

Production and Costs

Suppose you want to download more music from the Internet. There are (at least) a couple of ways to do that: You can spend more time downloading, or you can get a faster connection.

If the problem is to download more music *by tonight*, you've got fewer options; a fast connection can take several days to install. In that case, you're just going to have to devote more hours to downloading, and the value of those hours measures the cost (to you) of getting the music files.

But if the problem is to download more music *over the next several months*, then you might want to consider a faster connection. In the long run, you've got more options than in the short run.

Business executives and managers face the same set of issues all the time. Suppose you own a dressmaking factory and you want to ramp up your hourly output. You can do that by having more workers on the premises, or you can do it by investing in more (or better) sewing machines that will make the workers more efficient. In the short run, you've got to go with more workers, because it takes a while to get new machines ordered, delivered, and installed. But in the long run, you'll probably want to go with some combination of more workers *and* more machines.

Your costs depend on two things: How many dresses do you make, and how do you make them? In Chapter 5, we saw that firms choose quantities by equating marginal cost to marginal revenue. But we didn't say very much about where the firm's cost curves come from in the first place. The answer to that question depends on the technology available to the firm. In this chapter we'll see how the firm, taking the available technology as given, chooses a production process, and how that production process determines the firm's costs.

6.1 Production and Costs in the Short Run

In the short run, the firm has limited options. A car manufacturer can't build a new factory overnight and a dressmaker has to wait to install new sewing machines. We'll abstract from this situation somewhat by assuming that in the short run, the only way

Total product (TP)

The quantity of output produced by the firm in a given amount of time.

Total product depends on the quantity of labor the firm hires.

Short-run production function

The function that associates to each quantity of labor its total product.

for a dressmaker to produce more dresses is to hire more labor. Our first task is to be more explicit about the relationship between the quantity of labor and the quantity of dresses. Our next task will be to explore that relationship to understand how the firms' cost curves are determined.

The Total, Marginal, and Average Products of Labor

We'll start with a numerical example, illustrated in Exhibit 6.1.

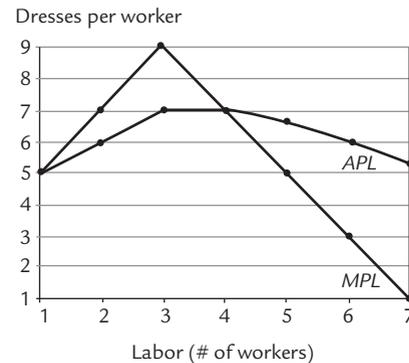
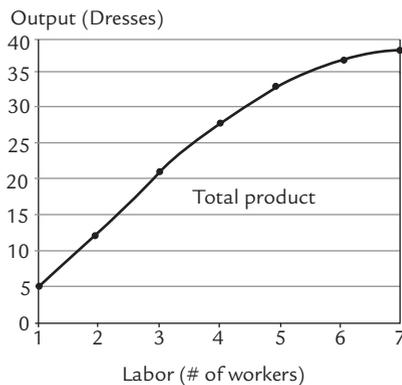
The first two columns relate the number of workers to the number of dresses the firm can produce in a given period of time (say an hour). The chart shows that 1 worker produces 5 dresses per hour; 2 workers produce 12 dresses, and so on. The number of dresses is called the **total product** (abbreviated **TP**) of labor.

The same information is recorded in the curve displayed underneath the first two columns in the exhibit. That curve is called the firm's **short-run production function**.

The short-run production function slopes upward because each additional worker contributes something to the production process. The number of dresses that each

EXHIBIT 6.1

Total, Marginal, and Average Products



Quantity of Labor	Total Product (TP)	Marginal Product of Labor (MPL)	Average Product of Labor (APL)
1 worker	5 dresses	5 dresses per worker	5 dresses per worker
2	12	7	6
3	21	9	7
4	28	7	7
5	33	5	6.6
6	36	3	6
7	37	1	5.3

Total product (TP) is the quantity of output (in this case, dresses) that a given number of workers can produce (in a prespecified amount of time). The marginal product of labor (MPL) is the additional output due to one additional worker, and the average product of labor is the total product divided by the number of workers. In this example, when there are fewer than 4 workers, the marginal product exceeds the average product, so the average product is rising. When there are more than 4 workers, marginal product is less than average product, so average product is falling.

worker adds is called the **marginal product of labor (MPL)**. In Exhibit 6.1, the sixth worker increases the total product by 3 dresses (from 33 to 36), so that worker's marginal product is 3 dresses. The marginal products of each worker are listed in the third column of the chart and plotted in the right-hand graph.

Exercise 6.1 Make sure that all the marginal products have been computed correctly.

The **average product of labor (APL)** is the number of dresses divided by the number of workers. Because the number of dresses is the same thing as the total product, we can write:

$$APL = TP/L$$

where L (which stands for labor) is the total number of workers employed. For example, when 4 workers produce 28 dresses, the average product of labor is 7 dresses per worker. In Exhibit 6.1, the average product of labor is computed in the last column of the chart and graphed in the right-hand graph.

Exercise 6.2 Check that all the average products have been computed correctly.

Total product is measured in *dresses* but marginal and average products are measured in *dresses per worker*. Thus the marginal and average products must be plotted on a separate graph from total product.

The total product and marginal product curves are related: Marginal product gives the *slope* of total product. For example, when the number of workers increases from 5 to 6 (an increase of 1), output increases from 33 to 36 (an increase of 3). The ratio $3/1$ is the *slope* of the total product curve near the point (6, 36), and $3 = 3/1$ is also the *height* of the marginal value curve at 6.

The Shape of the Average Product Curve

If 5 bakers can produce 500 cupcakes per day, how many cupcakes per day can 6 bakers produce? If this were an elementary school word problem, the answer would be 600. But in real life the answer might well be different. The sixth baker interacts with the first five bakers in ways that might make them all either more or less productive. For example, his presence might make it easier for all the bakers to specialize—one greases the pans while another mixes the batter and yet another prepares the frosting. In that case 6 bakers might produce *more* than 600 cupcakes. Or, the sixth baker might compete for counter space with the first five and get in their way; in that case, 6 bakers might produce *fewer* than 600 cupcakes.

There are plenty of examples in other industries as well. Two lumberjacks with a two-handed whipsaw can cut down a lot more than twice as many trees as either one could harvest individually, but a hundred lumberjacks might cut down far fewer than a hundred times as many trees, because there just aren't that many trees to cut down. A hundred auto workers are far more productive on average than a single auto worker because they can locate themselves at strategic points along an assembly line, whereas a single worker would have to run all over the factory performing a multitude of tasks. But a thousand auto workers in the same factory might be *less* productive on average,

Marginal product of labor (MPL)

The increase in total product due to hiring one additional worker (assuming that capital is held fixed).

Average product of labor (APL)

Total product divided by the number of workers.



Dangerous
Curve

for the simple reason that they crowd the factory and get in each others' way. An army of 10,000 is more than 10,000 times as powerful as an army of one, but even here the advantages of size are limited: The Roman poet Virgil tells us that his army was so crowded that many soldiers had no room to use their weapons.

The Shape of the Marginal Product Curve

Like the average product curve, and for similar reasons, the marginal product curve has the same general inverted U shape. The second baker contributes more than the first and the third contributes more than the second (so marginal product is increasing), but eventually additional bakers start getting in each others' way and marginal product begins to decrease.

In Exhibit 6.1, marginal product starts decreasing after the third worker comes on board. We say that three workers marks the **point of diminishing marginal returns** for this firm.

Point of diminishing marginal returns

The point after which the marginal product curve begins to decrease.

You can see the same phenomenon in Exhibit 6.2, where the curves are smoother and perhaps easier to look at. The point of diminishing marginal returns occurs when the firm has hired L_0 workers. After that point, as further workers are added, marginal product continues to fall. Eventually, after L_1 workers have been hired, average product begins falling also.

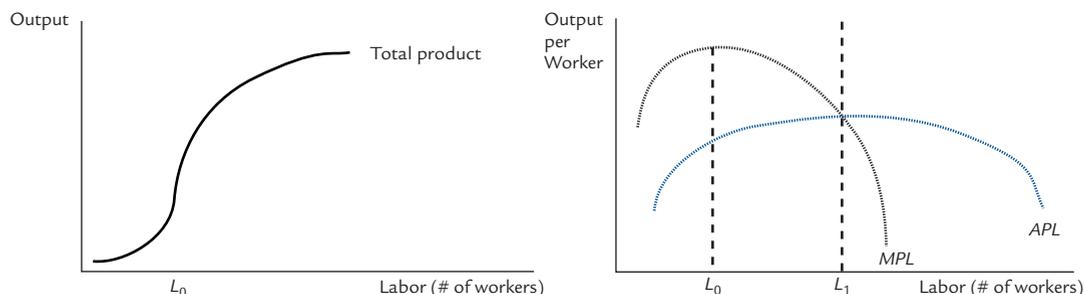
The Relationship Between the Average and Marginal Product Curves

Suppose your bakery employs 5 bakers to bake 500 cupcakes. The average product of labor is 100 cupcakes per baker.

Now you hire a sixth baker and output goes up to 630 cupcakes. The sixth baker's marginal product is 130 cupcakes, and hiring him raises the average product to $630/6 = 105$ cupcakes per baker.

EXHIBIT 6.2

The Stages of Production



When there are fewer than L_1 workers, marginal product (*MPL*) exceeds average product (*APL*) and average product is rising. When there are more than L_1 workers, marginal product is less than average product and average product is falling. Therefore, average product has the shape of an inverted U and marginal product cuts through average product at the top of the U. The marginal product curve also has an inverted U shape.

The sixth baker raises the average product (from 100 to 105) precisely because his marginal product (130) is above the existing average (100).

Whenever a worker's marginal product is greater than the average product, adding that worker causes average product to rise.

Now suppose you hire a seventh baker and output rises to 700 cupcakes. The seventh baker's marginal product (70) is below the existing average product of 105, so hiring that baker causes the average product to fall (from 105 to $700/7 = 100$).

Whenever a worker's marginal product is below the average product, adding that worker causes average product to fall.

You can see all this in Exhibit 6.2: Up to the point where L_1 workers are hired, each worker's marginal product exceeds the average product (i.e., the marginal product curve lies above the average cost curve). In this region, average cost is rising. After L_1 or more workers have been hired, each worker's marginal product is below the average product (i.e., the marginal product curve lies below the average cost curve). In this region, average cost is falling.

It follows that the marginal cost curve must *cross* the average cost curve right at the point where the average cost curve turns around, which is to say right at the top of the inverted U, as you can see in Exhibit 6.2.

Costs in the Short Run

Remember that firms have both *fixed costs* and *variable costs*. In the example of Exhibit 6.1, we've assumed that in the short run, the only thing the firm can vary is labor. Thus, *in the short run, the only variable cost is the cost of hiring labor*.

In a more realistic example, a dressmaker would have other variable costs, including the cost of buying fabric. In this example, we've implicitly assumed that fabric is free. Obviously the assumption is absurd, but fortunately it won't affect the lessons we'll draw from the example.

To figure out the firm's variable cost curve, you need to know the total product curve and the **wage rate**. Exhibit 6.3 shows the connection. The first two columns reproduce the total product curve from Exhibit 6.1, and we add the assumption that workers earn a wage rate of \$15 per hour. Then, to get the variable cost numbers, we multiply the number of workers by 15. This is done in the fourth column of Exhibit 6.3.

