University of Zurich[UZH]

Leseaufträge «Mikroökonomik I»
Modul 3: Produktion und Kosten

Unit 3:

- Langfristige Produktion

Quellen:

- **Chapter 10 – Production**
  Frank, Robert H, & Cartwright, Edward. (2016). *Microeconomics and Behaviour (2nd European ed.)*. London: McGraw-Hill Education.

PRODUCTION IN THE LONG RUN

The examples discussed thus far have involved production in the short run, where at least one productive input cannot be varied. In the long run, by contrast, all factors of production are by definition variable. In the short run, with K held fixed in the production function $Q = F(K, L)$, we were able to describe the production function in a simple two-dimensional diagram. With both K and L variable, however, we now require three dimensions instead of two. And when there are more than two variable inputs, we require even more dimensions.

This creates a problem similar to the one we encountered in Chapter 4 when the consumer was faced with a choice between multiple products: we are not very adept at graphical representations involving three or more dimensions. For production with two variable inputs, the solution to this problem is similar to the one adopted in Chapter 4.

To illustrate, consider again the production function discussed earlier in this chapter:

$$Q = F(K, L) = 2KL \tag{10.4}$$

and suppose we want to describe all possible combinations of K and L that give rise to a particular level of output—say, $Q = 16$. To do this, we solve $Q = 2KL = 16$ for K in terms of L, which yields

$$K = \frac{8}{L} \tag{10.5}$$

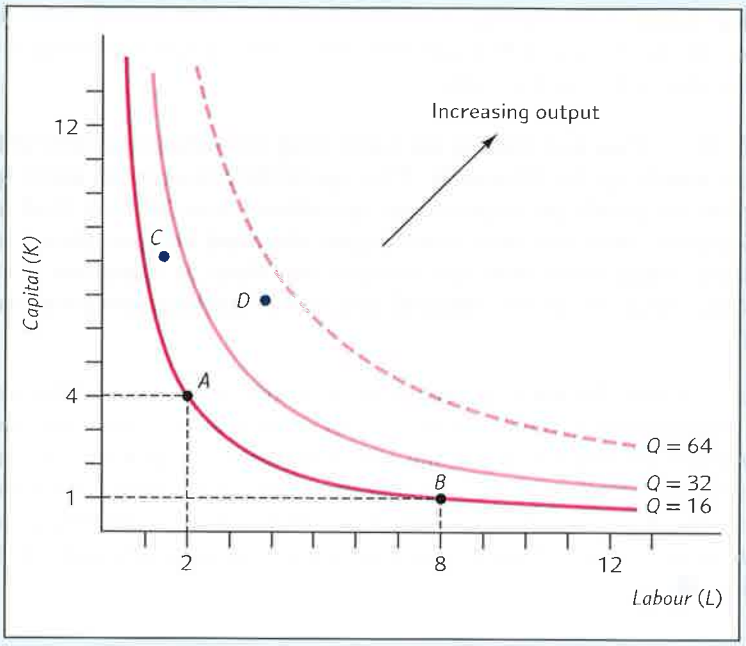
The (L, K) pairs that satisfy Equation 10.5 are shown by the curve labelled $Q = 16$ in Figure 10.7. The (L, K) pairs that yield 32 and 64 units of output are shown in Figure 10.7 as the curves labelled $Q = 32$ and $Q = 64$, respectively. Such curves are called **isoquants**, and are defined formally as *all combinations of variable inputs that yield a given level of output.*[7]

isoquants the set of all input combinations that yield a given level of output.

Note the clear analogy between the isoquant and the indifference curve of consumer theory. Just as an indifference map provides a concise representation of a consumer's preferences, an *isoquant map* provides a concise representation of a production process.

[7]'Iso' comes from the Greek word for 'same', which also appears, for example, in the meteorological term 'isobars', meaning lines of equal barometric pressure.

FIGURE 10.7
Part of an Isoquant Map for the Production Function $Q = 2KL$
An isoquant is the set of all (L, K) pairs that yield a given level of output. For example, each (L, K) pair on the curve labelled $Q = 32$ yields 32 units of output. The isoquant map describes the properties of a production process in much the same way as an indifference map describes a consumer's preferences.



On an indifference map, movements to the northeast correspond to increasing levels of satisfaction. Similar movements on an isoquant map correspond to increasing levels of output. A point on an indifference curve is preferred to any point that lies below that indifference curve, and less preferred than any point that lies above it. Likewise, any input bundle on an isoquant yields more output than any input bundle that lies below that isoquant, and less output than any input bundle that lies above it. Thus, bundle C in Figure 10.7 yields more output than bundle A, but less output than bundle D.

The only substantive respect in which the analogy between isoquant maps and indifference maps is incomplete is the significance of the labels attached to the two types of curve. From Chapter 4 recall that the actual numbers assigned to each indifference curve were used to indicate only the relative rankings of the bundles on different indifference curves. The number we assign to an isoquant, by contrast, corresponds to the actual level of output we get from an input bundle along that isoquant. With indifference maps, we are free to relabel the indifference curves in any way that preserves the original ranking of bundles. But with isoquant maps, the labels are determined uniquely by the production function.

The Marginal Rate of Technical Substitution

Recall from our discussion of consumer theory in Chapter 4 that the marginal rate of substitution is the rate at which the consumer is willing to exchange one good for another along an indifference curve. The analogous concept in production theory is called the **marginal rate of technical substitution**, or **MRTS**. It is the rate at which one input can be exchanged for another without altering output. In Figure 10.8, for example, the MRTS at A is defined as the absolute value of the slope of the isoquant at A, $|\Delta K/\Delta L|$.

marginal rate of technical substitution (MRTS) the rate at which one input can be exchanged for another without altering the total level of output.

