Module 5

Production analysis

Introduction

This module moves from the decisions of utility-maximising households to examine the factors governing the behaviour of perfectly competitive profit-maximising producers. It is a transition module that introduces many central topics, among which are productivity, costs and economic profits. The most profitable way to employ the firm's resources to produce a given product requires an understanding of production function. Production function is simply an input-output relationship between one or more factors of production (input) and the good or service produced (output). This relationship is analysed and quantified during a production study to determine the most economical combination of input resources to obtain a given level of output. A study of production functions is also fundamental to cost analysis. Once production function has been identified, its cost function can be derived from the production function. Hence the production affects the firm's status in its industry.

Upon completion of this module you will be able to:

Outcomes

- *outline* the major decisions that firms must take.
- *distinguish* the short run from the long run.
- *describe* the relationship between average product, marginal product and total product.
- *apply* the Iso-quant approach to production.
- *distinguish* (diminishing) returns to factor from (diminishing) returns to scale.
- *describe* the rule of costs minimisation.

Terminology

Long-run: All factors may be varied.

Production function: An input-output relationship between one or more factors of production (input) and the good or service produced (output). It is an equation or table or graph showing the amount of output of a product that a firm can produce per period of time for each set of inputs. Both inputs and outputs are measured in physical rather than in monetary units.

Short-run: The quantities of some inputs are variable while others are in fixed supply.

Production analysis

Once demand for a given product or service has been determined, management decides the most profitable way to employ the firm's resources to produce that good or service. Such decisions involve an understanding of production functions.

A production function is simply an input-output relationship between one or more factors of production (input) and the good or service produced (output). This relationship is analysed and quantified during a production study to determine the most economical combination of input resources to obtain a given level of output. Or conversely, it may involve the determination of the maximum output obtainable from a given level and mix of inputs. A production study may pertain not only to the production of goods, such as planes, computer games, scanners, ice cream, but also to the production of services, such as Internet services, hospitality industry or health care.

The production function

Production function is an equation, table, or graph showing the amount of output of a product that a firm can produce per period of time for each set of inputs. Both inputs and outputs are measured in physical rather than in monetary units. Technology is assumed to remain constant during the period of the analysis. Technology summarises the feasible means (know-how) of converting raw inputs, such as steel, labour and machinery, into an output such as a car.

A study of production functions is also fundamental to cost analysis. Once a firm's production function has been identified, its cost function can be derived from the production function, provided the market prices of the input factors are known. Hence, the production function strongly affects the firm's status in its industry, and the study of production functions is even more basic than the study of cost functions.

Like a demand function, a production function can be expressed as a schedule, graph, or an equation such as

$$Q = f(K, L) \tag{1}$$

where Q, a specific output, is a function of the input factors, K and L. Note that the general ideas presented here are valid for any two inputs, not just for labour and capital. It is important to remember that a production function relates to some given level of technology. If the technology changes through the upgrading of labour, materials, machinery, equipment, processes, or management, the production function changes accordingly.

The short run versus the long run

In production and cost theory, the distinction is made between the *short run,* in which the quantities of some inputs are variable while others are in fixed supply, and the *long run* in which all factors may be varied. Consequently, it is useful to classify the inputs on the basis of whether or not they are variable in the short run. Labour is typically regarded as variable and capital as fixed in the short run.

It is important to understand that *the long run* does not refer to a long period of time. It is a peculiarity of the economists' jargon that the term has no direct connection with time at all, and that the firm is likely to be in a long-run situation for relatively short periods. When intending to change its scale of production, the firm must continue to operate in a short-run situation until its most-fixed factor becomes variable.

Production in the short run

As a manager, your job is to use the available production function efficiently. This effectively means that you must determine how much of each input to use to produce output. In the short run, some factors of production are fixed, and this limits your choices in making input decisions. For example, it takes several years for Honda Motor Company to build an assembly line. The level of capital is generally fixed in the short run. However, in the short run, Honda can adjust its use of inputs such as labour and steel; such inputs are called variable factors of production.

Therefore, the short run by definition is that period of time during which at least one input is fixed. Variable inputs are those inputs whose quantities are directly related to the level of production.

