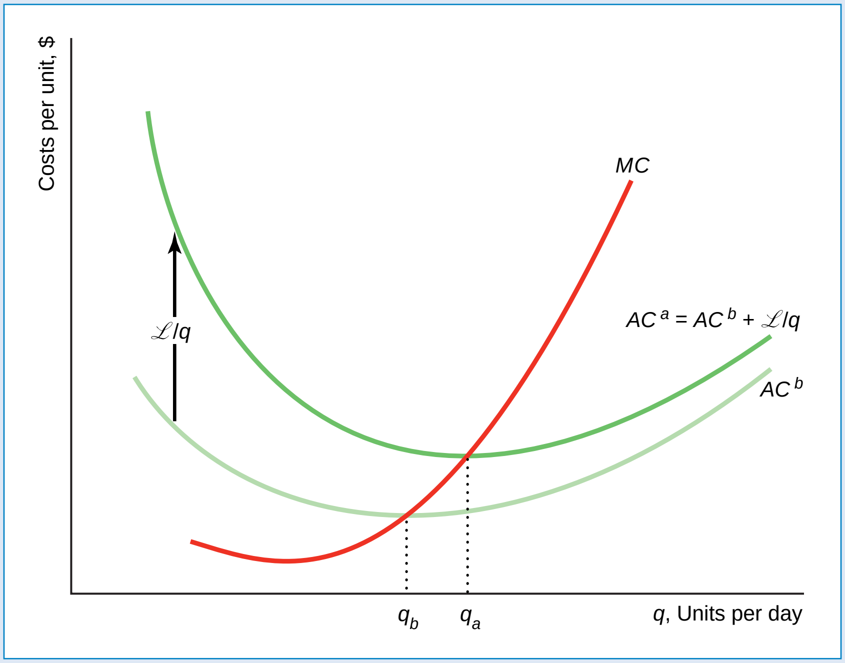
Solved Problem 7.1

What is the effect of a lump-sum franchise tax $\mathcal{L}$ on the quantity at which a firm's after-tax average cost curve reaches its minimum? (Assume that the firm's before-tax average cost curve is U-shaped.)

Answer

1. *Determine the average tax per unit of output*: Because the franchise tax is a lump-sum payment that does not vary with output, the more the firm produces, the less tax it pays per unit. The tax per unit is $\mathcal{L}/q$. If the firm sells only 1 unit, its cost is $\mathcal{L}$; however, if it sells 100 units, its tax payment per unit is only $\mathcal{L}/100$.

2. *Show how the tax per unit affects the average cost*: The firm's after-tax average cost, AC^a, is the sum of its before-tax average cost, AC^b, and its average tax payment per unit, $\mathcal{L}/q$. Because the average tax payment per unit falls with output, the gap between the after-tax average cost curve and the before-tax average cost curve also falls with output on the graph.



3. *Determine the effect of the tax on the marginal cost curve*: Because the franchise tax does not vary with output, it does not affect the marginal cost curve.

4. *Compare the minimum points of the two average cost curves*: The marginal cost curve crosses from below both average cost curves at their minimum points. Because the after-tax average cost lies above the before-tax average cost curve, the quantity at which the after-tax average cost curve reaches its minimum, q_a, is larger than the quantity, q_b, at which the before-tax average cost curve achieves a minimum.

Short-Run Cost Summary

We discussed three cost-level curves—total cost, fixed cost, and variable cost—and four cost-per-unit curves—average cost, average fixed cost, average variable cost, and marginal cost. Understanding the shapes of these curves and the relationships between them is crucial to understanding the analysis of firm behavior in the rest of this book. Fortunately, we can derive most of what we need to know about the shapes and the relationships between the curves using four basic concepts:

- In the short run, the cost associated with inputs that cannot be adjusted is fixed, while the cost from inputs that can be adjusted is variable.
- Given that input prices are constant, the shapes of the variable cost and cost curves are determined by the production function.
- Where there are diminishing marginal returns to a variable input, the variable cost and cost curves become relatively steep as output increases, so the average cost, average variable, and marginal cost curves rise with output.
- Because of the relationship between marginals and averages, both the average cost and average variable cost curves fall when marginal cost is below them and rise when marginal cost is above them, so the marginal cost cuts both these average cost curves at their minimum points.

7.3 LONG-RUN COSTS

In the long run, the firm adjusts all its inputs so that its cost of production is as low as possible. The firm can change its plant size, design and build new machines, and otherwise adjust inputs that were fixed in the short run.

Although firms may incur fixed costs in the long run, these fixed costs are *avoidable* (rather than *sunk*, as in the short run). The rent of F per month that a restaurant pays is a fixed cost because it does not vary with the number of meals (output) served. In the short run, this fixed cost is sunk: The firm must pay F even if the restaurant does not operate. In the long run, this fixed cost is avoidable: The firm does not have to pay this rent if it shuts down. The long run is determined by the length of the rental contract during which time the firm is obligated to pay rent.

In our examples throughout this chapter, we assume that all inputs can be varied in the long run so that there are no long-run fixed costs ($F = 0$). As a result, the long-run total cost equals the long-run variable cost: $C = VC$. Thus our firm is concerned about only three cost concepts in the long run—total cost, average cost, and marginal cost—instead of the seven cost concepts that it considers in the short run.

To produce a given quantity of output at minimum cost, our firm uses information about the production function and the price of labor and capital. The firm chooses how much labor and capital to use in the long run, whereas the firm chooses only how much labor to use in the short run when capital is fixed. As a consequence, the firm's long-run cost is lower than its short-run cost of production if it has to use the "wrong" level of capital in the short run. In this section, we show how a firm picks the cost-minimizing combinations of inputs in the long run.

Input Choice

A firm can produce a given level of output using many different *technologically efficient* combinations of inputs, as summarized by an isoquant (Chapter 6). From among the technologically efficient combinations of inputs, a firm wants to choose the particular bundle with the lowest cost of production, which is the *economically efficient* combination of inputs. To do so, the firm combines information about technology from the isoquant with information about the cost of labor and capital.

We now show how information about cost can be summarized in an *isocost line*. Then we show how a firm can combine the information in an isoquant and isocost lines to pick the economically efficient combination of inputs.

Isocost Line. The cost of producing a given level of output depends on the price of labor and capital. The firm hires L hours of labor services at a wage of w per hour, so its labor cost is wL. The firm rents K hours of machine services r per hour, so its capital cost is rK. (If the firm owns the capital, r is the implicit rental rate.) The firm's total cost is the sum of its labor and capital costs:

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

The firm can hire as much labor and capital as it wants at these constant input prices.

The firm can use many combinations of labor and capital that cost the same amount. Suppose that the wage rate, w, is \$5 an hour and the rental rate of capital, r, is \$10. Five of the many combinations of labor and capital that the firm can use that cost \$100 are listed in Table 7.3. These combinations of labor and capital are plotted on an **isocost line**, which is all the combinations of inputs that require the same (*iso-*) total expenditure (*cost*). Figure 7.4 shows three isocost lines. The \$100 isocost line represents all the combinations of labor and capital that the firm can buy for \$100, including the combinations *a* through *e* in Table 7.3.

Table 7.3 Bundles of Labor and Capital That Cost the Firm \$100

Bundle	Labor, L	Capital, K	Labor Cost, $wL = \$5L$	Capital Cost, $rK = \$10K$	Total Cost, $wL + rK$
a	20	0	\$100	\$0	\$100
b	14	3	\$70	\$30	\$100
c	10	5	\$50	\$50	\$100
d	6	7	\$30	\$70	\$100
e	0	10	\$0	\$10	\$100

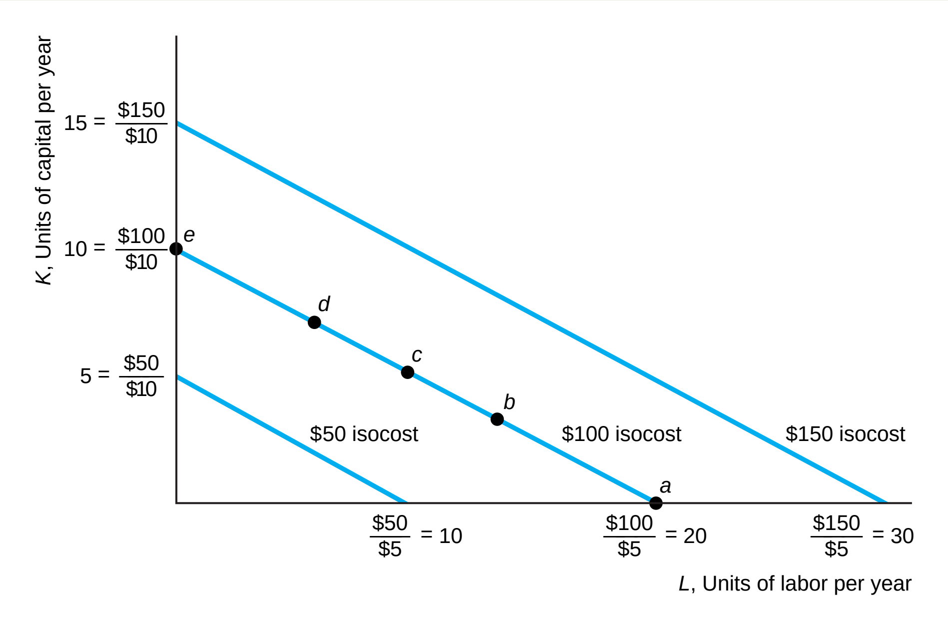


Figure 7.4 A Family of Isocost Lines. An isocost line shows all the combinations of labor and capital that cost the firm the same amount. The greater the total cost, the farther from the origin the isocost lies. All the isocosts have the same slope, $-w/r = -\frac{1}{2}$. The slope shows the rate at which the firm can substitute capital for labor holding total cost constant: For each extra unit of capital it uses, the firm must use two fewer units of labor to hold its cost constant.

Along an isocost line, cost is fixed at a particular level, $\overline{C}$, so by setting cost at $\overline{C}$ in Equation 7.3, we can write the equation for the $\overline{C}$ isocost line as

$$\overline{C} = wL + rK.$$

Using algebra, we can rewrite this equation to show how much capital the firm can buy if it spends a total of $\overline{C}$ and purchases L units of labor:

$$K = \frac{\overline{C}}{r} - \frac{w}{r}L. \tag{7.4}$$

By substituting $\overline{C}$ = \$100, w = \$5, and r = \$10 in Equation 7.4, we find that the \$100 isocost line is $K = 10 - \frac{1}{2}L$. We can use Equation 7.4 to derive three properties of isocost lines.