In consumer theory, we assumed that the marginal rate of substitution diminishes with downward movements along an indifference curve. For most production functions, the MRTS displays a similar property. Holding output constant, the less we have of one input, the more we must add of the other input to compensate for a one-unit reduction in the first input.

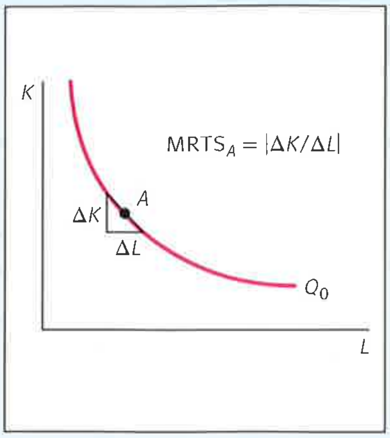


FIGURE 10.8
The Marginal Rate of Technical Substitution
The MRTS is the rate at which one input can be exchanged for another without altering total output. The MRTS at any point is the absolute value of the slope of the isoquant that passes through that point. If ΔK units of capital are removed at point A, and ΔL units of L are added, output will remain the same at Q_0 units.

A simple but important relationship exists between the MRTS at any point and the marginal products of the respective inputs at that point. In a small neighbourhood of point A in Figure 10.8, suppose we reduce K by ΔK and augment L by an amount ΔL just sufficient to maintain the original level of output. If MP_{KA} denotes the marginal product of capital at A, then the reduction in output caused by the loss of ΔK is equal to $\text{MP}_{KA}\,\Delta K$. Using MP_{LA} to denote the marginal product of L at A, it follows similarly that the gain in output resulting from the extra ΔL is equal to $\text{MP}_{LA}\,\Delta L$. Finally, since the reduction in output from having less K is exactly offset by the gain in output from having more L, it follows that

$$\text{MP}_{KA}\,\Delta K = \text{MP}_{LA}\,\Delta L \tag{10.6}$$

Cross-multiplying, we get

$$\frac{\text{MP}_{LA}}{\text{MP}_{KA}} = \frac{\Delta K}{\Delta L} \tag{10.7}$$

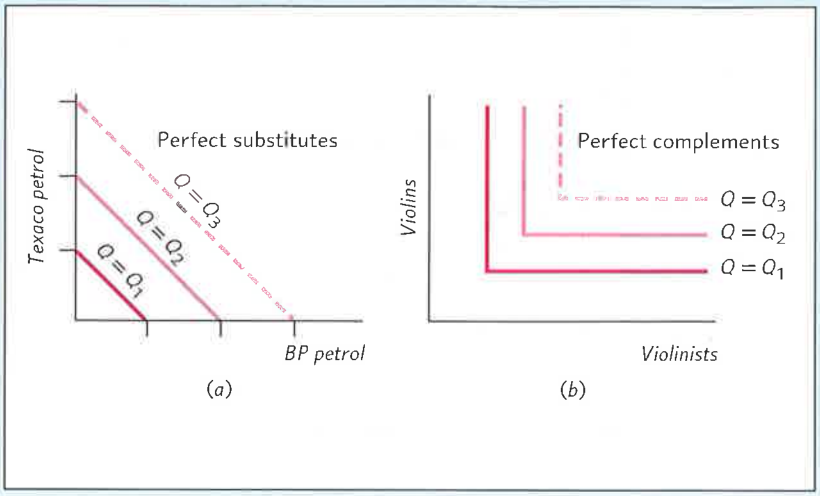
which says that the MRTS at A is simply the ratio of the marginal product of L to the marginal product of K. This relationship will have an important application in the next chapter, where we will take up the question of how to produce a given level of output at the lowest possible cost.

EXERCISE 10.5 Given a firm's current level of capital and labour inputs, the marginal product of labour for its production process is equal to 3 units of output. If the marginal rate of technical substitution between K and L is 9, what is the marginal product of capital?

In consumer theory, the shape of the indifference curve tells us how the consumer is willing to substitute one good for another. In production theory, an essentially similar story is told by the shape of the isoquant. Figure 10.9 illustrates the extreme cases of inputs that are perfect substitutes (a) and perfect complements (b). Figure 10.9(a) describes a production process in which cars and petrol are combined to produce trips. The input of petrol comes in two brands, Texaco and BP, which are perfect substitutes for one another. We can substitute 1 litre of BP petrol for 1 litre of Texaco petrol and still produce the same number of trips as before. The MRTS between Texaco and BP remains constant at 1 as we move downward along any isoquant.

Figure 10.9(b) describes a production process of an orchestra using the two inputs of musicians and musical instruments. In this process, the two inputs are perfect complements. Here, inputs are most effectively combined in fixed proportions. Having more than one violin per violinist doesn't augment production, nor does having more than one violinist per violin.

FIGURE 10.9
Isoquant Maps for Perfect Substitutes and Perfect Complements
In panel (*a*), we get the same number of trips from a given total quantity of petrol, no matter how we mix the two brands. BP and Texaco are perfect substitutes in the production of car trips. In panel (*b*), violinists and violins are perfect complements in the process of producing music.



RETURNS TO SCALE

A question of central importance for the organization of industry is whether production takes place most efficiently at large scale rather than small scale (where 'large' and 'small' are defined relative to the scale of the relevant market). This question is important because the answer dictates whether an industry will end up being served by many small firms or only a few large ones.