Mathematically, this function is captured by the following relationship:

$$Q = f(\overline{K}, L) \quad (2)$$

where, $\overline{K}$, refers to the fixed level of input of capital. The horizontal bar in this equation indicates that the input factors to its left are regarded as fixed in the production process under analysis, while the factor to its right is variable. The quantity of the output product, Q, is the result of combining a variable quantity of input factor, (skilled labour), with fixed quantities of other input (buildings or equipment).

Measures of productivity

The productivity of a factor of production refers to the amount of output that can be produced by that input, holding constant the input of all other factors of production. Obviously, an input, such as human resources, can be more productive if it works with modern mechanical and computerassisted equipment and high-quality raw materials. Similarly, the plant or equipment can be more productive if it is being operated by highly skilled and well-trained workers. We use the phrase *the state of technology* when we refer to the quality of the resources involved in the production. Table 5-1 is a numerical representative of this function.

Table 5-1 Alternative short-run production functions

Row 1	Capital inputs (machine month)	Labour inputs (person month) (L)								
2	(K)	0	1	2	3	4	5	6	7	8
3	1	0	15	35	77	112	130	145	155	162
4	2	0	22	48	80	125	161	188	208	223
5	3	0	35	72	120	180	255	310	358	387
6	4	0	500	102	162	230	305	385	444	493

Note that this table in fact represents four short-run production functions and not just one. Column 2 shows the amount of capital used per month. Accordingly, there are four possible *fixed* sizes of capital, varying from 1 to 4, $\overline{K}=1$, $\overline{K}=2$, $\overline{K}=3$ and $\overline{K}=4$, with 1 being the smallest and 4 the largest plant. Furthermore, each $\overline{K}$ is combined with the variable input of labour, varying from 0, column 3, to 8, the last column. The shaded area, rows 3, 4, 5, and 6, shows the level of output (total product). For example, row 3 represents a production function that combines 1 K with different Ls, 0 to 8, while row 4 represents a different production function that combines 2 units of K (a bigger plant) with labour, 0 to 8 units. Clearly, the manager has several plant options.

Demonstration problem

Referring to Table 5-1, suppose the firm is currently producing 160 units of output and using a production function that employs 2 K. Now suppose that the demand for the firm's output increases permanently to, say 240, with a possibility of a further 10 per cent increase in the following year. How many ways can this firm meet the market demand?

Answer:

The firm can produce 240 units of output using its current plant by stretching it almost to its limit. This will be done by combing 2K with 8L. Alternatively, the firm can produce the same output by expanding its plant (K) using a production function that combines either 3K (the next plant size) with 5L, or 4K with 4L. However, when we consider the possibility of the additional 10 per cent increase in demand, the manager is left with only two choices, plant 3, $\overline{K}=3$, and plant 4, $\overline{K}=4$.

Law of diminishing returns

Table 5-2, extracted from Table 5-1, shows that, in a given state of technology and keeping the input of capital constant, $\overline{K} = 2$, additional units of labour (the variable input) yield increasing output per unit of input up to a point. But eventually a point is reached beyond which further additions of labour yield diminishing returns per unit of input. In this illustration, the point of diminishing returns is reached at an input of five units of L. Put differently, the relative productivity of the marginal units of the variable inputs (L) initially increases (up to L = 4) and diminishes subsequently. Hence, the initial increasing returns to the variable input gives way to diminishing returns.

These, which are short-term phenomena, can be stated more broadly as the law of variable proportions. They state that as more and more of the variable factors are added to a given quantity of the fixed factors, the increment to output attributable to each of the additional units of the variable factor will increase at first, will later decrease, and will eventually become negative.

Table 5-2 The law of variable proportions

Units of input (L)	Units of output (for $\overline{K} = 2$)	Increment to output	Returns to the variable input
0	0	-	-
1	22	22	Increasing
2	48	26	Increasing
3	80	32	Increasing
4	125	45	Increasing
5	161	36	Diminishing
6	188	27	Diminishing
7	208	20	Diminishing
8	223	15	Diminishing

Total product, average product and marginal product

Total product (TP) is simply the maximum level of output that can be produced with a given amount of inputs, Column 2, in Table 5-2.

Average Product (AP_L), the most popular measure of productivity, is defined as total product divided by the quantity of the variable input (L) used, $AP_L = Q/L$.

Marginal Product (MP_L) is the change in total output attributable to the last unit of the input of labour, $MP_L = \frac{\Delta Q}{\Delta L}$.