The variable cost curve (shown in the exhibit) relates the number of dresses (*not* the number of workers!) to this variable cost. Thus a quantity of 5 dresses (which can be produced by 1 worker) corresponds to a variable cost of \$15; a quantity of 12 dresses (which can be produced by 2 workers) corresponds to a variable cost of \$30, and so forth.

Exercise 6.3 Verify that the other points on the variable cost curve have been computed and plotted correctly.

To get the firm's total cost curve, we have to know its fixed costs and then add those fixed costs to the variable costs. Typically, the firm's short-run fixed costs are the costs



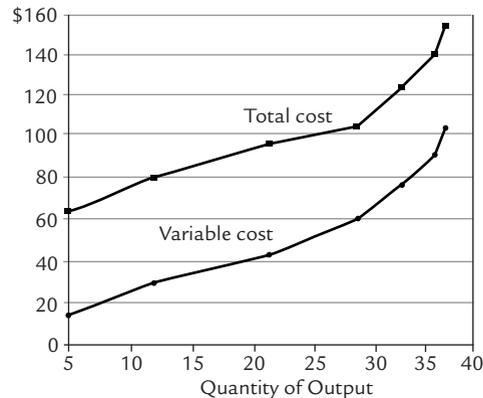
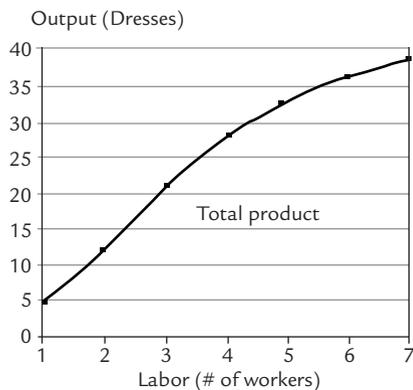
Dangerous
Curve

Wage rate

The price of hiring labor.

EXHIBIT 6.3

Variable Cost Curve



Quantity of Labor	Total Product (TP)	Quantity of Output	Variable Cost (VC)	Total Cost (TC)
1 worker	5 dresses	15	15	65
2	12	12	30	80
3	21	21	45	95
4	28	28	60	110
5	33	33	75	125
6	36	36	90	140
7	37	37	105	155

We take as given: the price of capital (\$10 per machine), the price of labor (\$15 per worker), the quantity of capital (5 machines), and the total product curve (shown on the left half of the exhibit). From this information, we compute points on the variable cost (VC) and total cost (TC) curves as follows: Given a quantity of output, use the total product curve to find the corresponding number of workers. Multiply by the wage rate (\$15 per worker) to get variable cost. Take variable cost and add fixed costs (in this case, 5 machines times \$10 per machine, or \$50) to get total cost.

Capital

Physical assets used as factors of production.

of **capital**, meaning the physical assets, such as machinery and factories, that are used in the production process. Examples of capital include a handyman's van, a secretary's computer, a professor's library, and a cowboy's lariat.

Because we've been talking about a dressmaker, let's assume that the relevant capital consists of sewing machines that can be rented for \$10 per hour. Let's also assume the firm has 5 sewing machines. Then the firm's fixed costs are \$50 per hour.



Dangerous Curve

The capital cost of \$10 per hour is the same for all firms, regardless of whether they own their own sewing machines. If Connie Daran's dress shop rents machines, Connie pays \$10 an hour for them. If Lauren Ralph's dress shop uses its own machines, then Lauren is forgoing the opportunity to rent those machines to Connie, making her opportunity cost \$10 per hour per machine.

To get the total cost numbers in Exhibit 6.3, we just take the variable cost numbers and add \$50. The resulting total cost curve lies exactly \$50 above the variable cost curve.

Computing Average Costs

The firm's **average variable cost (AVC)** is defined by the formula:

$$AVC = VC/Q$$

where *VC* is variable cost and *Q* is the quantity of output (the firm's total product). The firm's **average cost (AC)** is defined by the formula:

$$AC = TC/Q$$

where *TC* is total cost. Average cost is sometimes called **average total cost**.

In Exhibit 6.4, we compute *AVC* and *AC* for the same firm we studied in Exhibits 6.1 and 6.3. The left half of Exhibit 6.4 reproduces information on total, average, and marginal products from Exhibit 6.1. On the right side, the chart reproduces the Variable Cost and Total Cost columns from Exhibit 6.3. Average variable cost and average cost are computed directly from those columns. For example, at 5 dresses we have:

Average variable cost (AVC)

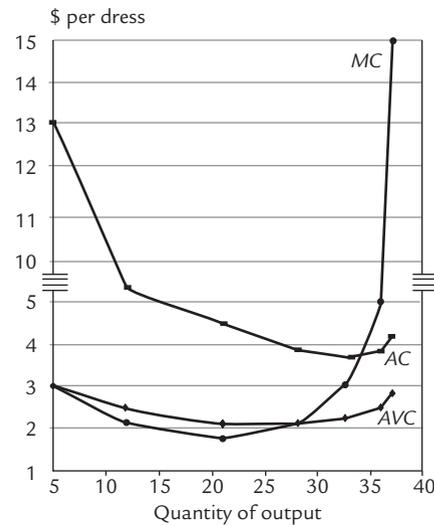
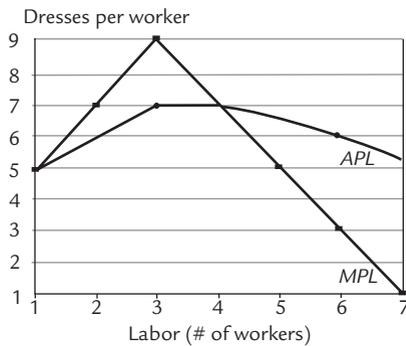
Variable cost divided by the quantity of output.

Average cost, or average total cost (AC)

Total cost divided by the quantity of output.

EXHIBIT 6.4

Deriving the Average and Marginal Cost Curves



Quantity of Labor	Total Product	Marginal Product of Labor (MPL)	Average Product of Labor (APL)	Quantity (Q)	Variable Cost (VC)	Total Cost (TC)	Average Variable Cost (AVC)	Average Cost (AC)	Marginal Cost (MC)
1 worker	5 dresses	5 dresses per worker	5 dresses per worker	5	\$15	\$65	\$3 per dress	\$13 per dress	\$3 per dress
2	12	7	6	12	30	80	2.50	6.67	2.14
3	21	9	7	21	45	95	2.14	4.52	1.67
4	28	7	7	28	60	110	2.14	3.93	2.14
5	33	5	6.6	33	75	125	2.27	3.79	3.00
6	36	3	6	36	90	140	2.50	3.89	5.00
7	37	1	5.3	37	105	155	2.84	4.19	15.00

The product curves on the left are taken from Exhibit 6.1. On the right, the variable cost and total cost data are taken from Exhibit 6.3. We compute *AVC*, *AC*, and *MC* from their definitions; namely, $AVC = VC/Q$ and $AC = TC/Q$. It turns out that we can also write $AVC = P_L/APL$. To compute *MC*, we use the formula $MC = P_L/MPL$.

$$AVC = VC/Q = \$15/5 = \$3 \text{ per dress}$$

and

$$AC = TC/Q = \$65/5 = \$13 \text{ per dress.}$$

All of the AVC and AC numbers are recorded on the curves below the chart.

Exercise 6.4 In Exhibit 6.4, verify that all the numbers in the AVC and AC columns have been computed correctly.

When labor is the only variable factor (as we have been assuming), there is another formula for average variable cost. Notice first that if the firm hires L workers, then its variable costs come to $P_L \cdot L$, where P_L is the wage rate of labor. Therefore,

$$AVC = \frac{VC}{Q} = \frac{(P_L \cdot L)}{Q} = \frac{P_L}{(Q/L)} = \frac{P_L}{APL}$$

or, more briefly,

$$AVC = P_L / APL$$

where $APL = Q/L$ is the average product of labor.

Exercise 6.5 Verify that $AVC = P_L/APL$ in every row of the charts in Exhibit 6.4. (Keep in mind that in this example, $P_L = \$15$.)

The Marginal Cost Curve

Now we want to construct the firm's marginal cost curve. Recall from Chapter 5 that marginal cost is the additional cost attributable to the last unit of output produced.

Thus, for example, we see in Exhibit 6.4 that the total cost of producing 36 dresses is \$140 and the total cost of producing 37 dresses is \$155. The difference, \$15 per dress, is the marginal cost when 37 dresses are produced. We have recorded the result of that calculation in the Marginal Cost column across from the quantity 37.

But how can we get the other numbers in the Marginal Cost column? For example, how can we compute marginal cost when the firm produces 33 dresses? In principle, we need to take the total cost of producing 33 dresses—which, according to the chart, is \$125—and subtract the cost of producing 32 dresses. Unfortunately, that information is missing from our incomplete chart, which lists only the quantities 5, 12, 21, 28, 33, 36, and 37.

But fortunately, there is another way to compute marginal cost. Here's the trick: First, use the total and marginal product curves to determine that when the total product is 33 dresses, the marginal product is 5 additional dresses per additional worker. Second, notice that "5 additional dresses per additional worker" is the same thing as $\frac{1}{5}$ additional workers per additional dress." So the marginal cost of producing an additional dress is equal to the cost of hiring $\frac{1}{5}$ of a worker. At the assumed going wage rate of \$15 per worker, that comes to \$3. So we record \$3 as the marginal cost of producing 33 units of output.

You might object that there is no such thing as $\frac{1}{5}$ of a worker. But don't forget that everything in our charts is implicitly measured "per hour." That makes it easy to hire $\frac{1}{5}$ of a worker—you hire someone to work 12 minutes out of every hour, or one day out of every five.



Dangerous
Curve

Similarly, we compute the marginal cost at a quantity of, say, 12: The marginal product of labor is now 7 dresses per worker, so it takes $\frac{1}{7}$ of a worker to produce an additional dress. Therefore, the marginal cost is $15 \times \frac{1}{7}$, or about \$2.14.

This method of calculating marginal costs can be summed up in a simple formula:

$$MC = P_L \times \frac{1}{MPL}$$

or

$$MC = \frac{P_L}{MPL}$$

Exercise 6.6 Check that all of the marginal cost numbers in Exhibit 6.4 have been derived correctly.