The technical property of the production function used to describe the relationship between scale and efficiency is called *returns to scale*. The term tells us what happens to output when all inputs are increased by exactly the same proportion. Because returns to scale refer to a situation in which all inputs are variable, *the concept of returns to scale is an inherently long-run concept.*

increasing returns to scale the property of a production process whereby a proportional increase in every input yields a more than proportional increase in output.

A production function for which any given proportional change in all inputs leads to a more than proportional change in output is said to exhibit **increasing returns to scale**. For example, if we double all inputs in a production function with increasing returns to scale, we get more than twice as much output as before. As we will see in Chapters 13 and 14, such production functions generally give rise to conditions in which a small number of firms supply most of the relevant market.

Increasing returns to scale often result from the greater possibilities for specialization in large organizations. Adam Smith illustrated this point by describing the division of labour in a pin factory:[8]

> One man draws out the wire, another straightens it, a third cuts it, a fourth points it, a fifth grinds it at the top for receiving the head; to make the head requires two or three distinct operations. . . . I have seen a small manufactory . . . of this kind where only ten men were employed . . . [who] could, when they exerted themselves, make among them about twelve pounds of pins in a day. There are in a pound upwards of four thousand pins of middling size. Those ten persons, therefore, could make among them upwards of forty-eight thousand pins in a day. Each person, therefore, making a tenth part of forty-eight thousand pins might be considered as making four thousand eight hundred pins in a day. But if they had all wrought separately and independently . . . they could not each of them have made twenty, perhaps not one pin in a day. . . .

[8] Adam Smith, *The Wealth of Nations*, New York: Everyman's Library, 1910 (1776), Book 1, p. 5.

The airline industry is often cited as one with increasing returns to scale. Industry professionals have long stressed that having a large number of flights helps an airline fill each flight by feeding passengers from its incoming flights to its outgoing flights. Local airport activities also exhibit increasing returns to scale. Because of the law of large numbers, moreover, (see Chapter 7) it follows that maintenance operations, flight crew scheduling and other inventory-related activities are all accomplished more efficiently on a large scale than on a small scale. Similarly, ticket-counter space, ticket agents, reservations equipment, baggage-handling equipment, ground crews and passenger-boarding facilities are all resources that are utilized more efficiently at high activity levels. Increasing returns to scale constitute the underlying explanation for why the industry has been moving toward ever-larger airlines in the last decade.

A production function for which a proportional change in all inputs causes output to change by the same proportion is said to exhibit **constant returns to scale**. In such cases, doubling all inputs results in a doubling of output. In industries in which production takes place under constant returns to scale, large size is neither an advantage nor a disadvantage.

constant returns to scale the property of a production process whereby a proportional increase in every input yields an equal proportional increase in output.

Finally, a production function for which a proportional change in all inputs causes a less than proportional change in output is said to exhibit **decreasing returns to scale**. Here, large size is a handicap, and we do not expect to see large firms in an industry in which production takes place with decreasing returns to scale. As we will see in Chapter 12, the constant and decreasing returns cases often enable many sellers to coexist within the same narrowly defined markets.

decreasing returns to scale the property of a production process whereby a proportional increase in every input yields a less than proportional increase in output.

A production function need not exhibit the same degree of returns to scale over the entire range of output. On the contrary, there may be increasing returns to scale at low levels of output, followed by constant returns to scale at intermediate levels of output, followed finally by decreasing returns to scale at high levels of output.

ECONOMIC NATURALIST 10.4

Why do builders use prefabricated frames for roofs but not for walls?

When construction crews build a wood-frame house, they usually construct framing for the walls at the construction site. By contrast, they often buy prefabricated framing for the roof. Why this difference?

There are two key differences between wall framing and roof framing: (1) cutting the timber for roof framing involves many complicated angle cuts, whereas the right-angle cuts required for wall framing are much simpler; and (2) sections of roof framing of a given size are all alike, whereas wall sections differ according to the placement of window and door openings. Both properties of roof framing lead to substantial economies of scale in production. First, the angle cuts they require can be made much more rapidly if a frame or 'jig' can be built that guides the lumber past the saw-blade at just the proper angle. It is economical to set up such jigs in a factory where thousands of cuts are made each day, but it usually does not pay to use this method for the limited number of cuts required at any one construction site. Likewise, automated methods are easy to employ for roof framing by virtue of its uniformity. The idiosyncratic nature of wall framing, by contrast, militates against the use of automated methods.