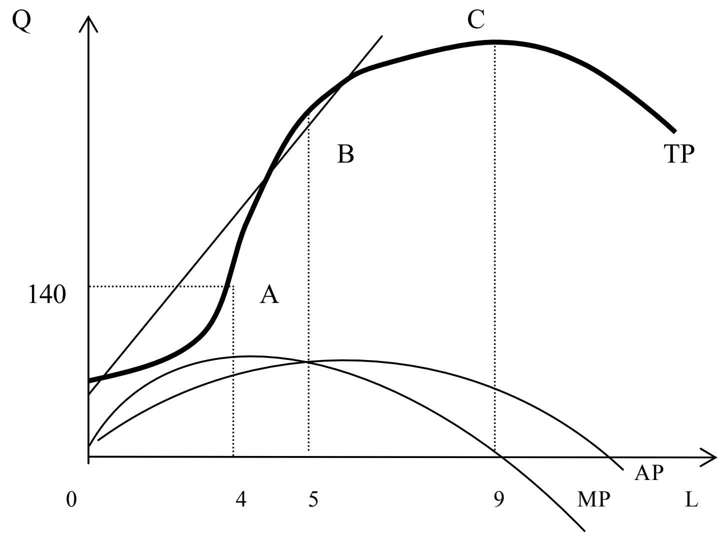
Note that the average and the marginal product concepts can also be defined for the fixed input of capital. Accordingly, AP_K, the average product of capital, is defined as total product divided by the quantity of the fixed input (K), whereas the marginal product of capital *(MP_K)* is the change in total output divided by the change in capital, $MP_K = \frac{\Delta Q}{\Delta L}$.

Figure 5-1 shows graphically the relationship among total product, marginal product and average product. The first thing to notice about these curves is that as the use of labour increases between points 0 and A, the slope of the total product curve increases (becomes steeper). That is, output increases at an increasing rate. Since the slope of the TP curve is $MP_{L,}$ the marginal product curve will be increasing over this range. The range over which marginal product increases is known as the range of *increasing marginal returns.*

Figure 5-1 also shows that marginal product reaches its maximum at point A, where four units of labour are employed. This is indicated on the TP curve by the inflection point, point A. That is, the point at which the curve changes from concave upward to concave downward. As the usage of labour increases from the fourth through to the eighth unit, total output increases, but at a decreasing rate. This is why marginal product declines between four and eight units of labour but is still positive. The range over which marginal product is positive but declining is known as the range of *diminishing marginal returns* to the variable input.

Marginal product becomes negative when more than nine units of labour are employed. After a point, using additional units of input actually reduces total product, which is what it means for marginal product to be negative. The range over which marginal product is negative is known as the range of *negative marginal returns.*

Figure 5-1



The relationship between marginal product and total product, exhibited in Figure 5-1, can be summarised as:

1. As long as the MP_L curve is rising, the total product curve increases at an increasing rate and is convex to the horizontal axis.
2. The quantity of input L at which the TP curve changes its curvature corresponds precisely to the point at which the MP_L curve peaks. This occurs at approximately four units of input, as shown by point A.
3. When the total-product curve reaches point C, its maximum MP_L is zero. Beyond this point, the marginal product is negative and the total product declines.

The law of variable proportions discussed above can be also discussed in the context of the average revenue. The average product of labour, AP_L, can be extracted from Table 5-2 as follows:

L	AP_L
0	0
1	22
2	24
3	26.67
4	31.25
5	32.2
6	31.33
7	29.7
8	27.87

AP_L is initially subject to the law of increasing returns, reaching a maximum at approximately 5 units of labour. Beyond that, it becomes subject to the law of diminishing returns. This relationship is also exhibited in Figure 5-1. To sum up:

1. The average product increases as variable input increases as long as the marginal product exceeds the average product.
2. When the marginal product is less than the average product, the average product decreases as variable input increases.
3. When the average product is at a maximum, the average product and marginal product are equal.

Figure 5-1 also illustrates the three stages of a typical production function.

Stage 1: This stage extends from zero input of the variable factor to the level of input where the average product is at a maximum. In this stage, the fixed factors are excessive relative to the variable input. Consequently, output can be increased by increasing the variable input relative to the fixed input.