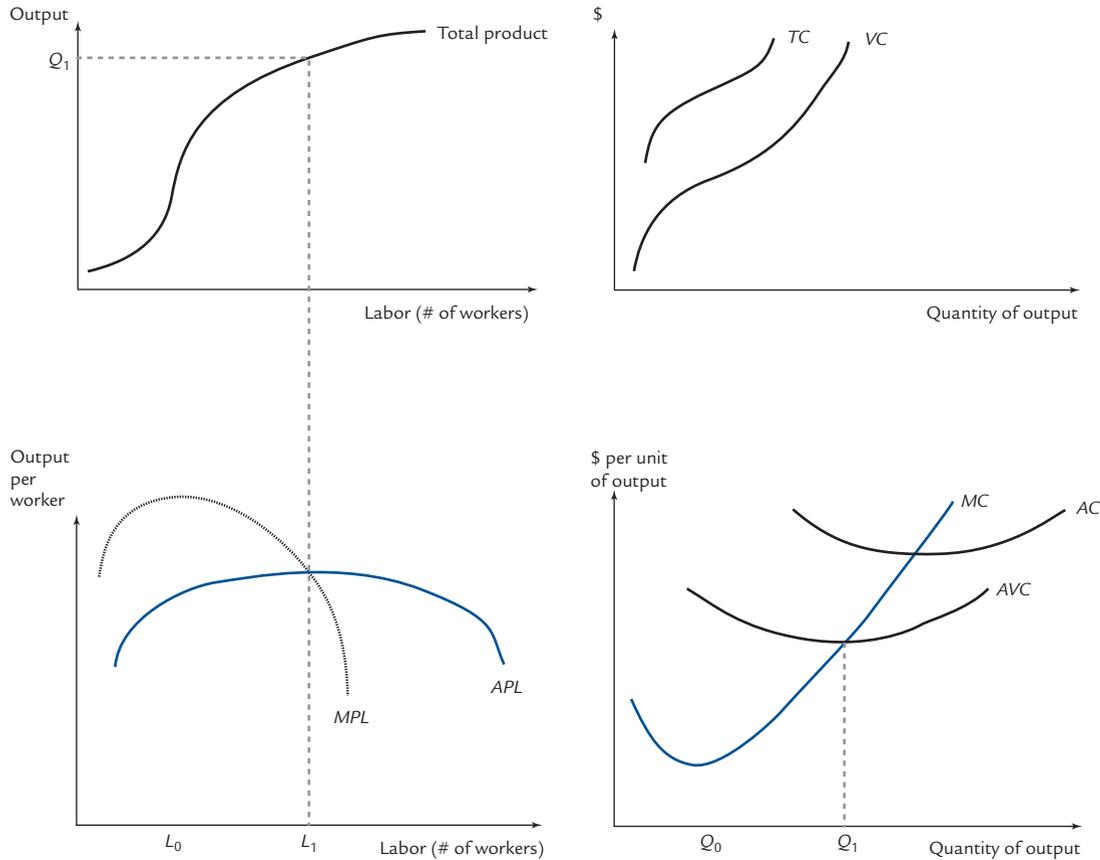
The Shapes of the Cost Curves

The right half of Exhibit 6.5 shows the shapes of the cost curves at a typical firm. The left half of the exhibit reproduces the product curves from Exhibit 6.2 for comparison. Here are the key facts about the geometry of the cost curves:

1. The variable cost (VC) curve is always increasing, because more output requires more labor and hence higher costs.
2. The total cost (TC) curve is determined by the formula $TC = FC + VC$, where FC (fixed cost) is constant. Therefore, it has exactly the same shape as the VC curve.
3. The marginal cost (MC) curve is U-shaped.
4. The average cost (AC) and average variable cost (AVC) curves are also U-shaped.
5. When marginal cost is below average variable cost, average variable cost is falling. In Exhibit 6.5, this refers to the region to the left of Q_1 . To see why, consider a situation where you've already produced, say, 10 items at an average variable cost of \$12 apiece. If the 11th item has a marginal cost below \$12 (i.e., if MC is below AVC), then it will lower the average variable cost below \$12 (i.e., average cost falls as the quantity increases from 10 to 11).
6. When marginal cost is above average variable cost, average variable cost is rising. In Exhibit 6.5, this occurs in the region to the right of Q_1 .
7. Marginal cost crosses average variable cost at the bottom of the average variable cost "U." This is a geometric consequence of points 5 and 6. When marginal cost is just equal to average variable cost, average variable cost is just changing from falling to rising.
8. The analogs of points 5, 6, and 7 hold when average variable cost is replaced by average cost, and they hold for the same reasons. Thus, when marginal cost is below average cost, average cost is falling; when marginal cost is above average cost, average cost is rising; marginal cost crosses average cost at the bottom of the average cost U.
9. The shapes of the cost curves are related to the shapes of the product curves. For example, we have $AVC = P_L/APL$ and $MC = P_L/MPL$, where P_L (the wage rate of labor) is a constant. These formulas convert the inverted U shapes of APL and MPL to the U shapes of AVC and MC .

EXHIBIT 6.5

The Geometry of Product Curves and Cost Curves



The product curves on the left are reproduced from Exhibit 6.2. Up to this point when there are L_1 workers and Q_1 units of output, marginal product exceeds average product, average product rises, marginal cost is below average variable cost, and average variable cost falls. Thereafter, marginal product is below average product, average product falls, marginal cost is above average variable cost, and average variable cost rises. Marginal cost cuts through both average variable cost and average cost at the bottom of the respective Us.



Dangerous
Curve

In drawing the cost curves, remember that TC and VC belong on a graph whose vertical axis shows “dollars,” while AVC , AC , and MC belong on a graph whose vertical axis shows “dollars per unit of output.” Remember, also, that *all* of these curves have an implicit unit of time built into them; thus, when we say that it takes 2 workers to produce 6 units of output, we really mean that it takes 2 workers to produce 6 units of output *in a given, prespecified period of time*.

6.2 Production and Costs in the Long Run

Typically, there are many ways to produce a unit of output. What can be done by 3 workers with 5 machines can perhaps also be done by 6 workers with only 1 machine. In the long run, the firm can adjust its employment of both labor and capital so as to achieve the least expensive method of producing a given quantity of output. Our first task will be to develop some geometry to help clarify the firm's considerations.

Isoquants

Exhibit 6.6 shows the set of all combinations that suffice to produce one unit of a certain good, which we will call X , in a given period of time. The vertical axis, labeled K , represents capital, and the horizontal axis, labeled L , represents labor. (K is traditionally used instead of C for *capital* in order to avoid any possible confusion with *consumption*.) The period of time is implicitly fixed; for example, we might be speaking of producing one unit of X per day. Appropriate units for labor and capital are, for example, “man-hours per day” and “machine-hours per day.”

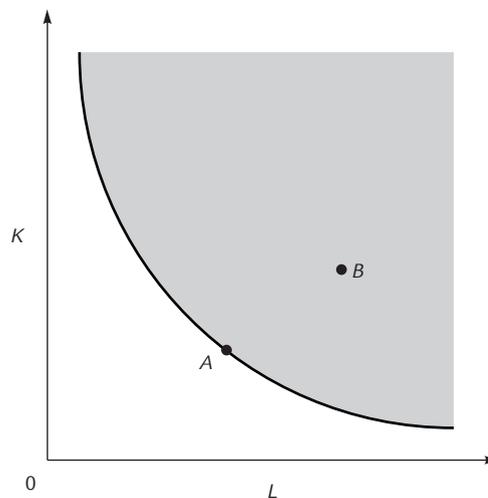
In Exhibit 6.6 every basket of inputs in the shaded part of the graph suffices to produce a unit of X . However, points that are off the boundary (like B) are **technologically inefficient**, in that there are other baskets of inputs, containing both less capital and less labor, that will also suffice to produce a unit of X . (For example,

Technologically inefficient

A production process that uses more inputs than necessary to produce a given output.

EXHIBIT 6.6

The Unit Isoquant



The shaded region represents all of the different baskets of capital and labor that can be used to produce one unit of X . Baskets that are off the boundary, like B , are technologically inefficient, in that a unit of X can be produced by a different basket (like A) containing smaller quantities of both inputs. The technologically efficient baskets for producing a unit of X are those on the unit isoquant, which is the heavy curve that bounds the shaded region.

basket A contains smaller quantities of both inputs than basket B does.) No firm would want to produce a unit of X using a technologically inefficient basket of inputs. Thus, we will ignore these baskets and concentrate on the technologically efficient ones. In Exhibit 6.6 the technologically efficient baskets for producing a unit of X are represented by the heavy curve that bounds the shaded region. That curve is called the **unit isoquant**.

Unit isoquant

The set of all technically efficient ways to produce one unit of output.

Why is the unit isoquant shaped as it is? Note first that no point to the northeast of A can be on the unit isoquant, because any such point (like B) is technologically inefficient. For the same reason, no point to the northeast of *any* point on the unit isoquant can also lie on the unit isoquant. It follows that the points on the isoquant must all be to either the northwest or the southeast of each other. Another way to say this is

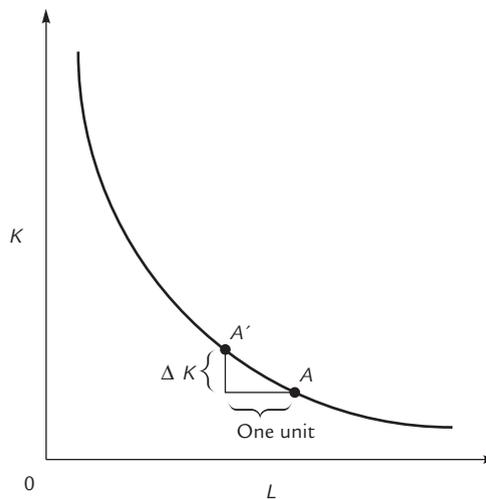
The unit isoquant is downward sloping.

The Marginal Rate of Technical Substitution

Suppose that each day a firm uses the basket of inputs A to produce one unit of X . One day an employee calls in sick, making it necessary to get by with one less unit of labor. How much additional capital will the firm need in order to maintain the daily output level? The answer is shown in Exhibit 6.7. Reducing labor input by one unit corresponds geometrically to moving one unit to the left; maintaining the output level corresponds geometrically to staying on the isoquant. Taken together, these requirements mandate that the firm move to point A' . The vertical distance between A and A' is the additional capital that must be added to the usual daily ration. That vertical distance has been labeled ΔK in Exhibit 6.7.

EXHIBIT 6.7

The Marginal Rate of Technical Substitution



The firm produces one unit of X per day using basket A of inputs. When labor input is reduced by one unit, capital input must be increased by ΔK units in order for the firm to remain on the isoquant and maintain its level of output. The number ΔK is the marginal rate of technical substitution of labor for capital.

For all practical purposes, the distance ΔK is equal to the slope of the isoquant at the point A .¹ The absolute value of this slope is called the **marginal rate of technical substitution of labor for capital** ($MRTS_{LK}$); it is the amount of capital necessary to replace one unit of labor while maintaining a constant level of output.²

Suppose that a construction firm produces 1 house per day by employing 100 carpenters and 10 power tools. Then it is reasonable to think that when a carpenter calls in sick, the firm can maintain its level of production through a small increase in power tool usage. On the other hand, if the same firm produces the same 1 house per day by employing 10 carpenters and 100 power tools, we expect it to need a much larger increase in tool usage to compensate for the same absent carpenter. In other words, when much labor and little capital are employed to produce a unit of output, $MRTS_{LK}$ is small, but when little labor and much capital are employed to produce the same unit of output, $MRTS_{LK}$ is large. Geometrically, this means that at points far to the southeast, the isoquant is shallow, while at points far to the northwest, it is steep. That is, the isoquant is convex.

Marginal rate of technical substitution of labor for capital ($MRTS_{LK}$)

The amount of capital that can be substituted for one unit of labor, holding output constant.