First, where the isocost lines hit the capital and labor axes depends on the firm's cost, $\overline{C}$, and on the input prices. The $\overline{C}$ isocost line intersects the capital axis where the firm is using only capital. Setting $L = 0$ in Equation 7.4, we find that the firm buys $K = \overline{C}/r$ units of capital. In the figure, the \$100 isocost line intersects the capital axis at \$100/\$10 = 10 units of capital. Similarly, the intersection of the isocost line with the labor axis is at $\overline{C}/w$, which is the amount of labor the firm hires if it uses only labor. In the figure, the intersection of the \$100 isocost line with the labor axis occurs at $L = 20$, where $K = 10 - \frac{1}{2} \times 20 = 0$.

Second, isocosts that are farther from the origin have higher costs than those that are closer to the origin. Because the isocost lines intersect the capital axis at $\bar{C}/r$ and the labor axis at $\bar{C}/w$, an increase in the cost shifts these intersections with the axes proportionately outward. The \$50 isocost line hits the capital axis at 5 and the labor axis at 10, whereas the \$100 isocost line intersects at 10 and 20.

Third, the slope of each isocost line is the same. From Equation 7.4, if the firm increases labor by ΔL, it must decrease capital by

$$\Delta K = -\frac{w}{r}\Delta L.$$

Dividing both sides of this expression by ΔL, we find that the slope of an isocost line, $\Delta K/\Delta L$, is $-w/r$. Thus the slope of the isocost line depends on the relative prices of the inputs. The slope of the isocost lines in the figure is $-w/r = -\$5/\$10 = -\frac{1}{2}$. If the firm uses two more units of labor, $\Delta L = 2$, it must reduce capital by one unit, $\Delta K = -\frac{1}{2}\Delta L = -1$, to keep its total cost constant. Because all isocost lines are based on the same relative prices, they all have the same slope, so they are parallel.

The isocost line plays a similar role in the firm's decision making as the budget line does in consumer decision making. Both an isocost line and a budget line are straight lines whose slopes depend on relative prices. There is an important difference between them, however. The consumer has a single budget line determined by the consumer's income. The firm faces many isocost lines, each of which corresponds to a different level of expenditures the firm might make. A firm may incur a relatively low cost by producing relatively little output with few inputs, or it may incur a relatively high cost by producing a relatively large quantity.

Combining Cost and Production Information. By combining the information about costs contained in the isocost lines with information about efficient production summarized by an isoquant, a firm chooses the lowest-cost way to produce a given level of output. We examine how our Norwegian printing firm picks the combination of labor and capital that minimizes its cost of producing 100 units of output. Figure 7.5 shows the isoquant for 100 units of output (based on Griliches and Ringstad, 1971) and the isocost lines where the rental rate of a unit of capital is 8 kroner (the Norwegian monetary unit, abbreviated *kr*) and the wage rate is 24 kr.

The firm can choose any of three equivalent approaches to minimize its cost:

- **Lowest-isocost rule**: Pick the bundle of inputs where the lowest isocost line touches the isoquant.
- **Tangency rule**: Pick the bundle of inputs where the isoquant is tangent to the isocost line.
- **Last-dollar rule**: Pick the bundle of inputs where the last dollar spent on one input gives as much extra output as the last dollar spent on any other input.

Using the *lowest-isocost rule*, the firm minimizes its cost by using the combination of inputs on the isoquant that is on the lowest isocost line that touches the isoquant. The lowest possible isoquant that will allow the printing firm to produce 100 units of output is the 2,000-kr isocost line. This isocost line touches the isoquant at the bundle of inputs x, where the firm uses $L = 50$ workers and $K = 100$ units of capital.

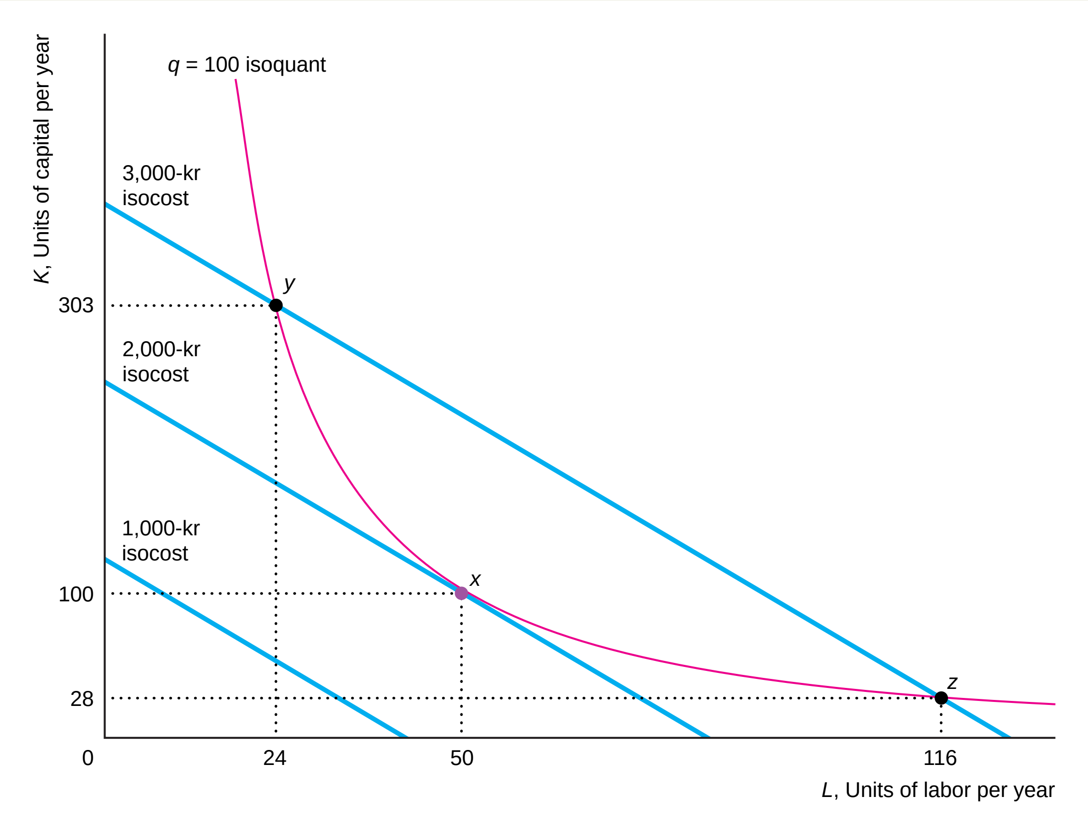


Figure 7.5 Cost Minimization. The Norwegian printing firm minimizes its cost of producing 100 units of output by producing at x (L = 50 and K = 100). This cost-minimizing combination of inputs is determined by the tangency between the q = 100 isoquant and the lowest isocost line, 2,000 kr, that touches that isoquant. At x, the isocost is tangent to the isoquant, so the slope of the isocost, $-w/r = -3$, equals the slope of the isoquant, which is the negative of the marginal rate of technical substitution. That is, the rate at which the firm can trade capital for labor in the input markets equals the rate at which it can substitute capital for labor in the production process.

How do we know that x is the least costly way to produce 100 units of output? We need to demonstrate that other practical combinations of input produce less than 100 units or produce 100 units at greater cost.

If the firm spent less than 2,000 kr, it could not produce 100 units of output. Each combination of inputs on the 1,000-kr isocost line lies below the isoquant, so the firm cannot produce 100 units of output for 1,000 kr.

The firm can produce 100 units of output using other combinations of inputs beside x; however, using these other bundles of inputs is more expensive. For example, the firm can produce 100 units of output using the combinations y (L = 24, K = 303) or z (L = 116, K = 28). Both these combinations, however, cost the firm 3,000 kr.

If an isocost line crosses the isoquant twice, as the 3,000-kr isocost line does, there must be another lower isocost line that also touches the isoquant. The lowest possible isocost line that touches the isoquant, the 2,000-kr isocost line, is tangent to the isoquant at a single bundle, x. Thus the firm may use the *tangency rule*: The firm chooses the input bundle where the relevant isoquant is tangent to an isocost line to produce a given level of output at the lowest cost.

We can interpret this tangency or cost minimization condition in two ways. At the point of tangency, the slope of the isoquant equals the slope of the isocost. As we showed in Chapter 6, the slope of the isoquant is the marginal rate of technical substitution (*MRTS*). The slope of the isocost is the negative of the ratio of the wage to the cost of capital, $-w/r$. Thus to minimize its cost of producing a given level of output, a firm chooses its inputs so that the marginal rate of technical substitution equals the negative of the relative input prices:

$$MRTS = -\frac{w}{r} \tag{7.5}$$

The firm picks inputs so that the rate at which it can substitute capital for labor in the production process, the *MRTS*, exactly equals the rate at which it can trade capital for labor in input markets, $-w/r$.

The printing company's marginal rate of technical substitution is $-1.5K/L$. At $K = 100$ and $L = 50$, its *MRTS* is -3, which equals the negative of the ratio of the input prices it faces, $-w/r = -24/8 = -3$. In contrast, at y, the isocost cuts the isoquant so the slopes are not equal. At y, the *MRTS* is -18.9375, which is greater than the ratio of the input price, 3. Because the slopes are not equal at y, the firm can produce the same output at lower cost. As the figure shows, the cost of producing at y is 3,000 kr, whereas the cost of producing at x is only 2,000 kr.

We can interpret the condition in Equation 7.5 in another way. We showed in Chapter 6 that the marginal rate of technical substitution equals the negative of the ratio of the marginal product of labor to that of capital: $MRTS = -MP_L/MP_K$. Thus the cost-minimizing condition in Equation 7.5 is (taking the absolute value of both sides)

$$\frac{MP_L}{MP_K} = \frac{w}{r}. \tag{7.6}$$

This expression may be rewritten as

$$\frac{MP_L}{w} = \frac{MP_K}{r}. \tag{7.7}$$

Equation 7.7 states the *last-dollar rule*: Cost is minimized if inputs are chosen so that the last dollar spent on labor adds as much extra output as the last dollar spent on capital.

The printing firm's marginal product of labor is $MP_L = 0.6q/L$, and its marginal product of capital is $MP_K = 0.4q/K$.[5] At Bundle x, the printing firm's marginal

[5]The printing firm's production function, $q = 1.52L^{0.6}K^{0.4}$, is a Cobb-Douglas production function. The marginal product formula for Cobb-Douglas production functions is derived in Appendix 6B.

product of labor is 1.2 (= 0.6 × 100/50) and its marginal product of capital is 0.4. The last krone spent on labor gets the firm

$$\frac{MP_L}{w} = \frac{1.2}{24} = 0.05$$

more output. The last krone spent on capital also gets the firm

$$\frac{MP_K}{r} = \frac{0.4}{8} = 0.05$$

extra output. Thus spending one more krone on labor at x gets the firm as much extra output as spending the same amount on capital. Equation 7.6 holds, so the firm is minimizing its cost of producing 100 units of output.