Stage 2: This stage extends from the end of Stage 1 (the point where the marginal product and the average product are equal) to the point where the marginal product is zero and the total product is at a maximum, point C in Figure 5-1. Stage 2 is a rational stage in which relatively good balance has been achieved between the variable and fixed inputs.

Stage 3: In this stage, in which input is greater than nine units, the variable-input factor is excessive relative to the fixed factors, the marginal product is negative, and the total product is falling. It is completely irrational to produce in this stage.

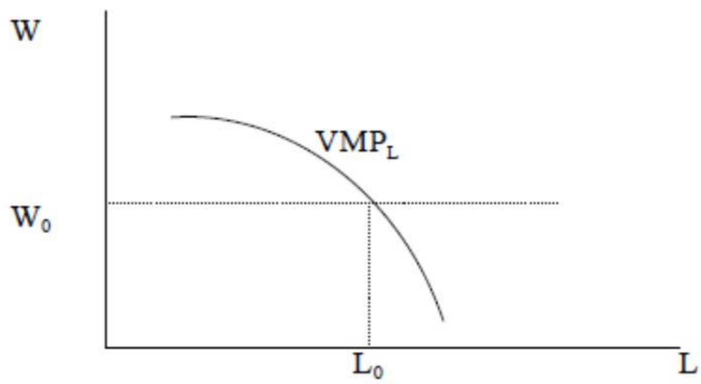
The optimal level of employment of a variable input

To determine the optimum input-output relationship for maximum profit, we must shift the analysis from physical input-output relationships to economic relationships. The precise amount of the input variable needed to produce maximum profit depends upon the revenue and cost associated with hiring an additional unit of the variable input, where the cost side depends on the price of the input variable, while the revenue side is a function of marginal product of the input variable as well as the selling price of the output.

The additional revenue associated with hiring an extra unit of the input of labour equals the value of output produced by this additional hiring. Being referred to as the *value of marginal product (VMP)*, this is defined as MP_L x P, where $MP_L = \Delta Q/\Delta L$ and P is the price of the product sold. On the other hand, the additional cost of hiring is the wage rate, W. Therefore, optimality requires that:

$MP_L \times P = W$ (3)

Figure 5-2



Recalling that the rational stage of production is stage 2, where MP_L is diminishing (Figure 5-2), the firm should hire labour up to a point where the value of its marginal product equals its marginal cost, L_0. Alternatively, equation (4) expresses the employment condition in real terms. The left hand side of this equation is the marginal physical product of labour, whereas the right hand side is the real wage:

$MP_L = W/P$ (4)

Demonstration problem

Suppose the production function is given by the following equation

$Q = 20L - 5L^2$

where L is defined in thousands of units of labour. What is the expression for the marginal product and average product of labour function? At what level of labour does TP reach its maximum? If P = $5 per unit and W = $2 per hour, what is the optimal level of employment of labour?

Answer:

$$MP_L = \frac{dQ}{dL} = 20 - 10L.$$

$$AP_L = \frac{Q}{L} = 20 - 5L$$

Note 'd' in the MP expression is the mathematical notation of derivative. The marginal product is the derivative of the total product with respect to quantity. That is, it shows the amount of change in output when the input of labour changes infinitesimally. TP is maximised when MP_L is set equal to zero: 20 – 10L = 0, hence L = 2 (000).

$VMP_L = W$, \$5(20 –10L) = \$2, and L = 1.96 (000) = 1960

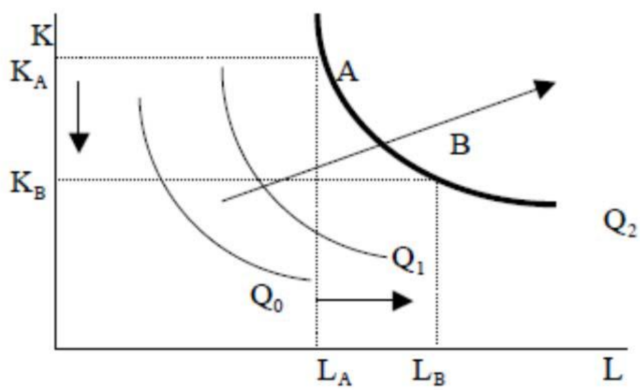
The production function of multiple variable inputs (Isoquants and Isocosts)

The firm's production problem is to choose the optimal levels of all inputs, not just a single input, that enter the production process such that output is maximised for a given budget constraint. Therefore, our next task is to examine the optimal choice of capital and labour in the long run, when both inputs are free to vary. In the presence of multiple variables of production, there exist various combinations of inputs that enable the manager to produce the same level of output.