Marginal Products and the $MRTS$

The marginal products of labor and capital are related to the marginal rate of technical substitution. Suppose labor input is reduced by one unit and capital input is increased by ΔK units, where ΔK is just enough to maintain the existing level of output. Then $\Delta K = MRTS_{LK}$

Consider the two steps in this experiment separately. When one unit of labor is sacrificed, output goes down by the marginal product of labor, MPL . When ΔK units of capital are hired, output goes up by $\Delta K \cdot MPK$, where MPK is the marginal product of capital. Because the existing level of output does not change, we must have

$$MPL = \Delta K \cdot MPK = MRTS_{LK} \cdot MPK$$

or

$$MRTS_{LK} = MPL / MPK$$

Thus, the marginal rate of technical substitution is closely related to the marginal products of labor and capital. Keep in mind the conceptual distinction, though: To measure $MRTS_{LK}$, we hold *output* fixed, vary L by one unit, and ask how much K must vary. To measure MPL , we hold *capital* (K) fixed, vary L by one unit, and ask how much output varies. To measure MPK , we hold *labor* (L) fixed, vary K by one unit, and ask how much output varies.³

¹ The line through A and A' is nearly tangent to the isoquant and can be made more nearly tangent by measuring labor in smaller units when it is desirable to do so. Its slope is equal to the rise over the run, which is $-\Delta K/1$, or $-\Delta K$.

² Some books call this the *marginal rate of technical substitution of capital for labor*; unfortunately, there is no standard accepted terminology.

³ The discussion in this section assumed a one-unit change in labor. More generally, if labor had changed by some amount ΔL , the equation would have been

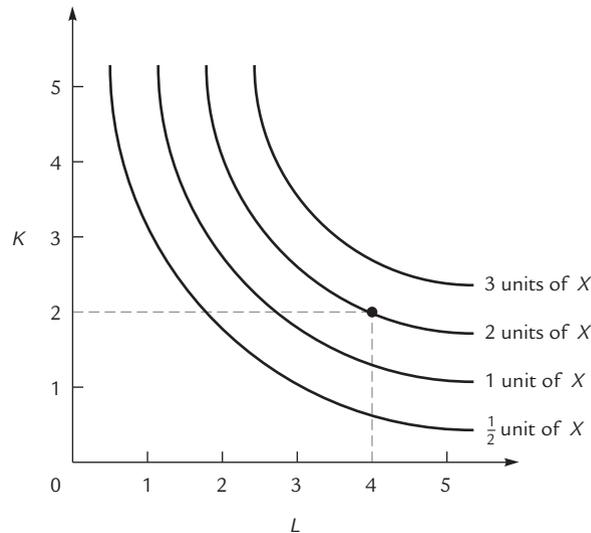
$$\Delta L \cdot MPL = \Delta K \cdot MPK$$

and we would still have reached the conclusion

$$MRTS_{LK} = \frac{\Delta K}{\Delta L} = \frac{MPL}{MPK}$$

EXHIBIT 6.8

The Production Function



The family of isoquants can be used to determine the maximum level of production that can be attained with any given level of inputs. For example, if the firm uses 4 units of labor and 2 of capital, then it can produce 2 units of output and no more. This rule for calculating the output that can be produced from a given basket of inputs is the firm's production function.

The Production Function

Suppose that the firm wants to produce 2 units of X instead of 1. We can draw an isoquant representing all of the technologically efficient input combinations that the firm can use. This “2-unit” isoquant lies above and to the right of the original “1-unit” isoquant. We can go on to draw isoquants for any given level of output, generating a family of isoquants such as the one shown in Exhibit 6.8.

The important facts about isoquants are these:

Isoquants slope downward, they fill the plane, they never cross, and they are convex.

You should recognize this list of properties; it characterizes families of indifference curves as well.

Exercise 6.7 Explain why isoquants never cross. Explain why they fill the plane.

Suppose that we want to know how much output the firm can produce with a given basket of inputs. We can use the family of isoquants to answer this question. For example, suppose that we want to know how much the firm can produce using 4 units of labor and 2 units of capital. From Exhibit 6.8 we see that this basket lies on the 2-unit isoquant; thus, the firm can use this basket to produce 2 units of X .

The rule for determining how much output can be produced with a given basket of inputs is called the firm's **production function**. If we know the family of isoquants, then we know the production function, and vice versa. Therefore, we can think of the graph in Exhibit 6.8 as providing a picture of the firm's production function.

Production function

The rule for determining how much output can be produced with a given basket of inputs.

Choosing a Production Process

In the long run, no factor of production is fixed, and the firm is free to use any production process. Given a level of output, the corresponding isoquant presents the firm with a menu of ways to produce that output, from which it chooses the option with the lowest cost. We will now develop a geometric device for keeping track of those costs.

Isocosts and Cost Minimization

Suppose that the firm can hire labor at a going wage rate of P_L and can hire capital at a going rental rate of P_K . Suppose also, for the moment, that the firm spends \$10 on inputs. Then the firm will be able to purchase L units of labor and K units of capital if and only if L and K satisfy the equation:

$$P_L \cdot L + P_K \cdot K = \$10$$

The collection of pairs (L, K) that satisfy this equation form a straight line with slope, $-(P_L/P_K)$. That line, called the \$10 **isocost**, is shown in Exhibit 6.9. Of the lines shown in the exhibit, the \$10 isocost is the one closest to the origin.

If the firm is willing to spend \$11 on inputs, then it can hire any combination of labor and capital that satisfies:

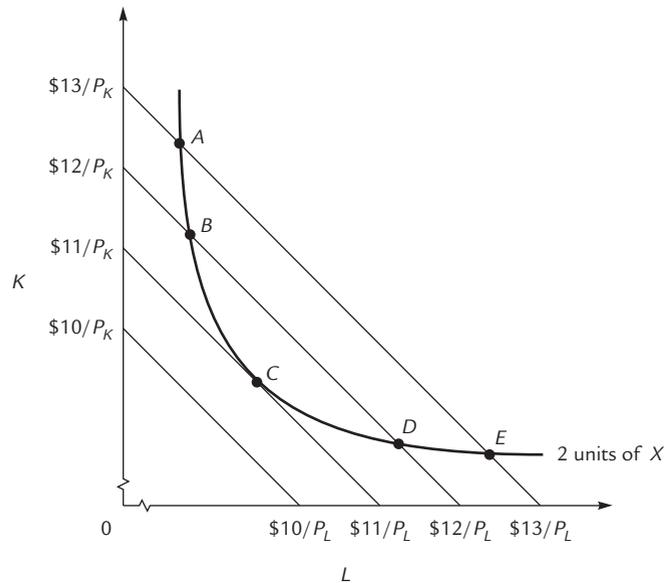
$$P_L \cdot L + P_K \cdot K = \$11$$

Isocost

The set of all baskets of inputs that can be employed at a given cost.

EXHIBIT 6.9

Cost Minimization



The isocost lines display all of the production processes that can be achieved for a given expenditure on inputs. Moving outward from the origin, the straight lines are the \$10, \$11, \$12, and \$13 isocosts. In order to produce 2 units of X , the firm must select a production process on the 2-unit isoquant. Of these processes, it will choose the one that is least costly, which is to say the one on the lowest isocost, namely, C .

The set of available points form another straight line, the \$11 isocost, which is also shown in Exhibit 6.9. The exhibit shows the \$12 and \$13 isocosts as well.

Now suppose that the firm wants to produce 2 units of output. Then it must select a production process that uses a basket of inputs on the 2-unit isoquant, shown in the exhibit. If it selects point *A*, on the \$13 isocost, then the cost of production is \$13. If it selects point *B*, the cost of production is \$12. If it selects point *C*, the cost of production is \$11. Of course, the firm wants to minimize its costs, and so it selects the production process corresponding to point *C*. The cost of producing 2 units of output is \$11.

Of course, the firm would prefer to spend only \$10 to produce its 2 units of output, but this is impossible: No point on the \$10 isocost is also on the 2-unit isoquant. The best it can do is to choose point *C*.

In order to minimize the cost of producing a given level of output, the firm always chooses a point of tangency between an isocost and the appropriate isoquant.

Cost Minimization and the Equimarginal Principle

There is another way to reach the same conclusion. Suppose that the firm considers hiring 1 less unit of labor and replacing it with sufficient capital so that it can continue producing 2 units of output. How much additional capital must it hire? The answer to this question is precisely the number that we have already called the marginal rate of technical substitution, or $MRTS_{LK}$. Recall that $MRTS_{LK}$ is also equal to the absolute value of the slope of the isoquant.

What are the marginal costs and benefits of such a decision? The marginal benefit is a saving of P_L when the firm hires 1 less unit of labor. The marginal cost arises from hiring $MRTS_{LK}$ additional units of capital at P_K each; the bill comes to $MRTS_{LK} \cdot P_K$.

The equimarginal principle tells us that the firm should seek to equate marginal cost with marginal benefit. That is, it should seek to set

$$MRTS_{LK} \cdot P_K = P_L$$

or

$$MRTS_{LK} = \frac{P_L}{P_K}$$

The left side of this equation is the absolute value of the slope of the isoquant, and the right side is the absolute value of the slope of the isocost. So the equation tells us that the firm should seek a point where the slopes of the isoquant and the isocost are equal, that is, a point of tangency.

To understand this better, let us think about what the firm can do if it is *not* at a point of tangency. What if the firm makes the mistake of operating at point *A* in Exhibit 6.9? Here the isoquant is steeper than the isocost; that is,

$$MRTS_{LK} > \frac{P_L}{P_K}$$

If the firm hires 1 more unit of labor and $MRTS_{LK}$ fewer units of capital, it can stay on the isoquant, decrease its capital costs by $MRTS_{LK} \cdot P_K$, and increase its labor costs by P_L . Because the last displayed inequality can be rewritten $MRTS_{LK} \cdot P_K > P_L$, this is a wise move for the firm to make. It shifts to the right and down along the isoquant to a point like *B*. Here, $MRTS_{LK}$ still exceeds P_L/P_K and the process is repeated; the firm

keeps moving southeast along the isoquant until it reaches point *C*, where $MRTS_{LK}$ and P_L/P_K are equal.