© Kevin Britland

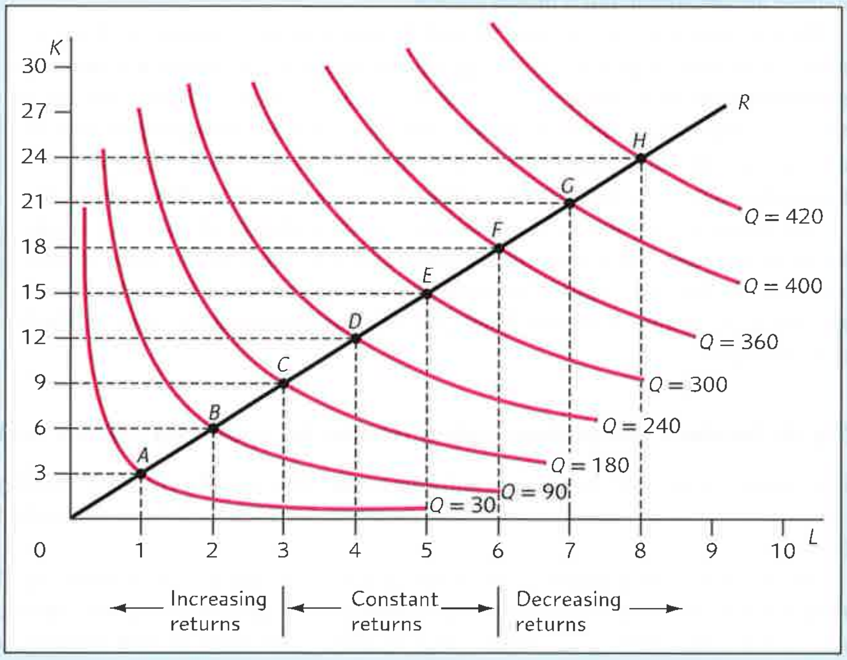
Why do builders build custom frames for walls but use prefabricated frames for roofs?

So the fact that there are much greater economies of scale in the construction of roof framing than wall framing helps account for why wall framing is usually built at the construction site while roof framing is more often prefabricated. ■

Showing Returns to Scale on the Isoquant Map

A simple relationship exists between a production function's returns to scale and the spacing of its isoquants.[9] Consider the isoquant map in Figure 10.10. As we move outward into the isoquant map along the ray labelled *R*, each input grows by exactly the same proportion. The particular production function whose isoquant map is shown in the diagram exhibits increasing returns to scale in the region from *A* to *C*. Note, for example, that when we move from *A* to *B*, both inputs double while output goes up by a factor of 3; likewise, when we move from *B* to *C*, both inputs grow by 50 per cent while output grows by 100 per cent. In the region from *C* to *F*, this same production function exhibits constant returns to scale. Note, for example, that when we move from *D* to *E*, both inputs grow by 25 per cent and output also grows by 25 per cent. Finally, the production function whose isoquant map is shown in Figure 10.10 exhibits decreasing returns to scale in the region to the northeast of *F*. Thus, when we move from *F* to *G*, both inputs increase by 16.7 per cent while output grows by only 11.1 per cent.

FIGURE 10.10
Returns to Scale Shown on the Isoquant Map
In the region from *A* to *C*, this production function has increasing returns to scale. Proportional increases in input yield more than proportional increases in output. In the region from *C* to *F*, there are constant returns to scale. Inputs and output grow by the same proportion in this region. In the region northeast of *F*, there are decreasing returns to scale. Proportional increases in both inputs yield less than proportional increases in output.



The Distinction between Diminishing Returns and Decreasing Returns to Scale

It is important to bear in mind that decreasing returns to scale have nothing whatsoever to do with the law of diminishing returns. Decreasing returns to scale refer to what happens when *all* inputs are varied by a given proportion. The law of diminishing returns, by contrast, refers to the case in which one input varies while all others are held fixed. As an empirical generalization, it applies with equal force to production functions having increasing, constant or decreasing returns to scale.

To illustrate the difference, consider again an airline company. If the company increases the number of pilots it employs while leaving unchanged everything else there will clearly come a point of diminishing returns to pilots. This does not, in any way, rule out increasing returns to scale. If we increase the number of pilots *and* the number of planes, ticket agents, etc., then there may be increasing returns to scale.

[9]The discussion in this section applies to *homothetic* production functions, an important class of production functions defined by the property that the slopes of all isoquants are constant at points along any ray.

The Logical Puzzle of Decreasing Returns to Scale

If the production function $Q = F(K, L)$ is a complete description of the corresponding production process, it is difficult to see how any production function could ever exhibit decreasing returns to scale in practice. The difficulty is that we ought to be able to duplicate the process used to produce any given level of output, and thereby achieve constant returns to scale. To illustrate, suppose first that $Q_0 = F(K_0, L_0)$. If we now want to produce $2Q_0$ units of output, we can always do so by again doing what we did the first time—namely, by again combining K_0 and L_0 to get Q_0 and adding that to the Q_0 we already have. Similarly, we can get $3Q_0$ by carrying out $F(K_0, L_0)$ three times in succession. Simply by carrying out the process again and again, we can get output to grow in the same proportion as inputs, which means constant returns to scale. And for reasons similar to the ones discussed above for the airline industry, it will often be possible to do even better than that.

In cases in which it is not possible to at least double our output by doubling both K and L, we seem forced to conclude that there must be some important input besides K and L that we are failing to increase at the same time. This input is variously referred to as 'organization' or 'communication', the idea being that when a firm gets past a certain size, it somehow starts to get out of control. Others claim that it is the shortage of managerial or entrepreneurial resources that creates bottlenecks in production. If there is indeed some unmeasured input that is being held fixed as we expand K and L, then we are still in the short run by definition. And there is no reason to expect to be able to double our output by doubling only *some* of our inputs.

SUMMARY

- Production is any activity that creates current or future utility. A production function summarizes the relationship between inputs and outputs. The short run is defined as that period during which at least some inputs are fixed. In the two-input case, it is the period during which one input is fixed, the other variable.

- The marginal product of a variable input is defined as the change in output brought forth by an additional unit of the variable input, all other inputs held fixed. The law of diminishing returns says that beyond some point the marginal product declines with additional units of the variable input.

- The average product of a variable input is the ratio of total output to the quantity of the variable input. Whenever marginal product lies above average product, the average product will increase with increases in the variable input. Conversely, when marginal product lies below average product, average product will decline with increases in the variable input.