The basic tools for understanding how alternative inputs can be used to produce output are *isoquants* and *isocosts*. An *isoquant* is a line joining combinations of inputs (K, L) that generate the same level of output. The word *isoquant* comes from the Greek word *iso* meaning equal and the Latin word *quantus* meaning quantity. As with indifference curves, there will be a family of isoquant curves, with higher curves preferred to lower curves. Similarly, the curves do not cross, and they are convex to the origin. Figure 5-3 depicts a typical set of isoquants. Because input bundles A and B both lie on the same isoquant, each will produce the same level of output, namely Q_2 units. Input mix A implies a more capital-intensive plant than the input mix B. As more of both inputs are used, a higher isoquant is obtained. Thus as we move in the northeast direction in the figure, each new isoquant is associated with higher and higher levels of output,

$$Q_2 > Q_1 > Q_0.$$

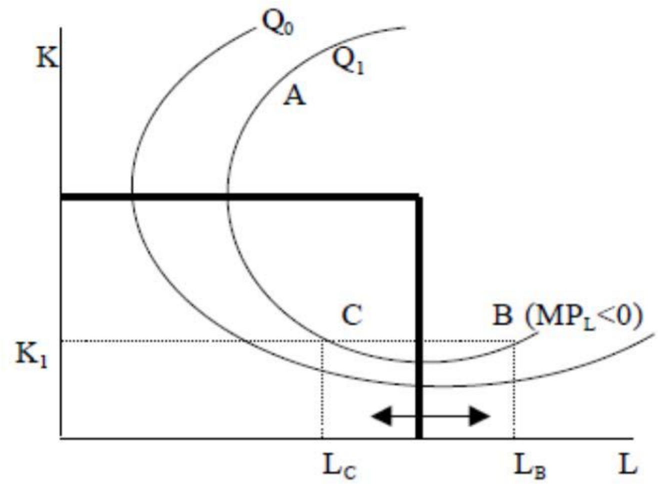
Figure 5-3



Two points need to be appreciated.

1. Unlike indifference curves, isoquant curves are not negatively sloped throughout their length. Isoquants may bend back, as shown in Figure 5-4, at both relatively high and relatively low ratios of capital to labour. The reason for this difference is that in consumer theory we assume that marginal utility would never be negative, whereas in producer theory we make no such restrictive assumption concerning marginal products.

Figure 5-4



To understand this better, we should distinguish between technical efficiency, whereby there is no wasted resources, and economic efficiency, whereby there is no waste of expenditure. Points A and B on the isoquant $Q_{1,}$ in Figure 5-4, are technically inefficient points. For example, the firm at point B is using too much labour relative to point C. By moving from B to C, the firm, while maintaining its output level (Q_1), can save ($L_B - L_C$) in labour. Note that in reducing the input of labour, employment of K remains unchanged. Similarly, point A is also technically inefficient. The two perpendicular (bold) lines drawn against the vertical and horizontal axis, separate the space into the economic and uneconomic region. Points above and the to the right of these lines (outside the box) such as B and A, are in the uneconomic region, while points below and to the left of these two lines (inside the box), such as C, are in economic region.

2. Furthermore, isoquants are convex. The reason isoquants are typically drawn with a convex shape is that inputs such as capital and labour are not perfectly substitutable. In Figure 5-3, for example, if we start at point A and begin substituting labour for capital, it takes increasing amounts of labour to replace each unit of capital that is taken away. The slope of an isoquant at any point can, therefore, be expressed as the ratio of the amount of capital that can be subtracted from the production process to the amount of labour that is added to the production process, such that output level remains constant. This ratio, known as the *marginal rate of substitution (MRTS)* is the rate at which labour and capital can substitute for each other:

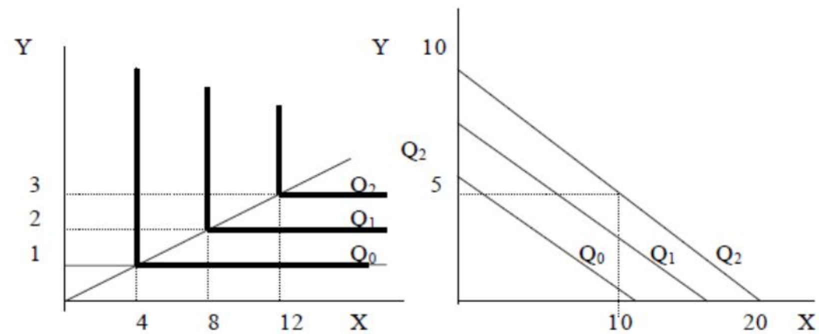
$$MRTS = -\frac{\Delta K}{\Delta L} \qquad (5)$$

where ΔK is the decrement to capital that will just allow output to remain unchanged given a one-unit increment, ΔL, to the labour input.

Shapes of Isoquants

The possibility of substitution of input L for input K and the degree of substitution between them determine the shape and slope of the isoquant. In addition to the normal shape of isoquants illustrated above, there are peculiar shapes associated with perfect substitutes and perfect complements, as shown in Figure 5-5. Panel (a) shows the right-angled isoquants resulting from two inputs that are perfectly complementary; that is, input of a single variable by itself will not produce any output. To obtain output, both Y and X must be input in a fixed ratio. For example, the manufacturing of a car requires an input of one body and four wheels, a ratio that never changes. Therefore, if two bodies are input, eight wheels must also be input to obtain two units of output, because the two inputs are complementary. Note that in panel (a), Y = body casting, and X = wheels.

Figure 5-5



(a) (b)

Panel (b) shows the isoquants resulting from two inputs that are perfect substitutes for one another. Suppose that a firm is choosing between two types of computers to store company data. One has a high-capacity hard drive that can store 10 gigabytes of data, while the other has a low-capacity hard drive that can store five gigabytes of data. If it needs to store 100 gigabytes of data, it could either purchase 10 high-capacity computers and no low-capacity computers (the vertical intercept), or it could purchase no high-capacity computers and 20 low-capacity computers (the horizontal intercept). Or it could purchase five high-capacity computers and 10 low-capacity computers. Note that in panel (b), Y = high-performance computers, while X = low-performance computers.

For the fixed-proportions production function (perfect complements), the expression for the production function is:

$$Q = \min(4W, \text{casting}) \qquad (7)$$

On the other hand, the production function for the linear (perfect substitutes) isoquants is represented by

$$Q = 10H + 5L \tag{8}$$

For this production function, the slope is constant and the MRST does not change as we move along the curve.

Cobb-Douglas production function

This is an intermediate between a linear production function and a fixed proportions production function. This production function is known as the *Cobb-Douglas* production function, and it is given by the formula

$$Q = AK^{a}L^{b} \tag{9}$$

where A, a, and b are positive constants. A is the scale variable, where a and b (the exponents) are the elasticity of output with respect to capital and labour respectively.

In the Cobb-Douglas production function, capital and labour can be substituted for each other. Unlike a fixed proportions production, capital and labour can be used in variable proportions. Unlike a linear production function, though, the rate at which labour can be substituted for capital is not constant as you move along an isoquant. Note, however, that this type of production function is not the only one that allows substitutability between L and K. It is, however, one that has several friendly mathematical properties.

Isocosts

An isocost line is a locus of combinations of inputs that require the same total expenditure. Let us express the firm's expenditure on inputs as

$$TC = K \times P_K + L \times P_L \tag{10}$$

where TC is the total dollar expenditure; P_K and P_L are the unit prices of capital and labour, respectively; and *K* and L are the number of physical units of capital and labour that are to be employed in the production process. This can be rearranged to appear as

$$K = \frac{TC}{P_K} - \frac{P_L}{P_K}L \tag{11}$$

in which form it is perhaps more recognisable as a linear equation explaining K in terms of L and three known values (TC, P_L, and P_K). It can therefore be plotted in the same space as the isoquant curves. The intercept of the isocost line on the capital axis occurs when $L = 0$, and this intercept value of *K* is simply the total expenditure divided by the price of the capital units, resulting in a particular physical quantity of capital units, TC/P_K. As we purchase units of labour, it is evident that our purchase of capital units, for the same expenditure (isocost) level, is drawn down in the ratio of the price of labour to the price of capital.