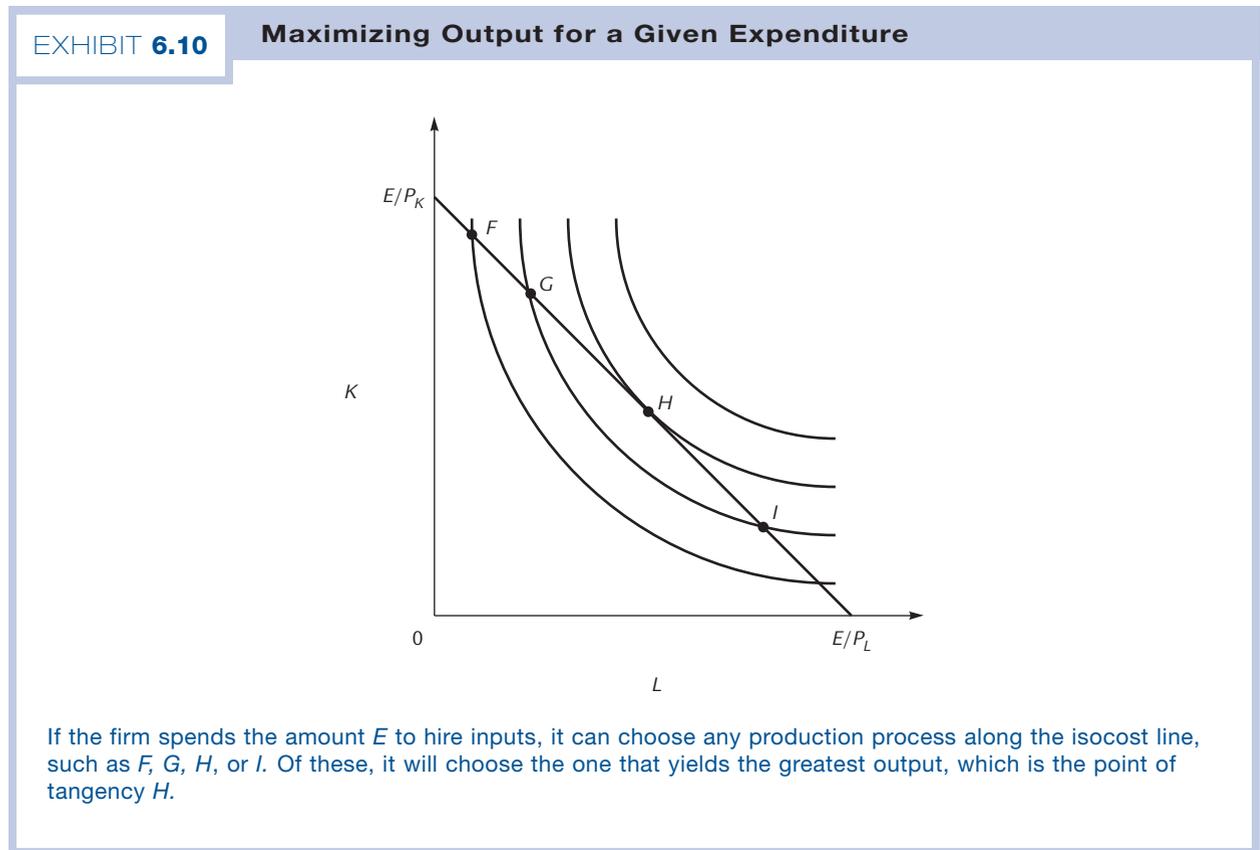
Exercise 6.8 Explain the adjustment process if the firm starts at a point like *E*.

Output Maximization

We will describe one more way to see that the firm always chooses to operate at a tangency. Exhibit 6.9 illustrates the problem of a firm that has chosen its level of output (in this case, 2 units) and seeks the least expensive way to produce it. Exhibit 6.10 illustrates the problem of a firm that has instead chosen its expenditure on inputs and is now deciding how much to produce.

If the chosen expenditure is *E*, then the firm must choose a production process on the *E* isocost, shown in Exhibit 6.10. How much does the firm want to produce? Surely, the most that it possibly can, which is to say that it wants to be on the highest available isoquant. In the figure, it is clear that this occurs at point *H*, the point of tangency.

In summary, there are two ways of looking at the firm's problem, but both lead to the same conclusion. Whether the firm wants to minimize the cost of producing a given output (as in Exhibit 6.9), or to maximize its output for a given expenditure (as in Exhibit 6.10), it is led to the same conclusion: Produce at a point where an isocost is tangent to an isoquant.



The Expansion Path

All this should have a familiar ring to it; it is reminiscent of the way consumers choose bundles of output goods to purchase. However, the analogy is less close than it first appears. There is one critical difference between the consumer (who seeks a tangency between his budget line and an indifference curve) and the firm (which seeks a tangency between an isocost and an isoquant).

The difference is this: A consumer has a *given* income to divide among consumption goods, whereas a firm can *choose* its level of expenditure on inputs. Put another way, a consumer is constrained to only one budget line, whereas a firm has a whole *family* of isocosts (one for each level of expenditure) from which it can choose.

Unlike an individual, a firm has no budget constraint. The reason is that individuals pursue consumption, whereas firms pursue profits. As a result, the firm can “afford” to spend any amount on inputs that is appropriate to its goal. Even when there is a limited amount of cash on hand, a profit-maximizing firm can borrow against its future profits to achieve whatever is the optimal level of expenditure and output.⁴ The same borrowing opportunities are not available to an individual who decides he wants to visit Hawaii.

In terms of our graphs, the consequence of all this is that we must consider the entire family of isocost lines available to the firm. They are parallel, because they all have the same slope, $-(P_L/P_K)$, but those reflecting higher levels of expenditure are farther out than others.⁵ This is shown in Exhibit 6.11.

The tangencies between isocosts and isoquants lie along a curve called the firm’s **expansion path**. We know that the firm chooses one of these tangencies. However, we have not yet said anything that allows us to determine *which* tangency the firm selects. In order to fully predict the firm’s behavior, we know from Chapter 5 that we need to take account of the marginal revenue curve, which is derived from the demand for the firm’s output. Because this information does not appear in the expansion path diagram, it is not surprising that we cannot use the diagram to predict the firm’s behavior, at least with respect to its output decision. We will not return to this question until Chapter 7.

Expansion path

The set of tangencies between isoquants and isocosts.

The Long-Run Cost Curves

To derive a firm’s long-run cost curves, we need to know its production function (i.e., the isoquants) and the input prices P_L and P_K (which determine the isocosts). By way of example, we will assume that the isoquants are as shown in Exhibit 6.11 and that the input prices are $P_K = \$10$, $P_L = \$15$. Given this information, we can plot the isocosts as in Exhibit 6.11 and draw in the expansion path by connecting the tangencies. All of this has been done in the exhibit.

Suppose the firm plans to produce 33 units of output per day. It selects a tangency on the 33 unit isoquant, which you can see from the exhibit occurs at the point where $K = 6$ and $L = 4$. Therefore, the firm hires 6 units of capital and 4 of labor for a total cost of $(6 \times \$10) + (4 \times \$15) = \$120$. This is the firm’s **long-run total cost** of producing 33 units.

Long-run total cost

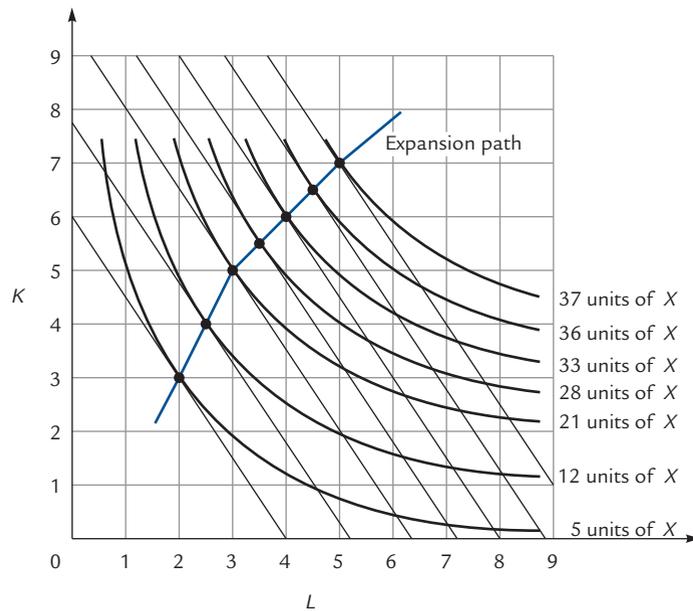
The cost of producing a given amount of output when the firm is able to operate on its expansion path.

⁴ In practice, there might actually be limitations on the firm’s ability to borrow that are not accounted for by our simple model. However, the standard assumption in elementary treatments of the theory of the firm is that all of the firm’s profits from production are available for the purchase of inputs, even before production takes place. Economists are aware that firms can face borrowing constraints and have intensely studied the consequences of those constraints, but this is a more advanced topic.

⁵ We are assuming that P_L and P_K are not affected by the actions of the firm. This assumption would fail only if the firm in question hired a significant proportion of either all the labor or all the capital in the economy.

EXHIBIT 6.11

Deriving Long-Run Total Cost



To produce 33 units of output, the firm selects the tangency, where $K = 6$ and $L = 4$. Because $P_K = \$10$ and $P_L = \$15$, the associated total cost is $(6 \times \$10) + (4 \times \$15) = \$120$.

Similarly, if the firm wants to produce 21 units of output, then it uses 5 units of capital and 3 of labor for a total cost of $(5 \times \$10) + (3 \times \$15) = \$95$.

These points can be plotted on a long-run total cost curve with output on the horizontal axis and total cost on the vertical. There is a point at (33, \$120) and another at (21, \$95).

We have discovered that *if* the firm wants to produce 33 units a day, the best way to do that is with 6 units of capital and 4 units of labor. We have *not* said that there's any reason the firm should want to produce exactly 33 units a day.

The logic goes as follows: For each possible quantity of output (e.g., 33 units a day), we figure out the cost-minimizing way to produce the output. Only *after* we have computed the cost for each quantity will we have enough information to begin thinking about what quantity the firm should actually produce.



Dangerous Curve

Exercise 6.9 What is the total cost of producing 37 units of output? 5 units of output? 12 units of output?

Long-Run Average and Marginal Costs

In Exhibit 6.4, we constructed the (short-run) average and marginal cost curves from our knowledge of the (short-run) total cost curve. We can follow exactly the same

Long-run average cost

Long-run total cost divided by quantity.

Long-run marginal cost

That part of long-run total cost attributable to the last unit produced.

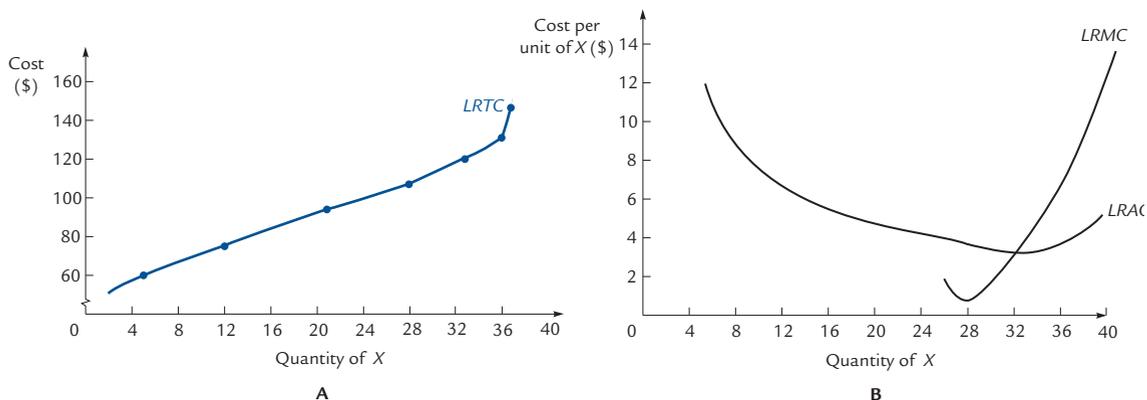
procedure with long-run costs. The long-run total cost curve of Exhibit 6.12, panel A, gives rise to the long-run average and marginal cost curves shown in panel B. **Long-run average cost** is given by the formula $LRAC = LRTC/Q$ and **long-run marginal cost** is the increment to long-run total cost attributable to the last unit of output produced. At a quantity of 33 units, we have $LRAC = TC/Q = \$120/(33 \text{ units}) = \3.63 per unit. At a quantity of 37 units, we have $LRMC = \$145 - \132.50 per unit = \$12.50 per unit. (All of the numbers here are taken from the table in Exhibit 6.12.)

If we want to compute the long-run marginal cost at a quantity of 28 units, we must subtract from \$107.50 the long-run total cost of producing 27 units, a number that is not shown in the table. However, you could in principle determine this number from Exhibit 6.11, if the 27-unit isoquant were drawn in.