If instead the firm produced at y, where it is using more capital and less labor, its MP_L is 2.5 (= 0.6 × 100/24) and the MP_K is approximately 0.13 (≈ 0.4 × 100/303). As a result, the last krone spent on labor gets $MP_L/w \approx 0.1$ more unit of output, whereas the last krone spent on capital gets only a fourth as much extra output, $MP_K/r \approx 0.017$. At y, if the firm shifts one krone from capital to labor, output falls by 0.017 because there is less capital and increases by 0.1 because there is more labor for a net gain of 0.083 more output at the same cost. The firm should shift even more resources from capital to labor—which increases the marginal product of capital and decreases the marginal product of labor—until Equation 7.6 holds with equality at x.

To summarize, we demonstrated that there are three equivalent rules that the firm can use to pick the lowest-cost combination of inputs to produce a given level of output when isoquants are smooth: the lowest-isocost rule, the tangency rule (Equations 7.5 and 7.6), and the last-dollar rule (Equation 7.7). If the isoquant is not smooth, the lowest-cost method of production cannot be determined by using the tangency rule or the last-dollar rule. The lowest-isocost rule always works—even when isoquants are not smooth—as we now illustrate.

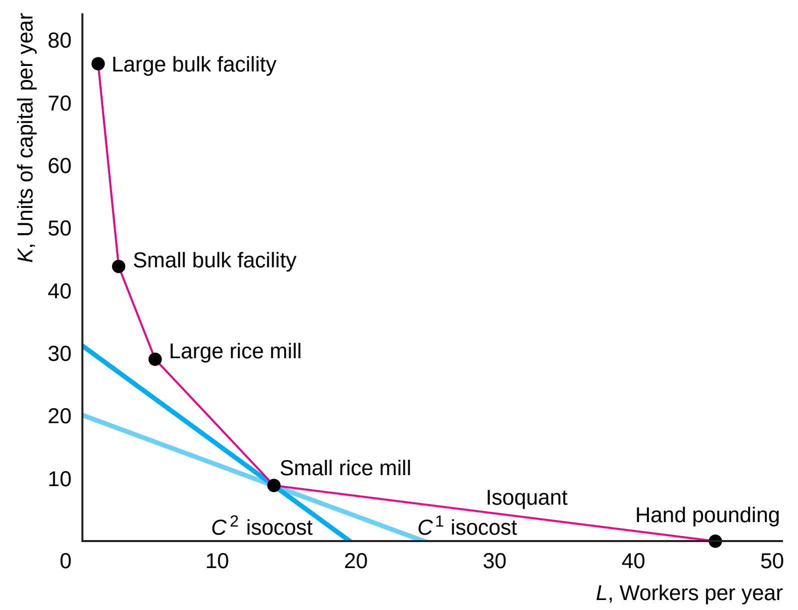
RICE MILLING ON JAVA

A dramatic change occurred in the process used for milling rice on Java in the early 1970s (Timmer, 1984). About 80% of that island's rice crop was hand-pounded in 1971. By 1973, less than 50%, and perhaps as little as 10%, was hand-pounded. During that period, thousands of new small, mechanical rice mills had been built. What prompted this change in technique? The change was due to a fall in the cost of capital, which increased the factor price of labor relative to capital.

Rice may be milled in five ways: by (1) hand pounding, (2) small-mill processing, (3) large-mill processing, (4) using a small bulk facility, or (5) using a large bulk facility. These methods vary in the amounts of labor and capital services required. Hand pounding, a technique that is at least a millennium old, is labor intensive: Workers use only a wooden pounding pole and pestle. Small-mill processing employs equipment to hull and polish the rice but relies on sun drying rather than mechanical drying equipment. Large mills (which are rare) hull, polish, and dry using a combination of machines and sunlight. Small and large bulk facilities mill and store high-quality rice; these were used elsewhere but not on Java.

What was the least expensive way to process rice on Java? The least-cost combination of laborers and capital is determined by finding the lowest isocost that touches the relevant isoquant.

The isoquant for 10 million rupees' worth of processed rice (see graph) has kinks. The point at each kink represents one of the five milling processes. The straight line connecting a pair of points shows how the same output could be produced by using a combination of those two processes. For example, at the midpoint of the line between hand pounding and small mills, half the rice is produced by hand pounding and half by the small mill.



Because the isoquant has kinks, small changes in the relative cost of the two inputs do not necessarily lead to a change in technique. Two different sloped isocost lines, C^1 and C^2 (which reflect the range of relative factor prices in 1973), touch the isoquant at the same point: The small rice mill is the least expensive technique. Prior to the early 1970s, the isocost line was flatter than C^2 (the price of capital was relatively high compared to that of labor) so a less capital-intensive technique was used.

Factor Price Changes. Once the Norwegian printing firm determines the lowest-cost combination of inputs to produce a given level of output, it uses that method as long as the input prices remain constant. How should the firm change its behavior if the cost of one of the factors changes? Suppose that the wage falls from 24 kr to 8 kr but the rental rate of capital stays constant at 8 kr.

The firm minimizes its new cost by substituting away from the now relatively more expensive input, capital, toward the now relatively less expensive input, labor. The change in the wage does not affect technological efficiency, so it does not affect the isoquant in Figure 7.6. Because of the wage decrease, the new isocost lines have a flatter slope, $-w/r = -8/8 = -1$, than the original isoquant lines, $-w/r = -24/8 = -3$.

The relatively steep original isocost line is tangent to the 100-unit isoquant at Bundle x ($L = 50$, $K = 100$). The new, flatter isocost line is tangent to the isoquant at Bundle v ($L = 77$, $K = 52$). Thus the firm uses more labor and less capital as labor becomes relatively less expensive. Moreover, the firm's cost of producing 100 units falls from 2,000 kr to 1,032 kr because of the fall in the wage.

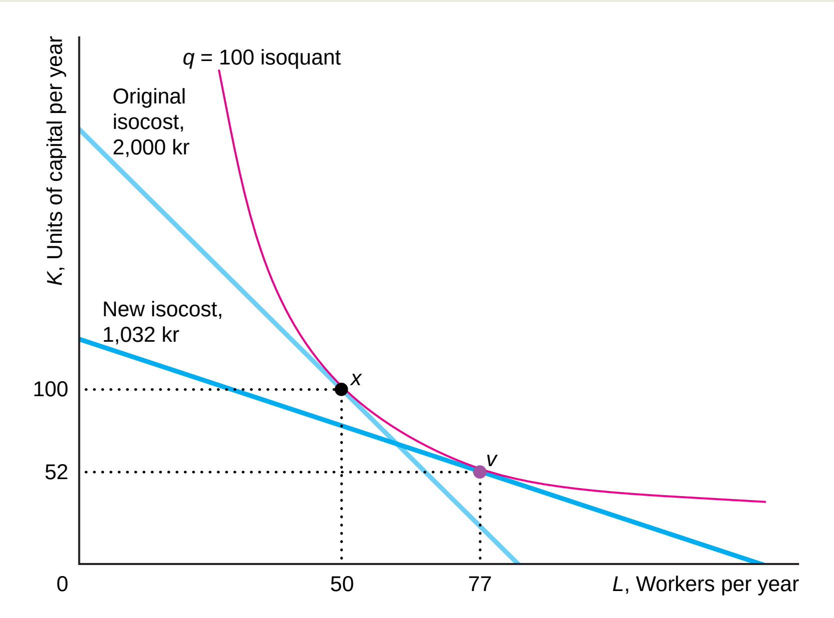


Figure 7.6 **Change in Factor Price.** Originally, the wage was 24 kr and the rental rate of capital was 8 kr, so the lowest isocost line (2,000 kr) was tangent to the $q = 100$ isoquant at x ($L = 50$, $K = 100$). When the wage falls to 8 kr, the isocost lines became flatter: Labor became relatively less expensive than capital. The slope of the isocost lines falls from $-w/r = -24/8 = -3$ to $-8/8 = -1$. The new lowest isocost line (1,032 kr) is tangent at v ($L = 77$, $K = 52$). Thus when the wage falls, the firm uses more labor and less capital to produce a given level of output, and the cost of production falls from 2,000 kr to 1,032 kr.

This example illustrates that a change in the relative prices of inputs affects the mix of inputs that a firm uses. In contrast, if all prices increase by the same amount, so that the relative prices do not change, the firm continues to produce any given amount of output using the same input combination as it did originally, as we now show.

Solved Problem 7.2

When the input prices were $\hat{w}$ and $\hat{r}$, a firm produced $\hat{q}$ units of output using $\hat{L}$ units of labor and $\hat{K}$ units of capital. Suppose that the prices of both its inputs doubles. If the firm continues to produce $\hat{q}$ units of output, will it change the amount of labor and capital it uses? What happens to its cost of producing $\hat{q}$?

Answer

1. *Determine whether the change affects the slopes of the isoquant or the isocost lines*: A change in input prices does not affect the isoquant, which depends only on technology (the production function). Moreover, the doubling of the input prices does not affect the slope of the isocost lines. The original slope was $-\hat{w}/\hat{r}$, and the new slope is $-(2\hat{w})/(2\hat{r}) = -\hat{w}/\hat{r}$.

2. *Using a rule for cost minimization, determine whether the firm changes its input mix*: A firm minimizes its cost by producing where its isoquant is tangent to the lowest possible isocost line. That is, the firm produces where the slope of its isoquant, *MRTS*, equals the slope of its isocost line, $-w/r$. Because the slopes of the isoquant and the isocost lines are unchanged after input prices doubled, the firm continues to produce $\hat{q}$ using the same amount of labor, $\hat{L}$, and capital, $\hat{K}$, as originally.

3. *Calculate the original cost and the new cost and compare them*: The firm's original cost of producing $\hat{q}$ units of output was $\hat{w}\hat{L} + \hat{r}\hat{K} = \hat{C}$. Its new cost of producing the same amount of output is $(2\hat{w})\hat{L} + (2\hat{r})\hat{K} = 2\hat{C}$. Thus its cost of producing $\hat{q}$ doubled when the input prices doubled. The isocost lines have the same slope as before, but the cost associated with each isocost line has doubled.

How Long-Run Cost Varies with Output

We now know how a firm determines the cost-minimizing output for any given level of output. By repeating this analysis for different output levels, the firm determines how its cost varies with output.

Panel a of Figure 7.7 shows the relationship between the lowest-cost factor combinations and various levels of output for the printing firm when input prices are held constant at w = 24 kr and r = 8 kr. The curve through the tangency points is the long-run **expansion path**: the cost-minimizing combination of labor and capital for each output level. The lowest-cost way to produce 100 units of output is to use the labor and capital combination x (L = 50 and K = 100), which lies on the 2,000-kr isocost line. Similarly, the lowest-cost way to produce 200 units is to use z, which is on the 4,000-kr isocost line. The expansion path goes through x and z.