- An important practical problem is that of how to allocate an input across two productive activities to generate the maximum possible output. In general, two types of solution are possible. A corner solution occurs when the marginal product of the input is always higher in one activity than in the other. In that case, the best thing to do is to concentrate all the input in the activity where it is more productive.

- An interior solution occurs whenever the marginal product of the variable input, when all of it is placed in one activity, is lower than the marginal product of the first unit of the input in the other activity. In this case, the output-maximizing rule is to distribute the input across the two activities in such a way that its marginal product is the same in both. Even experienced decision makers often violate this simple rule. The pitfall to be on guard against is the tendency to equate not marginal but average products in the two activities.

- The long run is defined as the period required for all inputs to be variable. The actual length of time that corresponds to the short and long runs will differ markedly in different cases. In the two-input case, all of the relevant information about production in the long run can be summarized graphically by the isoquant map. The marginal rate of technical substitution is defined as the rate at which one input can be substituted for another without altering the level of output. The MRTS at any point is simply the absolute value of the slope of the isoquant at that point. For most production functions, the MRTS will diminish as we move downward to the right along an isoquant.

- A production function is said to exhibit constant returns to scale if a given proportional increase in all inputs produces the same proportional increase in output, decreasing returns to scale if a given proportional increase in all inputs results in a smaller proportional increase in output, and increasing returns to scale if a given proportional increase in all inputs causes a greater proportional increase in output. Production functions with increasing returns to scale are also said to exhibit economies of scale. Returns to scale constitute a critically important factor in determining the structure of industrial organization.

APPENDIX

10

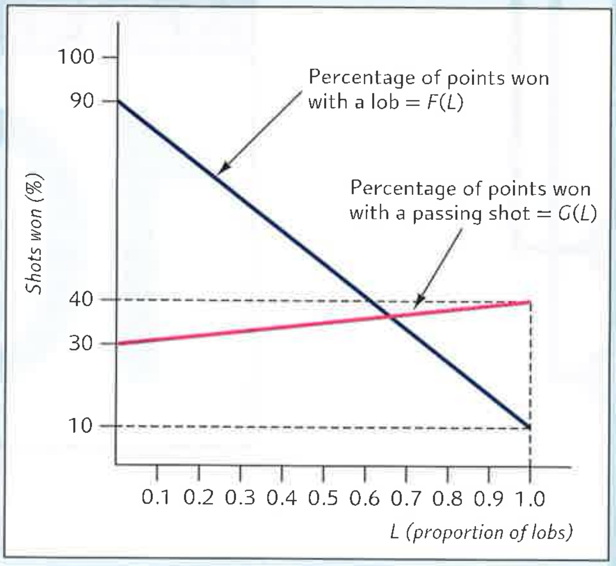
MATHEMATICAL EXTENSIONS OF PRODUCTION THEORY

THE AVERAGE–MARGINAL DISTINCTION

Calculus allows a fuller understanding of the average–marginal distinction. Here are three examples to illustrate.

Suppose that when your tennis opponent comes to the net, your best response is either to lob (hit the ball over his head) or to pass (hit the ball out of reach on either side). Each type of shot is more effective if it catches your opponent by surprise. Suppose someone who lobs all the time will win a given point only 10 per cent of the time with a lob, but that someone who virtually never lobs wins the point on 90 per cent of the rare occasions when she does lob. Similarly, suppose someone who tries passing shots all the time wins any given point only 30 per cent of the time with a passing shot, but someone who virtually never tries to pass wins 40 per cent of the time when she does try. Suppose, finally, that the rate at which each type of shot becomes less effective with use declines linearly

FIGURE A.10.1
Effectiveness versus Use: Lobs and Passing Shots



with the proportion of times a player uses it. What is the best proportion of lobs and passing shots to use when your opponent comes to the net?[1]

The payoffs from the two types of shot are summarized graphically in Figure A.10.1. Here, the 'production' problem is to produce the greatest possible percentage of winning shots when your opponent comes to the net. Let $F(L)$ be the percentage of points you will win with a lob as a function of the proportion of times you lob (L). $F(L)$ is thus, in effect, the average product of L. Let $G(L)$ be the percentage of points you will win with a passing shot, again as a function of the proportion of times you lob. The negative slope of $F(L)$ reflects the fact that lobs become less effective the more you use them. Similarly, the positive slope of $G(L)$ says that passing shots become more effective the more you lob. Your problem is to choose L^*, the best proportion of times to lob.

To find the optimal value of L, we must first discover how the percentage of total points won, denoted P, varies with L. For any value of L, P is simply a weighted average of the percentages won with each type of shot. The weight used for each type of shot is simply the proportion of times it is used. Noting that $(1 - L)$ is the proportion of passing shots when L is the proportion of lobs, we have

$$P = LF(L) + (1 - L)G(L) \tag{A.10.1}$$

The expression $LF(L)$ is the percentage of total points won on lobs. $(1 - L)G(L)$, similarly, is the percentage of total points won on passing shots. From Figure A.10.1, we see that the algebraic formulas for $F(L)$ and $G(L)$ are given by $F(L) = 90 - 80L$ and $G(L) = 30 + 10L$. Substituting these relationships into Equation A.10.1 gives

$$P = 30 + 70L - 90L^2 \tag{A.10.2}$$

which is plotted in Figure A.10.2. To find the optimum we need to find the value of L that maximizes P. Differentiating P with respect to L and setting equal to zero gives

$$\frac{dP}{dL} = 70 - 180L = 0. \tag{A.10.3}$$

[1]This example was suggested by Harvard psychologists Richard Herrnstein and James Mazur, in 'Making Up our Minds: A New Model of Economic Behavior', *The Sciences*, November/December 1987: 40–47.