Comparing the long-run Exhibit 6.12 with the short-run Exhibit 6.4, you will find that there is one fewer curve in Exhibit 6.11: In the long run, the average variable cost

EXHIBIT 6.12

Long-Run Total, Marginal, and Average Costs



Quantity of Output	Factors Employed		Cost of Factors		Total Cost
	K	L	K	L	
5	3	2	\$30	\$30	\$60
12	4	2.5	40	37.50	77.50
21	5	3	50	45	95
28	5.5	3.5	55	52.50	107.50
33	6	4	60	60	120
36	6.5	4.5	65	67.50	132.50
37	7	5	70	75	145

These cost curves are all derived from the graph in Exhibit 6.11. The table illustrates computations like the one in the caption to Exhibit 6.11. These computations yield points on the total cost curve. Points on the average cost curve are computed by dividing total cost by quantity: When 33 units are produced, the average cost is $\$120/33 = \3.63 . Points on the marginal cost curve are computed by taking differences in total cost: When 37 items are produced, the marginal cost is $\$145 - \$132.50 = \$12.50$. To compute the marginal cost when 28 items are produced, we must start with \$107.50 and subtract the total cost of producing 27 items. The latter number does not appear in the table, but could be computed from the graph in Exhibit 6.11.

curve has disappeared. This is because all costs are variable in the long run; therefore, in the long run there is no distinction between average cost and average variable cost.

Returns to Scale and the Shape of the Long-Run Cost Curves

Our goal is to determine the shape of the firm's long-run marginal and average cost curves. Because these curves are derived from the long-run total cost curve, which is in turn derived from the production function, it behooves us to start by thinking a little harder about the production function itself.

Here is an important question about the production function: When all input quantities are increased by 1%, does output go up by (1) more than 1%, (2) exactly 1%, or (3) less than 1%? Depending on the answer to this question, we say that the production function exhibits (1) **increasing returns to scale**, (2) **constant returns to scale**, or (3) **decreasing returns to scale**.

Students often confuse the concepts of diminishing marginal returns, on the one hand, and decreasing returns to scale, on the other. The two concepts are entirely different, and they are entirely different in each of two ways. The most important difference is that diminishing marginal returns is a short-run concept that describes the effect on output of increasing *one* input while holding other inputs fixed. Decreasing returns to scale is a long-run concept that describes the effect on output of increasing *all* inputs in the same proportion. The other difference is that the concept of diminishing marginal returns deals with marginal quantities, whereas the concept of decreasing returns to scale deals with total and average quantities. When we ask about diminishing marginal returns, we ask, "Will the next unit of this input yield more or less output *at the margin* than the last unit did?" When we ask about decreasing returns to scale, we ask, "Will a 1% increase in all inputs yield more or less than a 1% increase in *total* output?"

For given input prices, diminishing marginal returns are reflected by an increasing short-run marginal cost curve. Decreasing returns to scale, as we shall soon see, are reflected by an increasing long-run average cost curve.

Increasing Returns to Scale

Increasing returns to scale are likely to result when there are gains from specialization or when there are organizational advantages to size. Two men with two machines might be able to produce more than twice as much as one man with one machine, if each can occasionally use a helping hand from the other. At low levels of output, firms often experience increasing returns to scale.

Constant and Decreasing Returns to Scale

At higher levels of output, the gains from specialization and organization having been exhausted, firms tend to produce under conditions of constant or even decreasing returns to scale. Which of the two, constant or decreasing returns, is more likely? A good case can be made for constant returns. When a firm doubles all of its inputs, it can, if it chooses, simply set up a second plant, identical to the original one, and have each plant produce at the original level, yielding twice the original output. This strategy

Increasing returns to scale

A condition where increasing all input levels by the same proportion leads to a more than proportionate increase in output.

Constant returns to scale

A condition where increasing all input levels by the same proportion leads to a proportionate increase in output.



Dangerous
Curve

Decreasing returns to scale

A condition where increasing all input levels by the same proportion leads to a less than proportionate increase in output.

generates constant returns to scale and suggests that the firm should never have to settle for decreasing returns. This argument is often summed up in the slogan “What a firm can do once, it can do twice.”

Students sometimes object to this argument for constant returns. They argue that doubling the number of workers and the number of machines can lead to congestion in the factory and consequently to less than a doubling of output. This objection overlooks the fact that factory space is itself a productive input. When we measure returns to scale, we assume that *all* inputs are increased in the same proportion. In particular, we must double the space in the factory as well as the numbers of workers and machines.

A related objection is that when the scale of an operation is doubled, the owners can no longer keep as watchful an eye on the entire enterprise as they could previously. But if we view the owners’ supervisory talents as a productive input, this objection breaks down as well. Any measurement of returns to scale must involve the imaginary experiment of increasing these talents in the same proportion as all other productive inputs.

As long as *all* productive inputs are truly variable, the argument for constant returns is a convincing one. However, if there are some inputs (such as managerial skills or the owner’s cleverness as an entrepreneur) that are truly *fixed even in the long run*, then there may be decreasing returns to scale with respect to changes in all of the variable inputs. As a result, most economists are comfortable with the assumption that firms experience decreasing returns to scale at sufficiently high levels of output.



Dangerous
Curve

We assumed at the outset that in the long run every input is variable. When we now admit the possibility that some inputs may not be variable in the long run, we are admitting that our original model might not be a fully adequate description of reality.

Returns to Scale and the Average Cost Curve

Under conditions of increasing returns to scale, the firm’s long-run average cost curve is decreasing. This is because a 1% increase in output can be accomplished with less than a 1% increase in all inputs. It follows that an increase in output leads to a fall in the average cost of production.⁶

Under conditions of decreasing returns to scale, the firm’s long-run average cost curve is increasing.

Exercise 6.10 Justify the assertion of the preceding paragraph.

Under conditions of constant returns to scale, the firm’s long-run average cost curve is flat. This is the situation where “What a firm can do once it can do twice.” If the firm wants to double its output, it does so by doubling all of its inputs. The average cost per unit of output never changes.

If we assume that a firm experiences increasing returns to scale at low levels of output and decreasing returns thereafter, the firm’s long-run average cost curve is

⁶ This argument assumes that the firm can hire all of the inputs that it wants to at a going market price. Without this assumption, the long-run average cost curve could be increasing even in the presence of increasing returns to scale. The same caveat applies to all of our arguments in this subsection.

U-shaped, as in panel B of Exhibit 6.12. Only at one level of output (the quantity at which long-run average cost is minimized) does the firm face constant returns to scale.

When long-run marginal cost is below long-run average cost, long-run average cost is decreasing; and when long-run marginal cost is above long-run average cost, long-run average cost is increasing. Consequently, when long-run average cost is U-shaped, it is cut by long-run marginal cost at the bottom of the U. This is true in the long run for the same reason that it is true in the short run.

In general, the upward-sloping part of the firm's long-run marginal cost curve will be much more elastic than the upward-sloping part of its short-run marginal cost curve. Marginal cost rises much more quickly when the firm is constrained not to vary certain inputs (in the short run) than when it can vary all inputs to minimize costs for each level of output (in the long run).

6.3 Relations Between the Short Run and the Long Run

In Section 6.1, we studied the firm's short-run production function and cost curves; in Section 6.2, we studied the firm's long-run production function and cost curves. Our remaining task is to relate the two points of view.

From Isoquants to Short-Run Total Cost

Consider a firm that rents capital at a rate of $P_K = \$10$ and hires labor at a rate of $P_L = \$15$. The firm's production function is illustrated in Exhibit 6.13. Its capital is fixed in the short run at 5 units (thus, if a "unit" is a machine, the firm has the use of 5 machines; if a "unit" is 100 square feet of office space, the firm has the use of 500 square feet).

In the short run, the firm can only choose input baskets that contain exactly 5 units of capital, which is to say that it can only choose baskets that are located on the blackened horizontal line. To produce 5 units of output, it must select a basket that is both on this line and on the 5-unit isoquant; that is, it must select the point with 5 units of capital and 1 unit of labor. The firm's total cost is then $5 \times \$10 = \50 for capital plus $1 \times \$15 = \15 for labor, or \$65. (Of this \$65, the \$50 spent on capital is a fixed cost and the \$15 spent on labor is a variable cost.) This calculation is recorded in the first row of the table, under the columns headed "Short Run."

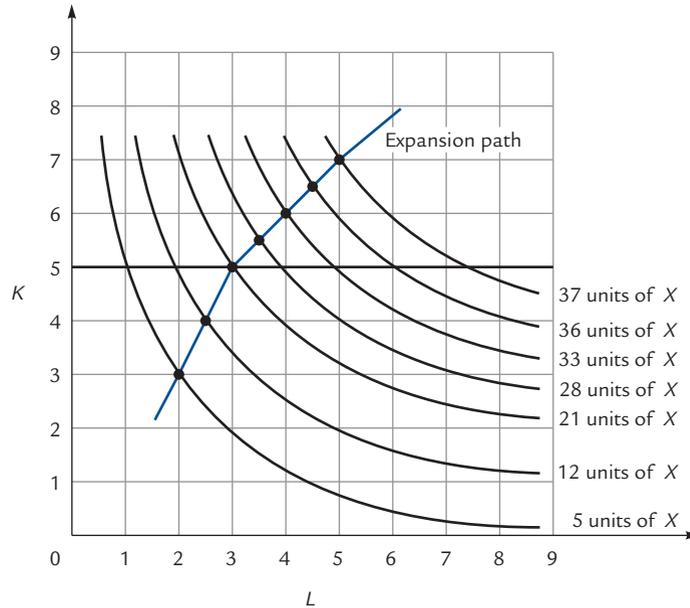
Similarly, if the firm wants to produce 12 units of output, it must select a point on both the blackened horizontal line and the 12-unit isoquant; that is, it must use 5 units of capital and 2 units of labor. Its total cost is \$80, as recorded in the second row of the table.

From the numbers in the Short Run half of the table, we can discover the firm's total product and total cost curves. The first column shows quantities of output, and the third shows the quantity of labor needed to produce that output. The information here is identical to the information in the first two columns of the table in Exhibit 6.1. The moral is this: If you know the isoquants and the fixed quantity of capital, you can derive the (short-run) total product curve.