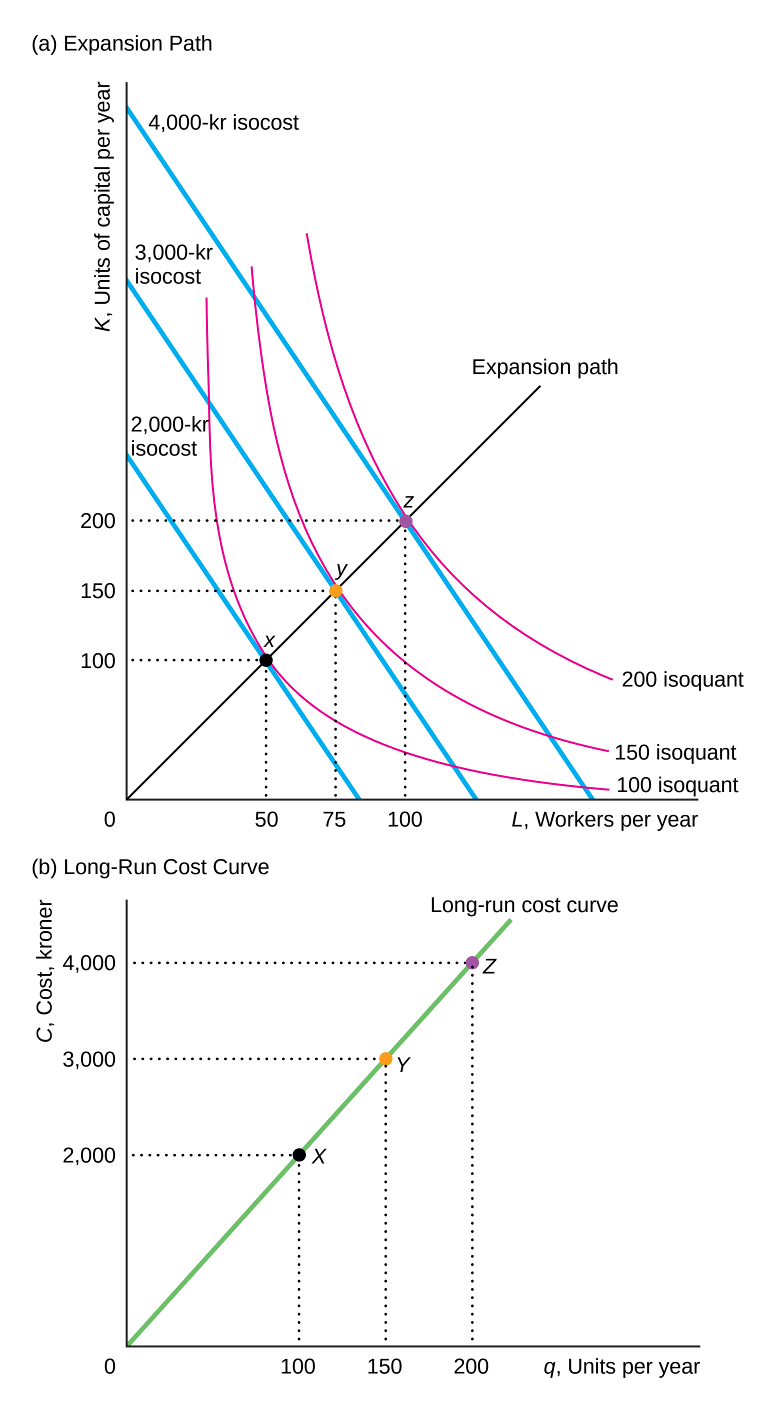


Figure 7.7 Expansion Path and Long-Run Cost Curve. (a) The curve through the tangency points between isocost lines and isoquants, such as *x*, *y*, and *z*, is called the expansion path. The points on the expansion path are the cost-minimizing combinations of labor and capital for each output level. (b) The expansion path shows the same relationship between long-run cost and output as the long-run cost curve.

The expansion path of the printing firm in the figure is a straight line through the origin with a slope of 2: At any given output level, the firm uses twice as much capital as labor.[6] To double its output from 100 to 200 units, the firm doubles the amount of labor from 50 to 100 workers and doubles the amount of capital from 100 to 200 units. Because both inputs double when output doubles from 100 to 200, cost also doubles.

The printing firm's expansion path contains the same information as its long-run cost function, $C(q)$, which shows the relationship between the cost of production and output. From inspection of the expansion path, to produce q units of output takes $K = q$ units of capital and $L = q/2$ units of labor. Thus the long-run cost of producing q units of output is

$$C(q) = wL + rK = wq/2 + rq = (w/2 + r)q = (24/2 + 8)q = 20q.$$

That is, the long-run cost function corresponding to this expansion path is $C(q) = 20q$. This cost function is consistent with the expansion path in panel a: $C(100) = 2{,}000$ kr at x on the expansion path, $C(150) = 3{,}000$ kr at y, and $C(200) = 4{,}000$ kr at z.

Panel b plots this long-run cost curve. Points X, Y, and Z on the cost curve correspond to points x, y, and z on the expansion path. For example, the 2,000-kr isocost line goes through x, which is the lowest-cost combination of labor and capital that can produce 100 units of output. Similarly, X on the long-run cost curve is at 2,000 kr and 100 units of output. Consistent with the expansion path, the cost curve shows that as output doubles, cost doubles.

7.3 **What is the long-run cost function for a fixed-proportions production function (Chapter 6) when it takes one unit of labor and one unit of capital to produce one unit of output? Describe the long-run cost curve.**

Answer

Multiply the inputs by their prices, and sum to determine total cost. The long-run cost of producing q units of output is $C(q) = wL + rK = wq + rq = (w + r)q$. Cost rises in proportion to output. The long-run cost curve is a straight line with a slope of $w + r$.

The Shape of Long-Run Cost Curves

The shapes of the average cost and marginal cost curves depend on the shape of the long-run cost curve. To illustrate these relationships, we examine the long-run cost curves of a typical firm that has a U-shaped long-run average cost curve.

The long-run cost curve in panel a of Figure 7.8 corresponds to the long-run average and marginal cost curves in panel b. Unlike the straight-line long-run cost curves of the printing firm and the firm with fixed-proportions production, the long-run

[6] In Appendix 7C, we show that the expansion path for a Cobb-Douglas production function is $K = [\beta w/(\alpha r)]L$. The expansion path for the printing firm is $K = [(0.4 \times 24)/(0.6 \times 8)]L = 2L$.

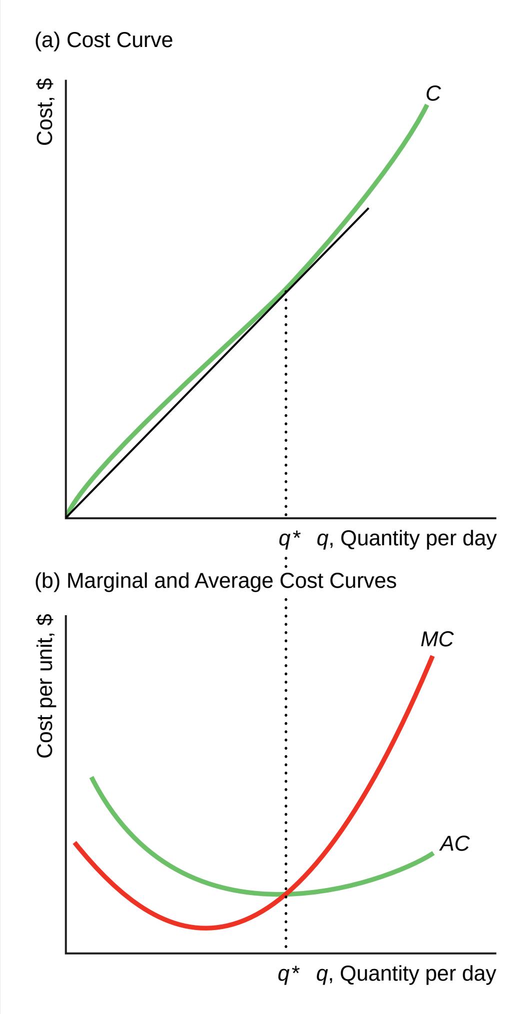


Figure 7.8 Long-Run Cost Curves. (a) The long-run cost curve rises less rapidly than output at output levels below q^* and more rapidly at higher output levels. (b) As a consequence, the marginal cost and average cost curves are U-shaped. The marginal cost crosses the average cost at its minimum at q^*.

cost curve of this firm rises less than in proportion to output at outputs below q^* and then rises more rapidly.

We can apply the same type of analysis that we used to study short-run curves to look at the geometric relationship between long-run total, average, and marginal curves. A line from the origin is tangent to the long-run cost curve at q^*, where the marginal cost curve crosses the average cost curve, because the slope of that line equals the marginal and average costs at that output. The long-run average cost curve falls when the long-run marginal cost curve is below it and rises when the long-run marginal cost curve is above it. Thus the marginal cost crosses the average cost curve at the lowest point on the average cost curve.

Why does the average cost curve first fall and then rise, as in panel b? The explanation differs from those given for why short-run average cost curves are U-shaped.

A key reason why the short-run average cost is initially downward sloping is that the average fixed cost curve is downward sloping: Spreading the fixed cost over more units of output lowers the average fixed cost per unit. There are no fixed costs in the long run, however, so fixed costs cannot explain the initial downward slope of the long-run average cost curve.

A major reason why the short-run average cost curve slopes upward at higher levels of output is diminishing marginal returns. In the long run, however, all factors can be varied, so diminishing marginal returns do not explain the upward slope of a long-run average cost curve.

Ultimately, as with the short-run curves, the shape of the long-run curves is determined by the production function relationship between output and inputs. In the long run, returns to scale play a major role in determining the shape of the average cost curve and other cost curves. As we discussed in Chapter 6, increasing all inputs in proportion may cause output to increase more than in proportion (increasing returns to scale) at low levels of output, in proportion (constant returns to scale) at intermediate levels of output, and less than in proportion (decreasing returns to scale) at high levels of output. If a production function has this returns-to-scale pattern and the prices of inputs are constant, long-run average cost must be U-shaped.

To illustrate the relationship between returns to scale and long-run average cost, we use the returns-to-scale example of Figure 6.5, the data for which are reproduced in Table 7.4. The firm produces one unit of output using a unit each of labor and capital. Given a wage and rental cost of capital of \$6 per unit, the total cost and average cost of producing this unit are both \$12. Doubling both inputs causes output to increase more than in proportion to 3 units, reflecting increasing returns to scale. Because cost only doubles and output triples, the average cost falls. A cost function is said to exhibit **economies of scale** if the average cost of production falls as output expands.