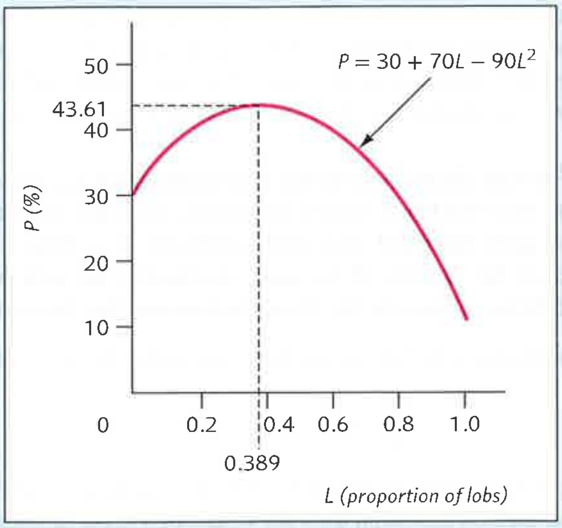


FIGURE A.10.2
The Optimal Proportion of Lobs

The value of L that maximizes P turns out to be $L^* = 0.389$, and the corresponding value of P is 43.61 per cent.

Note in Figure A.10.3 that at the optimal value of L, the likelihood of winning with a lob is almost twice as high (58.9 per cent) as that of winning with a passing shot (33.9 per cent). Many people seem to find this state of affairs extremely uncomfortable—so much so that they refuse to have anything to do with it. Instead they try to equate the *average product* of each type. Note in Figure A.10.3 that this occurs when $L = 2/3$, at which point the percentage of points won with either shot is 36.7. At this value of L, however, the *marginal product* of a passing shot will be much higher than for a lob, because it will strongly increase the effectiveness of all your *other* lobs. (Of course, an extra passing shot will also reduce the effectiveness of your other passing shots, but by a much smaller margin.)

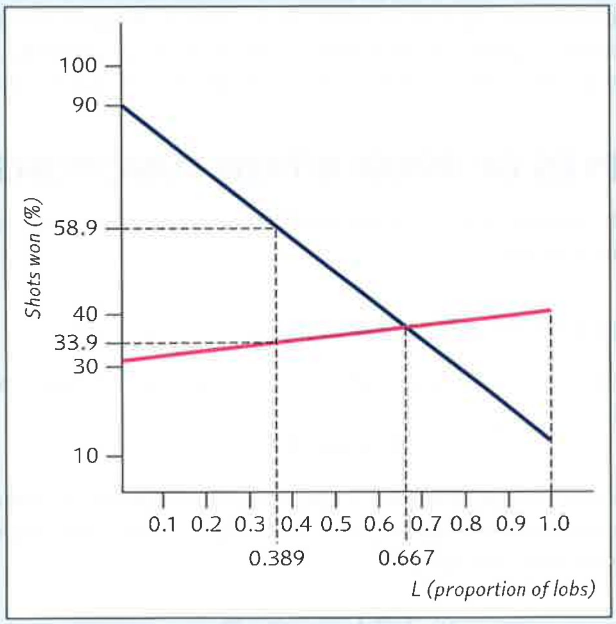


FIGURE A.10.3
At the Optimizing Point, the Likelihood of Winning with a Lob is Much Greater Than of Winning with a Passing Shot

The situation here is analogous to the allocation example involving the fishing boats mentioned in Chapter 10. There is no more reason to want the average return to each tennis shot to be the same than there is to want the average product on each end of the lake to be the same. And yet the tendency to equate average rather than marginal products is a very common pitfall, one that even experienced maximizers have to be on guard against. Let us consider two more final examples.

EXAMPLE A.10.1 Farmer Giles has been commissioned to produce 100 bottles of cider. There are two distinct ways in which he can collect the apples to make his cider. If he uses method 1 he will produce $Q_1 = 30L_1^{1/2}$ bottles where L_1 is the amount of hours he works. If he uses method 2 he will produce $Q_2 = 5L_2$ bottles. How should Giles allocate his time between the two methods?

We need to equate the marginal product of labour on method 1 with that of method 2. That is, we need

$$\frac{dQ_1}{dL_1} = \frac{15}{L_1^{0.5}} = \frac{dQ_2}{dL_2} = 5. \tag{A.10.4}$$

This implies that $L_1 = 9$. If $L_1 = 9$, then he produces 90 bottles using method 1. That leaves 10 bottles to be produced using method 2. Setting $Q_2 = 10$, we get $L_2 = 2$. So, he should spend 9 hours on method 1 and 2 hours on method 2.

Note that at the optimum the average product with method 1 is 90/9 = 10 bottles per hour. The average product with method 2 is only 5 bottles per hour. ◆

EXERCISE A.10.1 Same as Example A.10.1 except that Giles only needs to produce 80 bottles of cider.

EXAMPLE A.10.2 Suppose that there are two supermarkets in town. You are the manager of one of them. Each Monday morning you must decide prices for the following week. You can either price aggressively, by offering discounts, or non-aggressively. An analyst tells you that in weeks you price aggressively you make an average profit of €20,000. In weeks you price non-aggressively you make an average profit of €30,000. *True or false*: you should price non-aggressively each week?