Figure 5-6

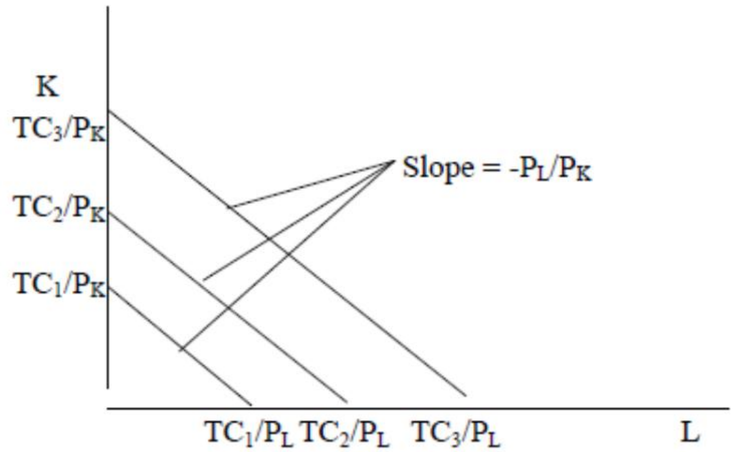


Figure 5-6 shows the graph of isocost lines for a variety of different total cost levels, TC_1, TC_2, and TC_3, where $TC_3 > TC_2 > TC_1$. In general, there are an infinite number of isocost lines, one corresponding to every possible level of total cost.

Demonstration problem

Suppose due to a general slowdown in the economy, the price of all factors of production, labour and capital, decline by the same percentage. How would this change impact the isocost line?

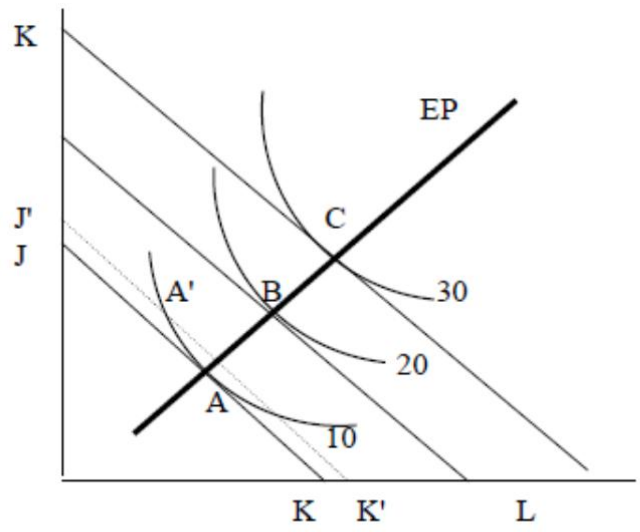
Answer:

A proportionate decrease in the price of both factors leaves the slope of the isocost line, PL/PK, intact while shifting it to the right in a parallel fashion

Cost minimisation (economic efficiency)

For each output level, there will be a minimum-cost combination of the factors necessary to produce that output level. In Figure 5-7, we show three isoquant curves representing 10, 20, and 30 units of output. The cost of producing 10 units is minimised at point A, since any other capital/labour combination producing 10 units, such as at A', will lie to the right of the isocost line *JK* and would thus require a larger total expenditure to purchase, J'K'. Recall that the intercepts on the capital and labour axes are equal to TC/P_K, where P_K and P_L are presumed to remain unchanged. Hence, larger total expenditures are represented by higher intercept points and higher isocost lines. Similarly, the output level 20 is produced at the least-cost by the input combination represented by point B, as 30 units of output is produced at the least-cost by the input combination at point C.

Figure 5-7



Input cost is minimised for a given output level, where the isoquant representing that output level is just tangent to the lowest attainable isocost line. Equivalently, output is maximised for a given input expenditure level where the isocost line representing that expenditure level is just tangent to the highest attainable isoquant. At the point of tangency between the isoquant and the isocost curve, the slopes of these curves are the same. That is, optimality requires that the rate at which the firm can technically substitute labour for capital equals the rate that the market allows it to.

$$MRTS = \frac{MP_L}{MP_K} = \frac{P_L}{P_K} \qquad (11)$$

Rearranging terms, we have

$$\frac{MP_L}{P_L} = \frac{MP_K}{P_K} \qquad (12)$$

This alternative arrangement holds that cost minimization or output maximisation requires that the ratios of the marginal products to prices of all inputs be equal. This is useful in that this way the rule of optimisation can be generalised to a larger number of inputs, 1, 2, 3,…..