Table 7.4 Returns to Scale and Long-Run Costs

Output, Q	Labor, L	Capital, K	Cost, $C = wL + rK$	Average Cost, $AC = C/q$	Returns to Scale
1	1	1	12	12	
3	2	2	24	8	Increasing
6	4	4	48	8	Constant
8	8	8	96	12	Decreasing

$w = r = \$6$ per unit.

Doubling the inputs again causes output to double as well—constant returns to scale—so the average cost remains constant. If an increase in output has no effect on average cost—the average cost curve is flat—there are *no economies of scale*.

Doubling the inputs once more causes only a small increase in output—decreasing returns to scale—so average cost increases. A firm suffers from **diseconomies of scale** if average cost rises when output increases.

Returns to scale in the production function are a sufficient but not necessary condition for economies of scale in the average cost curve. In the long run, a firm may change the ratio of capital to labor that it uses as it expands output. As a result, the firm could have economies of scale in costs without increasing returns to scale in production or could have diseconomies of scale in costs without decreasing returns to scale in production.

Consider a firm that has constant returns to scale in production at every output level. At small levels of output, the firm uses lots of labor and commonly available tools. At large levels of output, the firm designs and builds its own specialized equipment and uses relatively few workers, thereby lowering its average cost. Such a firm has economies of scale in cost despite having constant returns to scale in production.

Average cost curves can have many different shapes. Competitive firms typically have U-shaped average cost curves. Average cost curves in noncompetitive markets may be U-shaped, L-shaped (average cost at first falls rapidly and then levels off as output increases), everywhere downward sloping, or everywhere upward sloping or have other shapes. The shapes of the average cost curves indicate whether the production process has economies or diseconomies of scale.

Table 7.5 summarizes the shapes of average cost curves of firms in various Canadian manufacturing industries (as estimated by Robidoux and Lester, 1992). The table shows that U-shaped average cost curves are the exception rather than the rule in Canadian manufacturing and that nearly one-third of these average cost curves are L-shaped.

Some of these apparently L-shaped average cost curves may be part of a U-shaped curve with long, flat bottoms, where we don't observe any firm producing enough to exhibit diseconomies of scale. Cement firms provide an example of such a cost curve.

Table 7.5 Shape of Average Cost Curves in Canadian Manufacturing

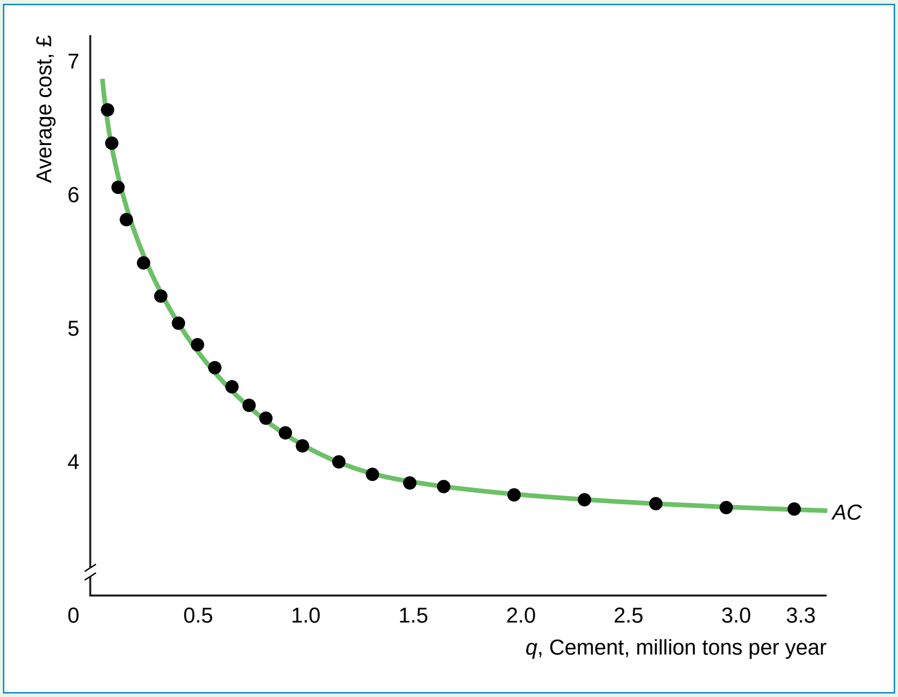
Scale Economies	Share of Manufacturing Industries, %	
Economies of scale: initially downward-sloping *AC*	57	
Everywhere downward-sloping *AC*		18
L-shaped *AC* (downward-sloping, then flat)		31
U-shaped *AC*		8
No economies of scale: flat *AC*	23	
Diseconomies of scale: upward-sloping *AC*	14	

Source: Robidoux and Lester (1992).

Application

AVERAGE COST OF CEMENT FIRMS

Cement producers enjoy substantial economies of scale. The dots in the figure show the observed average costs for British cement firms (Norman, 1979). The estimated average cost curve (fitted through these dots) shows that the



average cost curve is L-shaped: Average cost falls rapidly at first and then more slowly. Norman presented evidence that average cost curves have similar shapes in the United States and West Germany. Jha, Murty, Paul, and Sahni (1991) also found economies of scale in Indian cement, lime, and plaster plants. They estimated that a 1% increase in a plant's output lowers its average cost by 0.14%. Thus the long-run average cost in India is also L-shaped.

Estimating Cost Curves Versus Introspection

Economists use statistical methods to estimate a cost function. Sometimes, however, we can infer the shape by casual observation and deductive reasoning.

For example, in the good old days, the Good Humor company sent out fleets of ice-cream trucks to purvey its products. It seems likely that the company's production process had fixed proportions and constant returns to scale: If it wanted to sell more, Good Humor dispatched one more truck and one more driver. Drivers and trucks are almost certainly nonsubstitutable inputs (the isoquants are right angles). If the cost of a driver is w per day, the rental cost is r per day, and q quantity of ice cream is sold in a day, then the cost function is $C = (w + r)q$.

Such deductive reasoning can lead one astray, as I once discovered. A water heater manufacturing firm provided me with many years of data on the inputs it used and the amount of output it produced. I also talked to the company's engineers about the production process and toured the plant (which resembled a scene from Dante's *Inferno*, with staggering noise levels and flames everywhere).

A water heater consists of an outside cylinder of metal, a liner, an electronic control unit, hundreds of tiny parts (screws, washers, etc.), and a couple of rods that slow corrosion. Workers cut out the metal for the cylinder, weld it together, and add the other parts. "OK," I said to myself, "this production process must be one of fixed proportions because the firm needs one of everything to produce a water heater. How could you substitute a cylinder for an electronic control unit? Or how can you substitute labor for metal?"

I then used statistical techniques to estimate the production and cost functions. Following the usual procedure, however, I did not assume that I knew the exact form of the functions. Rather, I allowed the data to "tell" me the type of production and cost functions. To my surprise, the estimates indicated that the production process was not one of fixed proportions. Rather, the firm could readily substitute between labor and capital.

"Surely I've made a mistake," I said to the plant manager after describing these results. "No," he said, "that's correct. There's a great deal of substitutability between labor and metal."

"How can they be substitutes?"

"Easy," he said. "We can use a lot of labor and waste very little metal by cutting out exactly what we want and being very careful. Or we can use relatively little labor, cut quickly, and waste more metal. When the cost of labor is relatively high, we waste more metal. When the cost of metal is relatively high, we cut more carefully." This practice minimizes the firm's cost.

7.4 LOWER COSTS IN THE LONG RUN

In its long-run planning, a firm chooses a plant size and makes other investments so as to minimize its long-run cost on the basis of how many units it produces. Once it chooses its plant size and equipment, these inputs are fixed in the short run. Thus the firm's long-run decision determines its short-run cost. Because the firm cannot vary its capital in the short run but can vary it in the long run, short-run cost is at least as high as long-run cost and is higher if the "wrong" level of capital is used in the short run.

Long-Run Average Cost as the Envelope of Short-Run Average Cost Curves

As a result, the long-run average cost is always equal to or below the short-run average cost. Suppose, initially, that the firm in Figure 7.9 has only three possible plant sizes. The firm's short-run average cost curve is $SRAC^1$ for the smallest possible plant. The average cost of producing q_1 units of output using this plant, point a on $SRAC^1$, is \$10. If instead the plant used the next larger plant size, its cost of

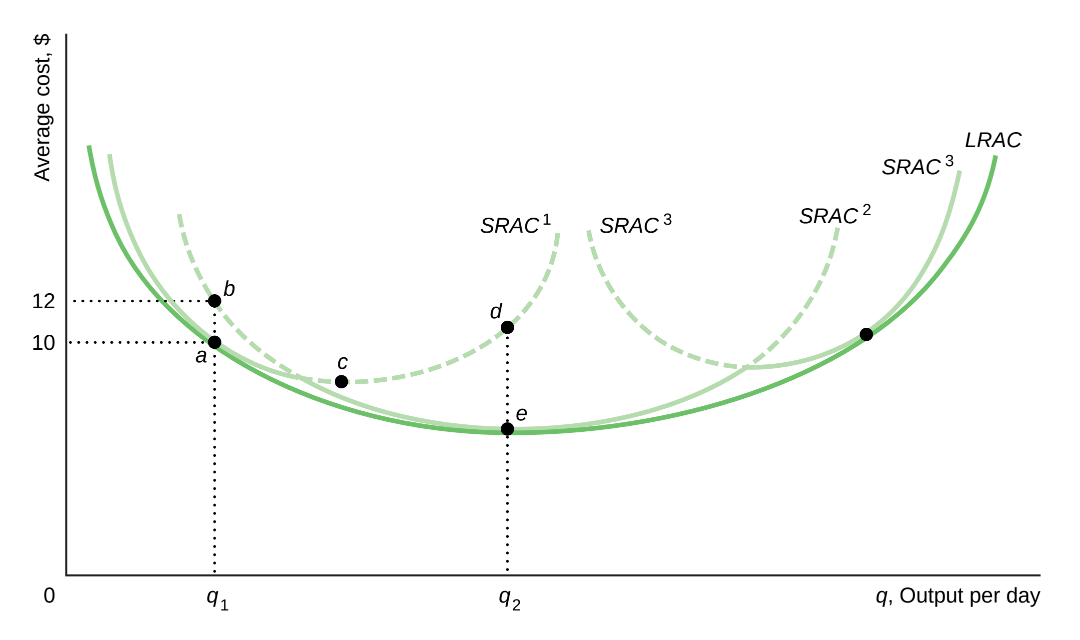


Figure 7.9 Long-Run Average Cost as the Envelope of Short-Run Average Cost Curves. If there are only three possible plant sizes, with short-run average costs $SRAC^1$, $SRAC^2$, and $SRAC^3$, the long-run average cost curve is the solid, scalloped portion of the three short-run curves. *LRAC* is the smooth and U-shaped long-run average cost curve if there are many possible short-run average cost curves.

producing q_1 units of output, point *b* on $SRAC^2$, would be \$12. Thus if the firm knows that it will produce only q_1 units of output, it minimizes its average cost by using the smaller plant size. If it expects to be producing q_2, its average cost is lower on the $SRAC^2$ curve, point *e*, than on the $SRAC^1$ curve, point *d*.