If, in addition, you know the factor prices, then you can also derive the short-run variable cost and total cost curves, as we showed in Exhibits 6.2 and 6.3. The same computations are shown again in Exhibit 6.13, under the Short Run columns showing

EXHIBIT 6.13

Short-Run and Long-Run Total Cost Curves



Quantity of Output	Short Run					Long Run				
	Factors Employed		Cost of Factors		Total Cost	Factors Employed		Cost of Factors		Total Cost
	K	L	K	L		K	L	K	L	
5	5	1	\$50	\$15	\$65	3	2	\$30	\$30	\$60
12	5	2	50	30	80	4	2.5	40	37.50	77.50
21	5	3	50	45	95	5	3	50	45	95
28	5	4	50	60	110	5.5	3.5	55	52.50	107.50
33	5	5	50	75	125	6	4	60	60	120
36	5	6	50	90	140	6.5	4.5	65	67.50	132.50
37	5	7	50	105	155	7	5	70	75	145

(Continues)

the cost of labor and total cost. The resulting short-run total cost curve, labeled *SRTC* in the second panel of Exhibit 6.13, is identical to the one shown in Exhibit 6.3.⁷

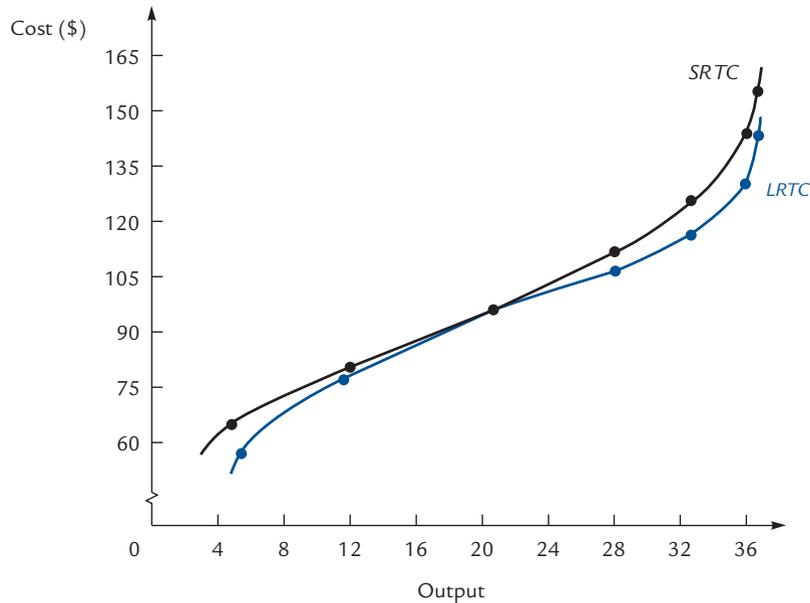
From Isoquants to Long-Run Total Cost

Exhibits 6.11 and 6.12 already illustrated the derivation of long-run total cost from isoquants and factor prices. These computations are repeated in the “Long Run” columns

⁷ In Section 6.1 we wrote *TC* for short-run total cost. We are now writing *SRTC* to distinguish the short-run total cost curve from the long-run total cost curve.

EXHIBIT 6.13

Short-Run and Long-Run Total Cost Curves (Continued)



With $P_K = \$10$ and $P_L = \$15$, the isoquant diagram gives rise to the table. Points from the table are plotted on the graph. The short-run total cost (*SRTC*) curve is drawn on the assumption that capital employment is fixed at 5 units. It is the same curve that was constructed in Exhibit 6.3. Because the firm always chooses the least expensive production process in the long run, long-run total cost is never greater than short-run total cost. If the firm happens to want to produce exactly 21 units of output, then its desired long-run capital employment is equal to its existing capital employment of 5 units. In this fortunate circumstance, the firm can produce at the lowest possible cost even in the short run. For any other level of output, short-run total cost exceeds long-run total cost.

of the table in Exhibit 6.13, and the resulting *LRTC* curve is redrawn in the second panel of that exhibit.

Short-Run Total Cost versus Long-Run Total Cost

To produce 12 units of output, the firm in Exhibit 6.13 selects the least expensive production process in the long run. Its costs total \$77.50. In the short run, the firm is forced to use a more expensive process, and so its costs are higher, totaling \$80. This illustrates something important:

Short-run total cost is always at least as great as long-run total cost.

The reason is simple. In the long run, the firm produces at the lowest possible cost. The short-run cost has no chance of being less than the lowest possible! Geometrically, this means that *SRTC* never dips below *LRTC*. You can see that this is true in Exhibit 6.13.

We can say even more. There is exactly one quantity of output for which the short-run and long-run total costs are equal. In Exhibit 6.13, that quantity is 21. This is the quantity at which the firm's long-run desired capital employment (in this case, 5 units)

happens to precisely equal the fixed amount of capital it has available. You can see in the exhibit that the *SRTC* and *LRTC* curves touch at a quantity of 21.

A Multitude of Short Runs

All of the *short-run* numbers in Exhibit 6.13 are derived on the assumption that the firm's capital is fixed at 5 units. What if capital is fixed at 4 units instead? Now what is the short-run total cost of producing 5 units of output? In order to achieve the 5-unit isoquant with 4 units of capital, the firm must employ 1.5 units of labor. The short-run total cost is $(4 \times \$10) + (1.5 \times \$15) = \$62.50$. To produce 12 units of output, the firm must employ 2.5 units of labor, and the short-run total cost is \$77.50.

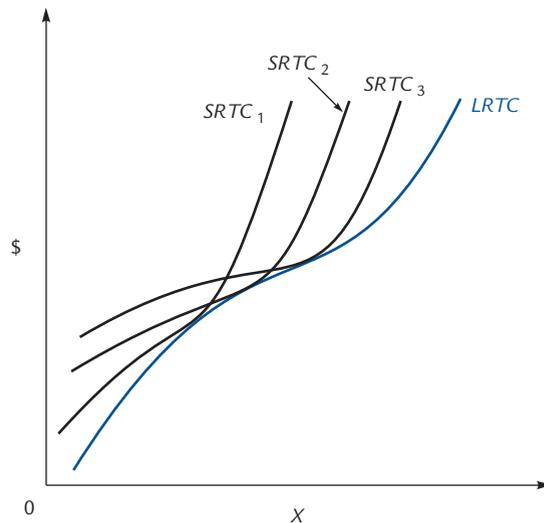
Exercise 6.11 With 4 units of capital, what is the *SRTC* when quantity is 28? When it is 33? When it is 36?

Plotting these points, we can construct a new short-run total cost curve, different from the one we constructed before. The new *SRTC* curve again touches the *LRTC* curve at exactly one point, this time at a quantity of 12.

For every quantity of capital, there is a corresponding *SRTC* curve, touching the *LRTC* curve at exactly one point. The geometry is illustrated in Exhibit 6.14.

EXHIBIT 6.14

Many Short-Run Total Cost Curves



When we draw a short-run total cost curve, we assume a fixed level of capital employment. If we assume a different fixed level of capital employment, we get a different short-run total cost curve. The graph shows the short-run total cost curves that result from various assumptions.

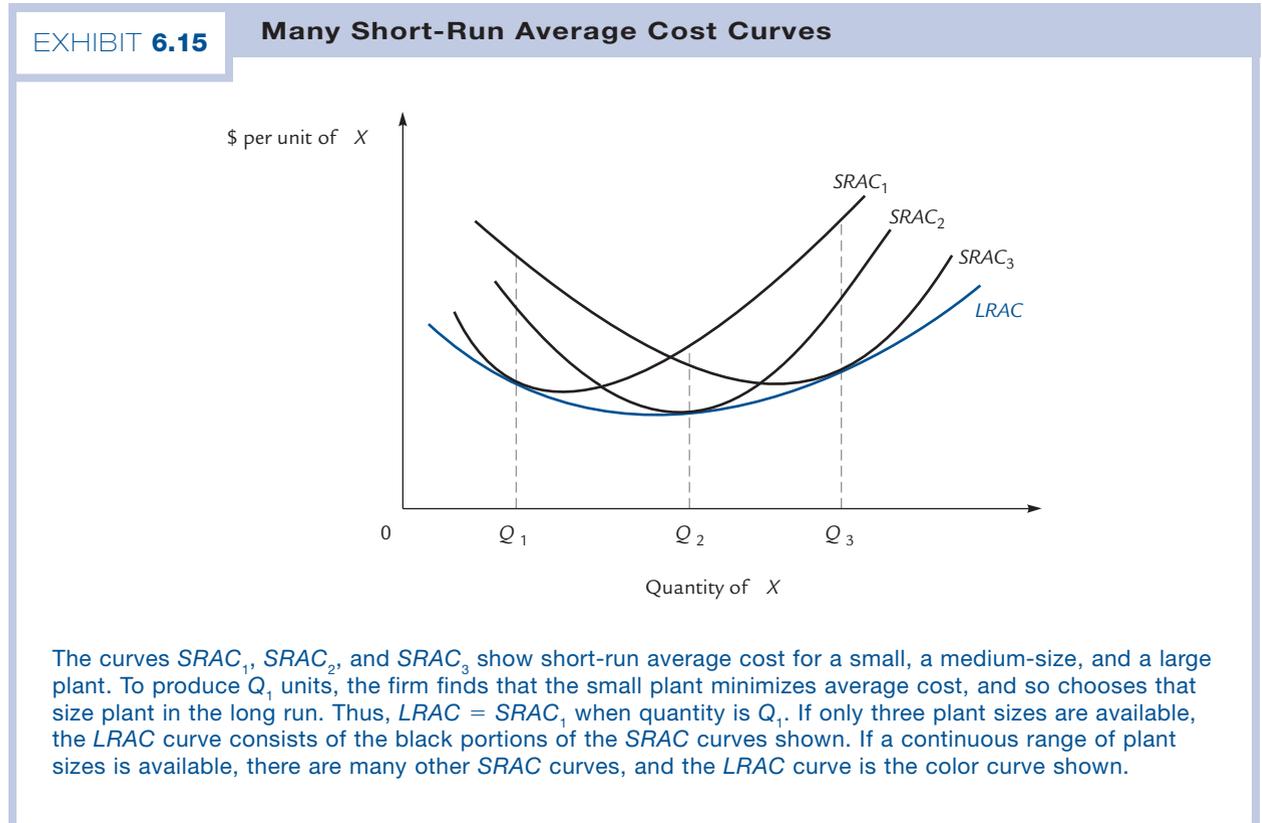
Each total cost curve touches the long-run total cost curve in one place, at that level of output for which the fixed capital stock happens to be optimal. In that case, the firm's long-run and short-run choices of production process coincide. The long-run total cost curve is the lower boundary of the region in which the various short-run total cost curves lie.

Short-Run Average Cost versus Long-Run Average Cost

Instead of plotting total cost curves, we can plot average cost curves. There is a different short-run average cost curve for each quantity of capital. You can think of capital as a measure of “plant size,” so that the short-run average cost curves in Exhibit 6.15 describe the situation for a small, a medium-size, and a large plant. If the firm wants to produce quantity Q_1 , average cost is minimized by the small plant represented by the curve $SRAC_1$. If the firm is required to operate with the medium-size plant represented by curve $SRAC_2$, its average cost is higher; if it operates with the large plant represented by $SRAC_3$, its average cost is even higher yet. In the long run, if Q_1 is the desired output, the firm chooses the small plant to minimize its average cost. Consequently, at Q_1 units, the long-run average cost is the same as the small plant’s short-run average cost. That is why the $SRAC_1$ and $LRAC$ curves touch at Q_1 .

If the firm wants to produce Q_3 units, it achieves the lowest average cost with the large plant, a somewhat higher average cost with the medium-size plant, and an even higher average cost with the small plant. In the long run, it chooses the large plant, so $LRAC$ is the same as $SRAC_3$ for Q_3 units of output.

Exercise 6.12 Suppose that the firm wants to produce Q_2 units of output. Which plant size is best? Which is second best? Which plant size will it choose in the long run? How is this fact reflected in the graph?



If the firm has only three possible plant sizes, then its long-run average cost curve consists of the black parts of the three short-run average cost curves. (For any quantity, the firm selects the optimal plant size and so achieves a point on one of the three *SRACs*.) In most of this chapter, we have assumed instead that the firm has a continuous range of plant sizes (i.e., it can choose any quantity of capital it desires). In this case, there are many other *SRAC* curves besides those pictured, and *LRAC* is the color curve in the graph. Each point on *LRAC* is then a point of tangency with some *SRAC* curve.

Summary

The role of the firm is to convert inputs into outputs. The cost of producing a given level of output depends on the technology available to the firm (which determines the quantities of inputs the firm will need) and the prices of the inputs.