If you answered 'true', you have not been paying attention. We need to know how often the supermarket has been pricing aggressively, and we need to know how a change in this proportion would affect profitability. Pricing aggressively less often may lower the profitability, not only of the additional weeks where non-aggressive pricing is used, but all other weeks where it is used. ◆

SOME EXAMPLES OF PRODUCTION FUNCTIONS

In this section we will examine two of the many different production functions that are commonly used in economic analysis.

The Cobb–Douglas Production Function

Perhaps the most widely used production function of all is the Cobb–Douglas, which in the two-input case takes the form

$$Q = mK^{\alpha}L^{\beta} \tag{A.10.5}$$

where α and β are numbers between zero and 1, and m can be any positive number.

To generate an equation for the Q_0 isoquant, we fix Q at Q_0 and then solve for K in terms of L. In the Cobb–Douglas case, this yields

$$K = \left(\frac{m}{Q_0}\right)^{-1/\alpha} (L)^{-\beta/\alpha} \tag{A.10.6}$$

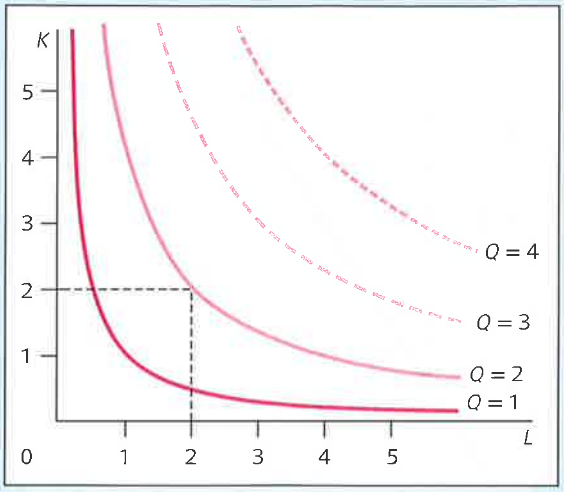


FIGURE A.10.4
Isoquant Map for the Cobb–Douglas Production Function $Q = K^{1/2} L^{1/2}$

For the particular Cobb–Douglas function $Q = K^{1/2} L^{1/2}$, the Q_0 isoquant will be

$$K = \frac{Q_0^2}{L} \tag{A.10.7}$$

A portion of the isoquant map for this particular Cobb–Douglas production function is shown in Figure A.10.4.

The number assigned to each particular isoquant in Figure A.10.4 is exactly the level of output to which it corresponds. For example, when we have 2 units of K and 2 units of L, we get $Q = \sqrt{2}\,\sqrt{2} = 2$ units of output. Recall from Chapter 4 that the numbers we used to label the indifference curves on an indifference map conveyed information only about *relative* levels of satisfaction. All that was required of our indexing scheme in that context was that the *order* of the numbers we assigned to the indifference curves reflects the proper ranking of the corresponding satisfaction levels. With isoquants, the situation is altogether different. We have, in effect, no choice about what labels to assign to them.

You can easily verify the following expressions for the marginal products of labour and capital in the Cobb–Douglas case:

$$\mathrm{MP}_K = \frac{\partial Q}{\partial K} = \alpha\, mK^{\alpha-1} L^{\beta} \tag{A.10.8}$$

and

$$\mathrm{MP}_L = \frac{\partial Q}{\partial L} = \beta\, mK^{\alpha} L^{\beta-1} \tag{A.10.9}$$

This means (recall that α and β are numbers between zero and one) that there are diminishing returns to labour and capital. The marginal rate of technical substitution is given by

$$\frac{\mathrm{MP_L}}{\mathrm{MP_K}} = \frac{\beta mK^{\alpha} L^{\beta-1}}{\alpha mK^{\alpha-1} L^{\beta}} = \frac{\beta K}{\alpha L}. \tag{A.10.10}$$

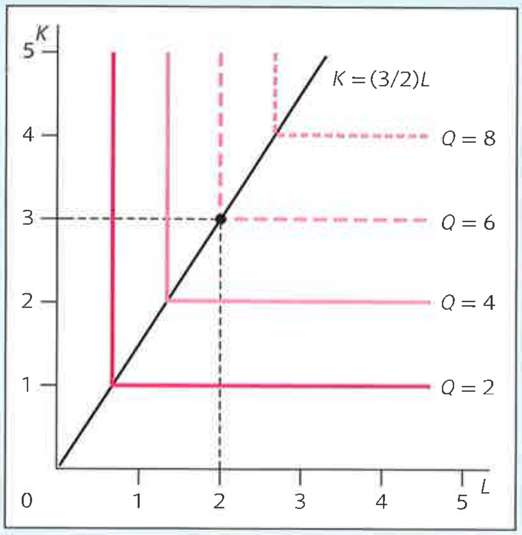
The Leontief, or Fixed-Proportions, Production Function

The simplest among all production functions that are widely used is the *Leontief*, named for the Nobel laureate Wassily Leontief, who devised it. For the two-input case, it is given by

$$Q = \min(aK, bL) \tag{A.10.11}$$

If you are unfamiliar with this curious functional form, its interpretation is simply that Q is equal to either aK or bL, whichever is smaller. Suppose, for example, that $a = 2$, $b = 3$, $K = 4$ and $L = 3$. Then, $Q = \min(2 \times 4, 3 \times 3) = \min(8, 9) = 8$. The isoquant map for $Q = \min(2K, 3L)$ is shown in Figure A.10.5.