In the long run, the firm chooses the plant size that minimizes its cost of production, so it picks the plant size that has the lowest average cost for each possible output level. At q_1, it opts for the small plant size, whereas at q_2, it uses the medium plant size. Thus the long-run average cost curve is the solid, scalloped section of the three short-run cost curves.

If there are many possible plant sizes, the long-run average curve, *LRAC*, is smooth and U-shaped. The *LRAC* includes one point from each possible short-run average cost curve. This point, however, is not necessarily the minimum point from a short-run curve. For example, the *LRAC* includes *a* on $SRAC^1$ and not its minimum point, *c*. A small plant operating at minimum average cost cannot produce at as low an average cost as a slightly larger plant that is taking advantage of economies of scale.

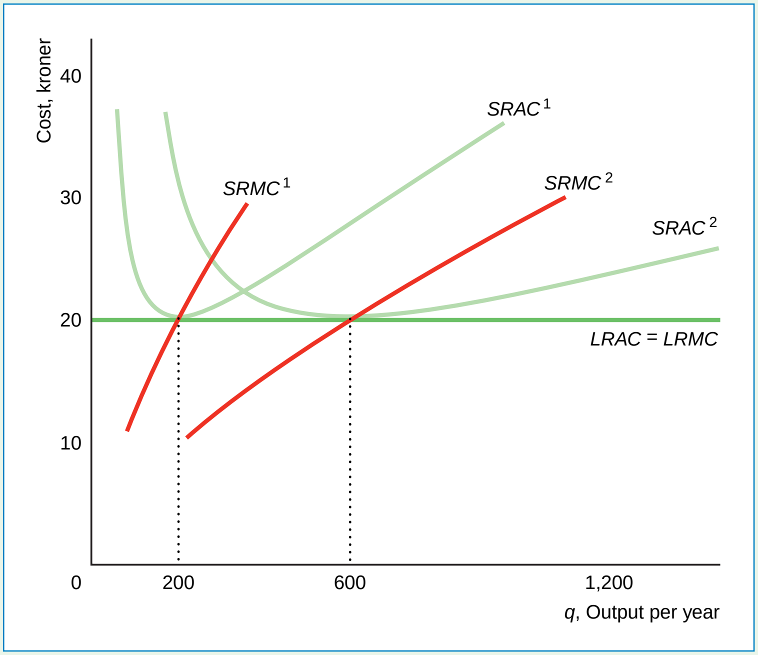
Application

LONG-RUN COST CURVES IN PRINTING AND OIL PIPELINES

Here we illustrate the relationship between long-run and short-run cost curves for our Norwegian printing firm and for oil pipelines. In the next application, we show the long-run cost when you or a firm chooses between a laser printer and an ink-jet printer, depending on how many pages will be printed.

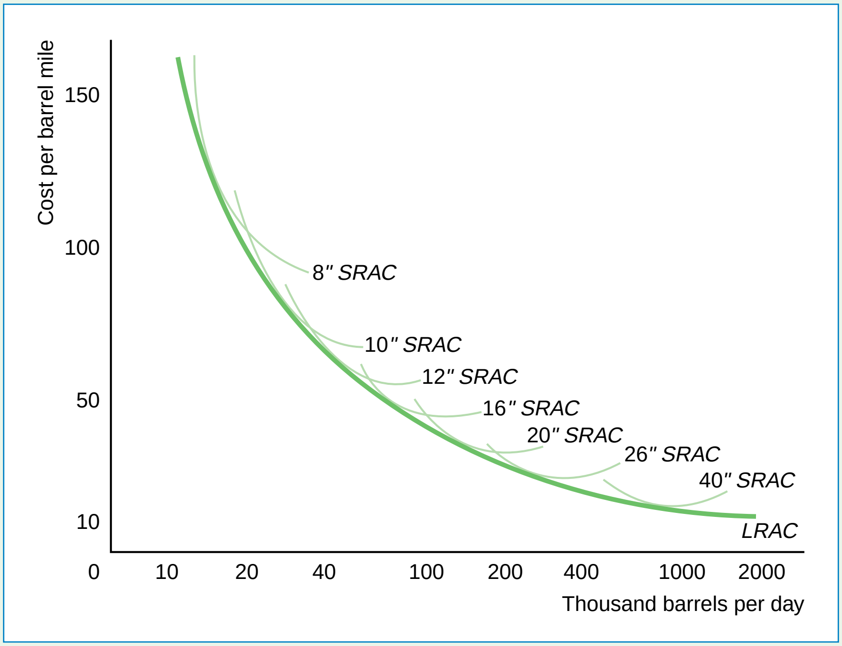
Printing Firm The first graph shows the relationship between short-run and long-run average cost curves for the Norwegian printing firm. Because this production function has constant returns to scale, doubling both inputs doubles output, so the long-run average cost, *LRAC*, is constant at 20 kr, as we saw earlier. If capital is fixed at 200 units, the firm's short-run average cost curve is $SRAC^1$. If the firm produces 200 units of output, its short-run and long-run average costs are equal. At any other output, its short-run cost is higher than its long-run cost.

The short-run marginal cost curves, $SRMC^1$ and $SRMC^2$, are upward sloping and equal the corresponding U-shaped short-run average cost curves, $SRAC^1$ and $SRAC^2$, only at their minimum points, 20 kr. In contrast, because the long-run average cost is horizontal at 20 kr, the long-run marginal cost curve, *SRMC*, is horizontal at 20 kr. Thus the long-run marginal cost curve is *not* the envelope of the short-run marginal cost curves.



Oil Pipelines Oil companies use the information in the second graph[7] to choose what size pipe to use to deliver oil. The 8" *SRAC* is the short-run average cost of a pipe with an 8-inch diameter. The long-run average cost curve, *LRAC*, is the envelope of all possible short-run average cost curves. It is more expensive to lay larger pipes than smaller ones, so a firm does not want to install unnecessarily large pipes. The average cost of sending a substantial quantity through a single large pipe is lower than that of sending it through two smaller pipes. For example, the average cost per barrel of sending 200,000 barrels per day through two 16-inch pipes is 1.67 (= \$50/\$30) greater than through a single 26-inch pipe.

Because the company incurs large fixed costs in laying miles and miles of pipelines and because pipes last for years, it does not vary the size of pipes in the short run. In the long run, the oil company installs the ideal pipe size to handle its "throughput" of oil. As Exxon (1975, p. 16) notes, several oil companies share interstate pipelines because of the large economies of scale.

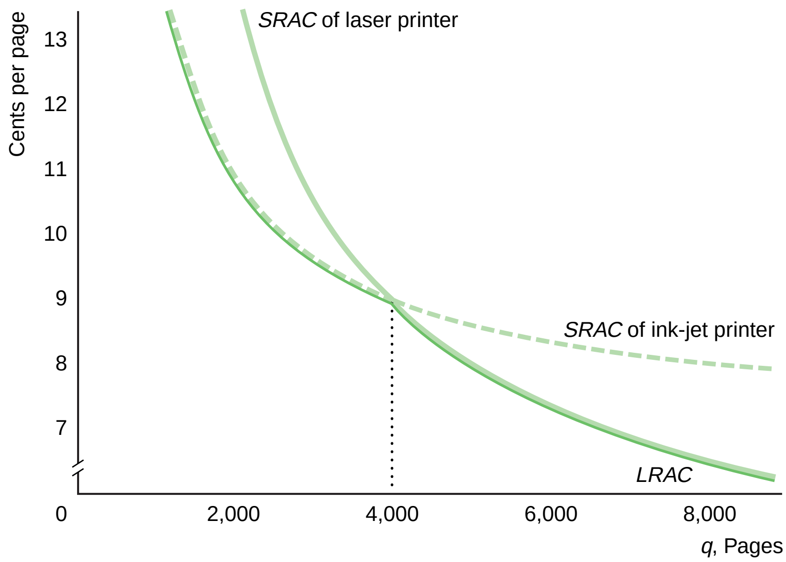


[7]Exxon Company, U.S.A., *Competition in the Petroleum Industry*, 1975, p. 30. Reprinted with permission.

Application

CHOOSING AN INK-JET OR A LASER PRINTER

You decide to buy a printer for your college assignments. You need to print in black and white. In 2002, you can buy a personal laser printer for \$200 or an ink-jet printer for \$80 that prints 10 pages a minute at the same density (1,200 dots per inch).



If you buy the ink jet, you save \$120 right off the bat. The laser costs less per page to operate, however. The cost of ink and paper is about 4¢ per page for a laser compared to about 7¢ per page for an ink jet. That means that the average cost per page of operating a laser (\$200/q + 0.04, where q is the number of pages) is less than that of an ink jet (\$80/q + 0.07) after q reaches about 4,000 pages.

The graph shows the short-run average cost curves for the laser printer and the ink-jet printer. The lower-cost choice is the ink jet if you're printing fewer than 4,000 pages and the laser if you're printing more.

So should you buy the laser printer? If you print more than 4,000 pages over its lifetime, the laser is less expensive to own and operate than the ink jet. If the printers last two years and you print 39 or more pages per week, then the laser printer is cost effective.

Short-Run and Long-Run Expansion Paths

Long-run cost is lower than short-run cost because the firm has more flexibility in the long run. To show the advantage of flexibility, we can compare the short-run and long-run expansion paths, which correspond to the short-run and long-run cost curves.

The Norwegian printing firm has greater long-run flexibility. The tangency of the firm's isoquants and isocost lines determines the long-run expansion path in Figure 7.10. The firm expands output by increasing both its labor and its capital, so its long-run expansion path is upward sloping. To increase its output from 100 to 200 units (move from *x* to *z*), it doubles its capital from 100 to 200 units and its labor from 50 to 100 workers. Its cost increases from 2,000 kr to 4,000 kr.