In the short run, the firm is committed to employing some inputs in fixed amounts. In the long run, it is free to vary its employment of every input, always producing at the lowest possible cost.

For illustrative purposes, we consider a firm that employs labor and capital, with capital fixed in the short run. The options available to the firm are then illustrated by its total product (*TP*) curve, also called its *short-run production function*. From the *TP* curve, we can derive the marginal product of labor (*MPL*) curve by computing the additional output derived from each additional unit of labor: The value of *MPL* is the slope of *TP*.

The average product of labor (*APL*) is defined to be TP/L , where L is the amount of labor employed. At low levels of output (the first stage of production), each additional worker increases the productivity of his colleagues. Therefore, marginal product exceeds average product and average product is rising. At higher levels of output (the second stage of production), each additional worker reduces the productivity of his colleagues. Therefore, marginal product is below average product and average product is falling. The average product curve has the shape of an inverted U, with the marginal product curve cutting through it at the highest point.

For a given level of output, the firm faces a fixed cost (*FC*), which is the cost of renting capital, and a variable cost (*VC*), which is the cost of hiring labor. *FC* can be computed as $P_K \cdot K$, where P_K is the price of capital and K is the firm's (fixed) capital usage. *VC* can be computed as $P_L \cdot L$, where P_L is the wage rate of labor and L is the quantity of labor needed to produce the desired output; the value of L that corresponds to a given quantity of output can be found by examining the *TP* curve.

The firm's total cost (*TC*) is the sum of *FC* and *VC*. Its average cost (*AC*) is TC/Q , where Q is the quantity of its output. Its average variable cost (*AVC*) is VC/Q . Its marginal cost is the increment to total cost attributable to the last unit of output.

Typically, the average, average variable, and marginal cost curves are U-shaped. *MC* cuts through both *AC* and *AVC* at their minimum points.

In the long run, the firm's technology is embodied in its production function, which is illustrated by the isoquant diagram. The slope of an isoquant is equal to the marginal rate of technical substitution between labor and capital. We expect $MRTS_{LK}$ to decrease as we move down and to the right along the isoquant, with the result that isoquants are convex.

In the long run, the firm minimizes costs for a given level of output, which leads it to choose a point of tangency between an isocost and an isoquant. Alternatively, we can think of the firm as maximizing output for a given expenditure on inputs; this reasoning also leads to the conclusion that the firm operates at a tangency. The set of all such tangencies forms the firm's expansion path.

To compute the long-run total cost for Q units of output, find the tangency of the Q -unit isoquant with an isocost, and compute the price of the corresponding input basket.

Long-run average and marginal costs can be computed from long-run total cost.

The long-run average cost curve is downward sloping, flat, or upward sloping, depending on whether the firm experiences increasing, constant, or decreasing returns to scale. We expect increasing returns (decreasing average cost) at low levels of output because of the advantages of specialization. At higher levels of output, there will be constant returns to scale unless some factor is fixed even in the long run; however, this case is very common because of limits on things like the skills and supervisory ability of the entrepreneur. Therefore, we often draw the long-run average cost curve increasing at high levels of output, making the entire curve U-shaped. (That is, we assume decreasing returns to scale at high levels of output.) Long-run marginal cost cuts through long-run average cost at the bottom of the U.

The same isoquant diagram that is used to derive long-run total cost can be used to derive short-run total product and total cost curves as well. Each possible plant size for the firm results in a different short-run total cost curve and consequently a different short-run average cost curve. The short-run cost curves never dip below the long-run cost curves. The short-run total cost curve associated with a given plant size touches the long-run total cost curve only at that quantity for which the plant size is optimal; the same is true for average cost curves.

Author Commentary

www.cengage.com/economics/landsburg

AC1. The efficient design of biological organisms has much in common with the efficient choice of a production process.

Review Questions

- R1.** What are the first and second stages of production?
- R2.** What is the shape of the *APL* curve? Why?
- R3.** Where does the *MPL* curve cross the *APL* curve? Why?
- R4.** What is the relationship between the *MPL* curve and the total product curve?
- R5.** Explain how to derive the firm's *VC* and *TC* curves from its *TP* curve.
- R6.** Explain how to derive the firm's *AC*, *AVC*, and *MC* curves.
- R7.** What geometric relationships hold among *AC*, *AVC*, and *MC*? Why?

- R8.** Define the marginal rate of technical substitution.
- R9.** What is the relationship between the marginal products of the factors of production and the marginal rate of technical substitution?
- R10.** What are the geometric properties of isoquants? Why do we expect these properties to hold?
- R11.** Explain why firms want to operate at a tangency between an isoquant and an isocost.
- R12.** Explain how to derive a firm's long-run total cost curve from its isoquant diagram and knowledge of the factor prices.
- R13.** What are increasing, constant, and decreasing returns to scale? How are they related to the shape of the long-run average cost curve?
- R14.** Explain how to derive the firm's (short-run) total product and total cost curves from the isoquant diagram. How would these curves be affected by a change in the rental rate on capital? How would they be affected by a change in the wage rate of labor?
- R15.** What is the relationship between the firm's long-run and short-run total cost curves?

Numerical Exercises

- N1.** A firm discovers that when it uses K units of capital and L units of labor, it is able to produce \sqrt{KL} units of output.
- Draw the isoquants corresponding to 1, 2, 3, and 4 units of output.
 - Suppose that the firm produces 10 units of output using 20 units of capital and 5 units of labor. Compute the $MRTS_{LK}$. Compute the MPL . Compute the MPK .
 - On the basis of your answers to part (b), is the equation $MRTS_{LK} = MPL/MPK$ approximately true? (It would become closer to being true if we measured inputs in smaller units.)
 - Suppose that capital and labor can each be hired at \$1 per unit and that the firm uses 20 units of capital in the short run. What is the short-run total cost to produce 10 units of output?
 - Continue to assume that capital and labor can each be hired at \$1 per unit. Show that in the long run, if the firm produces 10 units of output, it will employ 10 units of capital and 10 units of labor. (*Hint:* Remember that in the long run the firm chooses to set $MPK/P_K = MPL/P_L$.) What is the long-run total cost to produce 10 units of output?
 - Does this production function exhibit constant, increasing, or decreasing returns to scale?
- N2.** Repeat problem N1, replacing the function \sqrt{KL} with the function $K^{1/3}L^{2/3}$.

Problem Set

1. Suppose that you hire workers to address and stamp envelopes. Each worker earns \$5 per hour and produces 50 addressed, stamped envelopes per hour. You have unlimited free office space and can therefore add as many workers as you want to with no fall-off in productivity. You have no expenses other than paying workers. Draw the total product, marginal product, average product, total cost, average cost, average variable cost, and marginal cost curves.
2. Suppose in the preceding problem that you rent a stamping machine with unlimited capacity, for \$10 per hour. This makes it possible for workers to increase their output to 100 addressed, stamped envelopes per hour. Draw the new total product, marginal product, average product, total cost, average cost, average variable cost, and marginal cost curves.
3. In the situation of problems 1 and 2, suppose that you have a choice between renting the machine or not renting it. For what levels of output will you choose to rent the machine? For what levels of output will you choose not to? Suppose that in the long run you can decide whether or not to rent the machine. Draw your long-run total and average cost curves.
4. Suppose that your factory faces a total product curve that contains the following points:

Quantity of Labor	Total Product
6	1
10	2
13	3
15	4
18	5
23	6
30	7
40	8

If labor costs \$2 per unit, and you have fixed costs of \$30, construct tables showing your variable cost, total cost, average cost, and average variable cost curves.

5. Suppose that in the short run, capital is fixed and labor is variable. **True or False:** If the price of capital goes up, the firm's (short-run) average cost, average variable cost, and marginal cost curves will remain unaffected.
6. Suppose that in the short run, capital is fixed and labor is variable. **True or False:** If the price of labor goes up, the firm's (short-run) average cost, average variable cost, and marginal cost curves will all shift upward.
7. **True or False:** A wise entrepreneur will minimize costs for a given output rather than maximize output for a given cost.
8. Suppose that a firm is operating at a point off its expansion path, where

$$MRTS_{LK} > \frac{P_L}{P_K}$$

Explain how this firm could increase its output without changing its total expenditure on inputs. Use this to give an additional argument for why a firm operating off its expansion path would want to move toward its expansion path.

9. Widgets are produced using thingamabobs and doohickeys. For some reason, a certain firm always produces exactly three widgets per day. **True or False:** If the price of thingamabobs increases, then in the long run the firm is certain to switch to a production process that uses fewer thingamabobs and more doohickeys.
10. A firm faces the following total product curves depending on how much capital it employs:

<i>K</i> = 1 Unit		<i>K</i> = 2 Units		<i>K</i> = 3 Units	
Quantity of Labor	Total Product	Quantity of Labor	Total Product	Quantity of Labor	Total Product
1	100	1	123	1	139
2	152	2	187	2	193
3	193	3	237	3	263
4	215	4	263	4	319
5	233	5	286	5	366
6	249	6	306	6	407
7	263	7	323	7	410

- a. Suppose that the firm currently employs 1 unit of capital and 3 of labor. Compute $MRTS_{LK}$. Compute MPL . Compute MPK .
- b. Suppose that the firm currently employs 2 units of capital. The price of capital is \$4 per unit and the price of labor is \$10 per unit. What is the short-run total cost of producing 263 units of output? What is the long-run total cost of producing 263 units of output?
- c. Suppose that the price of capital increases to \$20 per unit and the price of labor falls to \$5 per unit. Now what is the long-run total cost of producing 263 units of output?
- d. Beginning with 1 unit of capital and 2 units of labor, does this production function exhibit increasing, constant, or decreasing returns to scale? Which way does the long-run average cost curve slope?
11. Terry's Typing Service produces manuscripts. The only way to produce a manuscript is for 1 secretary to use 1 typewriter for 1 day. Two secretaries with 1 typewriter or 1 secretary with 2 typewriters can still produce only 1 manuscript per day.
- a. Draw Terry's 1-unit isoquant.
- b. Assuming that Terry's technology exhibits constant returns to scale, draw several more isoquants.
- c. Assuming that Terry rents typewriters for \$4 apiece per day and pays secretaries \$6 apiece per day, draw some of Terry's isocosts. Draw the expansion path.
- d. Terry has signed a contract to rent exactly 5 typewriters. Illustrate the following, using tables, graphs, or both: the total product and marginal product of labor; the short-run total cost, variable cost, average cost, average variable cost, and marginal cost; the long-run total cost, long-run average cost, and long-run marginal cost.

- 12.** The desert town of Dry Gulch buys its water from LowTech Inc. LowTech hires residents to walk to the nearest oasis and carry back buckets of water. Thus, the inputs to the production of water are workers and buckets. The walk to the oasis and back takes one full day. Each worker can carry either 1 or 2 buckets of water but no more.
- a.** Draw some of LowTech's isoquants. With buckets renting for \$1 a day and workers earning \$2 per day, draw some of LowTech's isocosts. Draw the expansion path.
 - b.** LowTech owns 5 buckets. It could rent these out to another firm at \$1 per day, or it could rent additional buckets for \$1 per day, but neither transaction could be arranged without some delay. Illustrate the following, using tables, graphs, or both: the total product and marginal product of labor; the short-run total cost, variable cost, average cost, average variable cost, and marginal cost; the long-run total cost, long-run average cost, and long-run marginal cost.