FIGURE A.10.5
Isoquant Map for the Leontief Production Function $Q = \min(2K, 3L)$



To see why the Leontief is also called the fixed-proportions production function, note first in Figure A.10.5 that if we start with 3 units of K and 2 units of L, we get 6 units of output. If we then add more L—so that we have, say, 3 units of L instead of 2—we still get only $Q = \min(2 \times 3, 3 \times 3) = \min(6, 9) = 6$ units of output. By the same token, adding more K when we are at $K = 3$ and $L = 2$ will not lead to any additional output. In the Leontief case, K and L are used most effectively when $aK = bL$—in the example at hand, when $2K = 3L$. In Figure A.10.5, the locus of points for which $2K = 3L$ is shown as the ray $K = (3/2)L$. It is along this ray that the cusps of all the right-angled isoquants of this Leontief production function will lie.

Recall from Chapter 4 that in the case of perfect complements, the indifference curves had the same right-angled shape as the isoquants for the Leontief production function. This meant that the MRS was infinite on the vertical arm of the indifference curve, zero on the horizontal arm, and undefined at the cusp. For exactly parallel reasons, the MRTS in the Leontief case will be infinite on the vertical arm of the isoquant, zero on the horizontal and undefined at the cusp.

A MATHEMATICAL DEFINITION OF RETURNS TO SCALE

Mathematically, to increase all inputs in the same proportion means simply to multiply all inputs by the same number $c > 1$. By way of illustration, consider the production function we discussed in Chapter 10, $Q = F(K, L) = 2KL$. For this particular function, when we multiply each input by c we get

$$F(cK, cL) = 2(cK)(cL) = c^2 2KL = c^2 F(K, L) \qquad \text{(A.10.12)}$$

The result of multiplying each input by c in this production function is thus to multiply the original output level by c^2. Output thus grows more than in proportion to input growth in this case (with proportional growth, we would have had output equal to $cF(K, L)$), so this production function has increasing returns to scale. Thus, for example, if $c = 2$ (a doubling of each input), we get $F(2K, 2L) = 2(2K)(2L) = 4(2KL)$, a quadrupling of output.

Drawing on these observations, the definitions of our three cases may be summarized as follows (for any K, L and $c > 1$):

$$\text{Increasing returns: } F(cK, cL) > cF(K, L) \qquad \text{(A.10.13)}$$

$$\text{Constant returns: } F(cK, cL) = cF(K, L) \qquad \text{(A.10.14)}$$

and

$$\text{Decreasing returns: } F(cK, cL) < cF(K, L) \qquad \text{(A.10.15)}$$

The following two exercises will help cement your ability to apply these definitions to specific examples.

EXERCISE A.10.2 Does the production function $Q = \sqrt{K}\sqrt{L}$ have increasing, constant or decreasing returns to scale?

EXERCISE A.10.3 Does the production function $Q = K^{1/3} L^{1/3}$ have increasing, constant or decreasing returns to scale?

In the case of the Cobb–Douglas production function, $Q = mK^{\alpha} L^{\beta}$, Equations A.10.13 to A.10.15 imply a simple relationship between the parameters α and β and the degree of returns to scale. Specifically, if $\alpha + \beta > 1$, there are increasing returns to scale; $\alpha + \beta = 1$ means constant returns to scale; and $\alpha + \beta < 1$ means decreasing returns to scale. To illustrate for the constant returns case, suppose $Q = F(K, L) = mK^{\alpha}L^{\beta}$, with $\alpha + \beta = 1$. Then we have

$$F(cK, cL) = m(cK)^{\alpha}(cL)^{\beta} \qquad \text{(A.10.16)}$$

which reduces to

$$c^{(\alpha+\beta)}\, mK^{\alpha} L^{\beta} = cmK^{\alpha} L^{\beta} = cF(K, L) \qquad \text{(A.10.17)}$$

which, by Equation A.10.14, is the defining characteristic of constant returns to scale.

We finish with an example that illustrates the distinction between diminishing returns and decreasing returns to scale.

EXAMPLE A.10.3 Does the production function $Q = F(K, L) = K^{0.6}L^{0.5}$ have increasing, decreasing or constant returns to scale? Does it have diminishing returns to labour and capital?

This is a Cobb–Douglas production function with increasing returns to scale. The marginal product of capital is

$$MP_K = \frac{\partial Q}{\partial K} = 0.6\,\frac{L^{0.5}}{K^{0.4}}. \qquad \text{(A.10.18)}$$

This implies diminishing returns to capital as the bigger is K the smaller is the marginal product. The marginal product of labour is

$$MP_L = \frac{\partial Q}{\partial K} = 0.5\,\frac{K^{0.6}}{L^{0.5}}. \qquad \text{(A.10.19)}$$

So, we also have diminishing returns to labour. ◆

PROBLEMS

*1. Do the following production functions have increasing, decreasing or constant returns to scale? Which ones fail to satisfy the law of diminishing returns?

a. $Q = 4K^{1/2} L^{1/2}$
b. $Q = aK^2 + bL^2$
c. $Q = \min(aK, bL)$
d. $Q = 4K + 2L$
e. $Q = K^{0.5} L^{0.6}$
f. $Q = K_1^{0.3} K_2^{0.3} L^{0.3}$

*2. What is the marginal product of labour in the production function $Q = 2K^{1/3} L^{1/3}$ if K is fixed at 27?

3. Can the Cobb–Douglas production function be used to portray a production process in which returns to scale are increasing at low output levels and are constant or decreasing at high output levels?

*This problem is most easily solved using the calculus definition of marginal product.