In the short run, the firm cannot increase its capital, which is fixed at 100 units. The firm can increase its output only by using more labor, so its short-run expansion

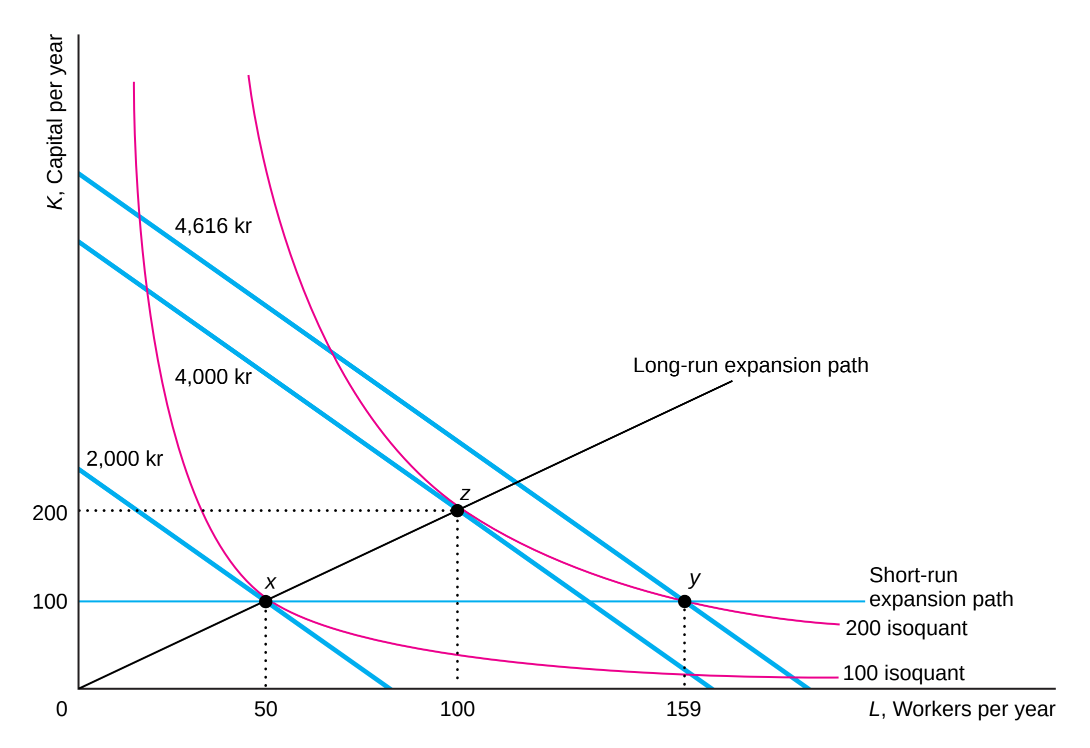


Figure 7.10 Long-Run and Short-Run Expansion Paths. In the long run, the Norwegian printing firm increases its output by using more of both inputs, so its long-run expansion path is upward sloping. In the short run, the firm cannot vary its capital, so its short-run expansion path is horizontal at the fixed level of output. That is, it increases its output by increasing the amount of labor it uses. Expanding output from 100 to 200 raises the printing firm's long-run cost from 2,000 kr to 4,000 kr but raises its short-run cost from 2,000 kr to 4,616 kr.

path is horizontal at $K = 100$. To expand its output from 100 to 200 units (move from x to y), the firm must increase its labor from 50 to 159 workers, and its cost rises from 2,000 kr to 4,616 kr. Doubling output increases long-run cost by a factor of 2 and short-run cost by approximately 2.3.

How Learning by Doing Lowers Costs

Two reasons why long-run cost is lower than short-run cost are that firms have more flexibility in the long run and that technical progress (Chapter 6) may lower cost over time. A third reason is **learning by doing**: the productive skills and knowledge of better ways to produce that workers and managers gain from experience.

In some firms, learning by doing is a function of the time since the product was introduced. In others, learning by doing is a function of *cumulative output*: the total number of units of output produced since the product was introduced. Learning is connected to cumulative output if workers become increasingly adept the more times they perform a task. As a consequence, workers become more productive if they make many units over a short period than if they produce a few units over a longer period. For example, the average labor cost of producing a C-141 plane (panel a of Figure 7.11) fell with cumulative output (based on Womer and Patterson, 1983).

If a firm is operating in the economies of scale section of its average cost curve, expanding output lowers its cost for two reasons. Its average cost falls today because of economies of scale, and for any given level of output, its average cost is lower in the next period due to learning by doing.

In panel b of Figure 7.11, the firm is currently producing q_1 units of output at point A on average cost curve AC^1. If it expands its output to q_2, its average cost falls in this period to B because of economies of scale. The learning by doing in this period results in a lower average cost, AC^2, in the next period. If the firm continues to produce q_2 units of output in the next period, its average cost falls to b on AC^2.

If instead of expanding output to q_2 in this period, the firm expands to q_3, its average cost is even lower in this period (C on AC^1) due to even more economies of scale. Moreover, its average cost in the next period is even lower, AC^3, due to the extra experience in this period. If the firm continues to produce q_3 in the next period, its average cost is c on AC^3. Thus all else the same, if learning by doing depends on cumulative output, firms have an incentive to produce more in the short run than they otherwise would to lower their costs in the future.

LEARNING BY DOING IN COMPUTER CHIPS

The cost of producing a computer memory chip falls substantially due to learning by doing. There are several different types of MOS (metal oxide on silicon) memory chips: EPROM (erasable programmable read-only memory), DRAM (dynamic random-access memory), and fast SRAM (static random-access memory). From one generation to another, EPROM's storage capacity doubles, whereas the storage capacity of DRAM and SRAM increases by a

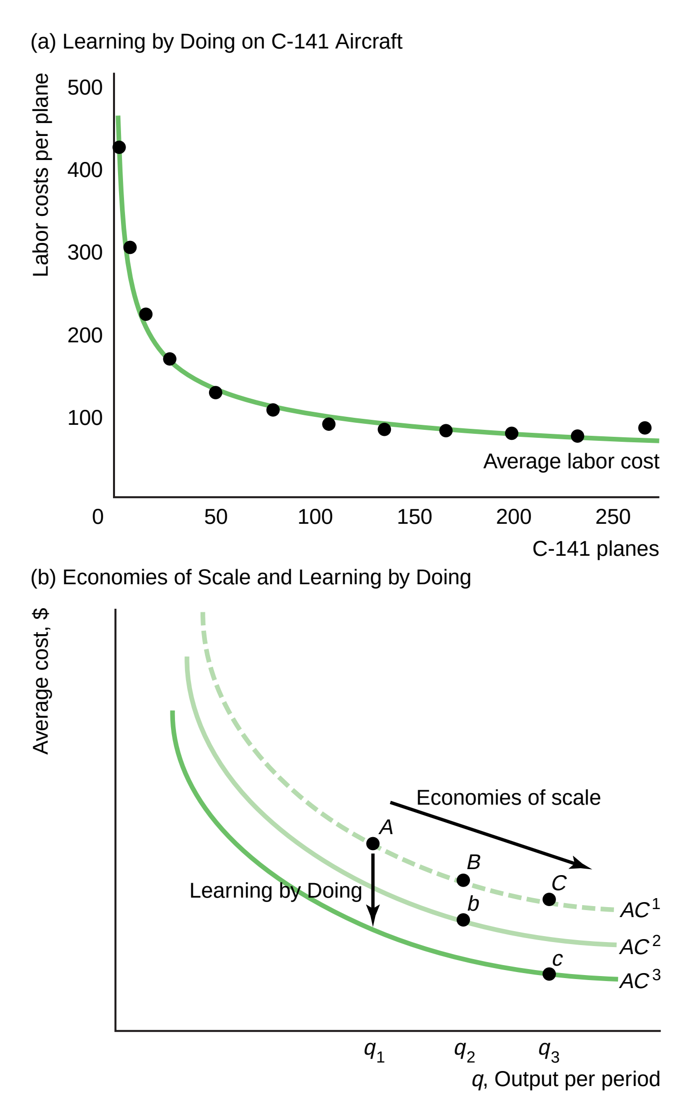


Figure 7.11 Learning by Doing. (a) As more C-141 aircraft were produced, the average labor cost per plane fell (Womer and Patterson, 1983). The horizontal axis shows the cumulative number of planes produced over time. (b) In the short run, extra production reduces a firm's average cost owing to economies of scale: because $q_1 < q_2 < q_3$, A is higher than B, which is higher than C. In the long run, extra production reduces average cost because of learning by doing. To produce q_2 this period costs B on AC^1, but to produce that same output in the next period would cost only b on AC^2. If the firm produces q_3 instead of q_2 in this period, its average cost in the next period is AC^3 instead of AC^2 because of additional learning by doing. Thus extra output in this period lowers the firm's cost in two ways: It lowers average cost in this period due to economies of scale and lowers average cost for any given output level in the next period due to learning by doing.

factor of 4. Like clockwork, a new EPROM generation appears every 18 months; a new DRAM, every three years.

Gruber (1992) finds that the average cost of EPROM chips falls with cumulative output but does not decrease over time or with the scale of production. With each doubling in the cumulative output of an EPROM chip, its average cost falls by 22%. This effect may encourage firms to produce more EPROM chips in the first few months after a new generation is introduced than they would without learning. By doing so, the firm gains experience more rapidly, which causes its average cost to fall more rapidly.

Irwin and Klenow (1994) find an average of 20% learning curve effect on cumulative output for DRAMs. Chung (2001) reports 17% learning for 64K DRAM and 9% for 256K DRAM in Korea.

Although the type and speed of learning by doing vary across chips, they are the same across generations of the same chip. Thus firms know that they can count on their costs falling and build these predictable cost reductions into their planning over time.

7.5 COST OF PRODUCING MULTIPLE GOODS

Few firms produce only a single good. We discuss single-output firms for simplicity. If a firm produces two or more goods, the cost of one good may depend on the output level of the other.

Outputs are linked if a single input is used to produce both of them. For example, mutton and wool both come from sheep, cattle provide beef and hides, and oil supplies both heating fuel and gasoline. It is less expensive to produce beef and hides together than separately. If the goods are produced together, a single steer yields one unit of beef and one hide. If beef and hides are produced separately (throwing away the unused good), the same amount of output requires two steers and more labor.

We say that there are **economies of scope** if it is less expensive to produce goods jointly than separately (Panzar and Willig, 1977, 1981). A measure of the degree to which there are economies of *s*cope (*SC*) is

$$SC = \frac{C(q_1, 0) + C(0, q_2) - C(q_1, q_2)}{C(q_1, q_2)},$$

where $C(q_1, 0)$ is the cost of producing q_1 units of the first good by itself, $C(0, q_2)$ is the cost of producing q_2 units of the second good, and $C(q_1, q_2)$ is the cost of producing both goods together. If the cost of producing the two goods separately, $C(q_1, 0) + C(0, q_2)$, is the same as producing them together, $C(q_1, q_2)$, then SC is zero. If it is cheaper to produce the goods jointly, SC is positive. If SC is negative, there are diseconomies of scope, and the two goods should be produced separately.

To illustrate this idea, suppose that Laura spends one day collecting mushrooms and wild strawberries in the woods. Her **production possibility frontier**—the maximum amounts of outputs (mushrooms and strawberries) that can be produced from a fixed amount of input (Laura's effort during one day)—is PPF^1 in Figure 7.12. The production possibility frontier summarizes the trade-off Laura faces: She picks fewer mushrooms if she collects more strawberries in a day.

If Laura spends all day collecting only mushrooms, she picks 8 pints; if she spends all day picking strawberries, she collects 6 pints. If she picks some of each, however, she can harvest more total pints: 6 pints of mushrooms and 4 pints of strawberries. The product possibility frontier is concave (the middle of the curve is farther from the origin than it would be if it were a straight line) because of the diminishing marginal returns to collecting only one of the two goods. If she collects only mushrooms, she must walk past wild strawberries without picking them. As a result, she

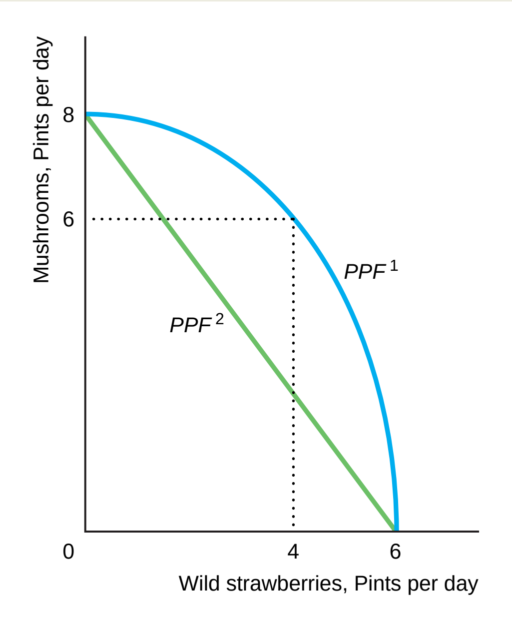


Figure 7.12 Joint Production. If there are economies of scope, the production possibility frontier is bowed away from the origin, PPF^1. If instead the production possibility frontier is a straight line, PPF^2, the cost of producing both goods does not fall if they are produced together.

has to walk farther if she collects only mushrooms than if she picks both. Thus there are economies of scope in jointly collecting mushrooms and strawberries.

If instead the production possibility frontier were a straight line, the cost of producing the two goods jointly would not be lower. Suppose, for example, that mushrooms grow in one section of the woods and strawberries in another section. In that case, Laura can collect only mushrooms without passing any strawberries. That production possibility frontier is a straight line, PPF^2 in Figure 7.12. By allocating her time between the two sections of the woods, Laura can collect any combination of mushrooms and strawberries by spending part of her day in one section of the woods and part in the other.

A number of empirical studies show that some processes have economies of scope, others have none, and some have diseconomies of scope. Shoesmith (1988) found that there are economies of scope in refining. It is less expensive to produce gasoline, distillate fuels, and other refined products together than separately. Akridge and Hertel (1986) observed large economies of scope in producing various fertilizers together.

Friedlaender, Winston, and Wang (1983) found that for American automobile manufacturers, it is 25% less expensive (SC = 0.25) to produce large cars together with small cars and trucks than to produce large cars separately and small cars and trucks together. However, there are no economies of scope from producing trucks together with small and large cars. Producing trucks separately from cars is efficient.

Kim (1987) found substantial diseconomies of scope in using railroads to transport freight and passengers together. It is 41% less expensive (SC = –0.41) to transport passengers and freight separately than together. In the early 1970s, passenger service in the United States was transferred from the private railroad companies to Amtrak, and the services are now separate. Kim's estimates suggest that this separation is cost effective.

DEAD END

Finally, dead people are pulling their weight—by providing "fuel" for heating thousands of homes in Sweden. The ovens of two high-tech crematoriums send power to local energy companies. The firms benefit from economies of scope because the costs of cremating and of producing energy are lower if the two activities are combined.

Summary

From all technologically efficient production processes, a firm chooses the one that is economically efficient. The economically efficient production process is the technologically efficient process for which the cost of producing a given quantity of output is lowest, or the one that produces the most output for a given cost.

1. **Measuring costs:** The economic or opportunity

cost of a good is the value of its next best alternative use. Economic cost includes both explicit and implicit costs.

2. **Short-run costs:** In the short run, the firm can vary the costs of the factors that it can adjust, but the costs of other factors are fixed. The firm's average fixed cost falls as its output rises. If a firm has a short-run average cost curve that is U-shaped, its marginal cost curve is below the average cost curve when average cost is falling and above average cost when it is rising, so the marginal cost curve cuts the average cost curve at its minimum.
3. **Long-run costs:** In the long run, all factors can be varied, so all costs are variable. As a result, average cost and average variable cost are identical. The firm chooses the combination of inputs it uses to minimize its cost. To produce a given output level, it chooses the lowest isocost line that touches the relevant isoquant, which is tangent to the isoquant. Equivalently, to minimize cost, the firm adjusts inputs until the last dollar spent on any input increases output by as much as the last dollar spent on any other input. If the firm calculates the cost of producing every possible output level given current input prices, it knows its cost function: Cost is a function of the input prices and the output level. If the firm's average cost falls as output expands, it has economies of scale. If its average cost rises as output expands, there are diseconomies of scale.
4. **Lower costs in the long run:** The firm can always do in the long run what it does in the short run, so its long-run cost can never be greater than its short-run cost. Because some factors are fixed in the short run, to expand output, the firm must greatly increase its use of other factors, which is relatively costly. In the long run, the firm can adjust all factors, a process that keeps its cost down. Long-run cost may also be lower than short-run cost if there is technological progress or learning by doing.
5. **Cost of producing multiple goods:** If it is less expensive for a firm to produce two goods jointly rather than separately, there are economies of scope. If there are diseconomies of scope, it is less expensive to produce the goods separately.

QUIZ

Questions

1. "There are certain fixed costs when you own a plane," [Andre] Agassi explained last week during a break in the action at the Volvo/San Francisco tennis tournament, "so the more you fly it, the more economic sense it makes. . . . The first flight after I bought it, I took some friends to Palm Springs for lunch."[8] Discuss Agassi's statement.
2. The only variable input a janitorial service firm uses to clean offices is workers who are paid a wage, w, of \$8 an hour. Each worker can clean four offices in an hour. Use math to determine the variable cost, the average variable cost, and the marginal cost of cleaning one more office. Draw a diagram like Figure 7.1 to show the variable cost, average variable cost, and marginal cost curves.
3. Using the information in Table 7.1, construct another table showing how a lump-sum franchise tax of \$30 affects the various average cost curves of the firm.
4. A firm builds shipping crates out of wood. How does the cost of producing a 1-cubic-foot crate (each side is 1 foot square) compare to the cost of building an 8-cubic-foot crate if wood costs \$1 a square foot and the firm has no labor or other costs? More generally, how does cost vary with volume?
5. You have 60 minutes to take an exam with two questions. You want to maximize your score. Toward the end of the exam, the more time you spend on either question, the fewer extra points per minute you get for that question. How should you allocate your time between the two questions? (*Hint*: Think about producing an output of a score on the exam using inputs of time spent on each of the problems. Then use Equation 7.6.)
6. Boxes of cereal are produced by using a fixed-proportion production function: One box and one unit (8 ounces) of cereal produce one box of cereal. What is the expansion path?

[8]Ostler, Scott, "Andre Even Flies like a Champ," *San Francisco Chronicle*, February 8, 1993, C1.

7. Suppose that your firm's production function has constant returns to scale. What is the long-run expansion path?
8. The production process of the firm you manage uses labor and capital services. How does the long-run expansion path change when the wage increases while the rental rate of capital stays constant?
9. A U-shaped long-run average cost curve is the envelope of U-shaped short-run average cost curves. On what part of the curve (downward sloping, flat, or upward sloping) does a short-run curve touch the long-run curve? (*Hint*: Your answer should depend on where on the long-run curve the two curves touch.)
10. Suppose that the government subsidizes the cost of workers by paying for 25% of the wage (the rate offered by the U.S. government in the late 1970s under the New Jobs Tax Credit program). What effect will this subsidy have on the firm's choice of labor and capital to produce a given level of output?
11. Suppose in Solved Problem 7.1 that the government charges the firm a franchise tax each year (instead of only once). Describe the effect of this tax on the marginal cost, average variable cost, short-run average cost, and long-run average cost curves.
12. What can you say about Laura's economies of scope if her time is valued at \$5 an hour and her production possibility frontier is PPF^1 in Figure 7.12?
13. *Review* (Chapter 6): What might cause the marginal product of capital to fall as output increases?

Problems

14. Give the formulas for and plot *AFC*, *MC*, *AVC*, and *AC* if the cost function is
    a. $C = 10 + 10q$
    b. $C = 10 + q^2$
    c. $C = 10 + 10q - 4q^2 + q^3$
15. What is the long-run cost function if the production function is $q = L + K$?
16. Gail works in a flower shop, where she produces 10 floral arrangements per hour. She is paid \$10 an hour for the first eight hours she works and \$15 an hour for each additional hour she works. What is the firm's cost function? What are its *AC*, *AVC*, and *MC* functions? Draw the *AC*, *AVC*, and *MC* curves.
17. A firm's cost curve is $C = F + 10q - bq^2 + q^3$, where $b > 0$.
    a. For what values of b are cost, average cost, and average variable cost positive? (From now on, assume that all these measures of cost are positive at every output level.)
    b. What is the shape of the *AC* curve? At what output level is the *AC* minimized?
    c. At what output levels does the *MC* curve cross the *AC* and the *AVC* curves?
    d. Use calculus to show that the *MC* curve must cross the *AVC* at its minimum point.
18. A firm has two plants that produce identical output. The cost functions are $C_1 = 10q - 4q^2 + q^3$ and $C_2 = 10q - 2q^2 + q^3$.
    a. At what output levels does the average cost curve of each plant reach its minimum?
    b. If the firm wants to produce 4 units of output, how much should it produce in each plant?
19. For a Cobb-Douglas production function, how does the expansion path change if the wage increases while the rental rate of capital stays the same? (*Hint*: See Appendix 7C.)
20. A firm has a Cobb-Douglas production function, $Q = AL^{\alpha}K^{\beta}$, where $\alpha + \beta < 1$. On the basis of this information, what properties does its cost function have?
21. A firm's average cost is $AC = \alpha q^{\beta}$, where $\alpha > 0$. How can you interpret α? (*Hint*: Suppose that $q = 1$.) What sign must β have if there is learning by doing? What happens to average cost as q gets large? Draw the average cost curve as a function of output for a particular set of α and β.