$$\frac{MP_1}{P_1} = \frac{MP_2}{P_2} = \frac{MP_3}{P_3} = ... \qquad (13)$$

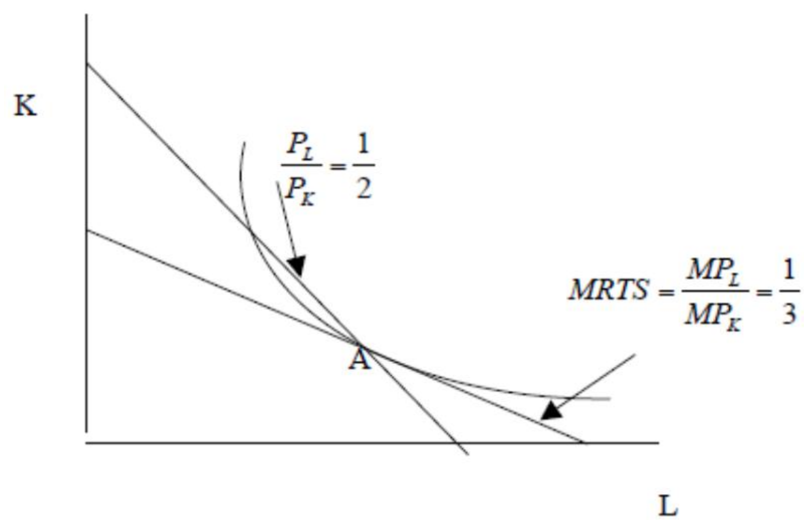
Demonstration problem

Suppose $MP_L = 1$, $MP_K = 3$, whereas P_L and P_K are respectively \$1 and \$2. Is the firm that faces these factors optimising information?

Answer:

Optimality requires $\frac{MP_L}{MP_K} = \frac{P_L}{P_K}$. Plugging in: $\frac{MP_L}{MP_K} = \frac{1}{3}$, whereas $\frac{P_L}{P_K} = \frac{1}{2}$. Therefore, the rate by which the firm is able to substitute labour for capital, 1/3, is lower than the rate the market allows it to, 1/2. As such, the firm is not minimising its costs. Specifically, the firm is using too much labour and too little capital. In terms of a diagram, Figure 5-8, below, portrays this situation. Note that, at point A, $MRTS = \frac{MP_L}{MP_K} = 1/3$ is the slope of the isoquant, and $\frac{P_L}{P_K} = 1/2$ is the slope of the isocost line. The firm needs to substitute K for labour. Doing so will increase MP_L and decreases MP_K, and hence will increase MRTS toward 1/2. The process of substitution stops when the MRTS equals the price ratio (1/2).

Figure 5-8



Demonstration problem (mathematically challenging)

Suppose the production is represented by a Cobb-Douglas production function, $Q = K^{0.5}L^{0.5}$. Also, suppose that P_K and P_L are \$10 and \$5, respectively.

a. Determine the optimal combination of inputs.
b. If the manager has \$500 to spend on labour and capital, how many units of K and L should the firm acquire?

Answer:

a. First we need to take the (partial) derivative of Q to find the marginal products of labour and capital: $MP_K = (0.5)K^{-0.5}L^{0.5}$ and $MP_L = (0.5)K^{0.5}L^{-0.5}$. Hence, $\frac{MP_L}{MP_K} = \frac{0.5K^{0.5}L^{-0.5}}{0.5K^{-0.5}L^{0.5}} = \frac{K}{L}$. Setting this equal to $\frac{P_L}{P_K} = \frac{1}{2}$, we have: $\frac{K}{L} = \frac{1}{2}$. Meaning that for every one unit of K, the firm needs to hire two units of L in order to minimise its costs, L = 2K.

b. The isocost line is represented by $TC = P_K xK + P_L xL = 10.K + 5.L = \$500$. Substituting in for either L or K: \$500 = 10(K) + 5(2K) = 20K, hence K = 25, L = 50.

In Figure 5-7, the line that connects A, B, and C is referred to as the *expansion path (EP)*. The expansion path, or more precisely, the *long run expansion path*, is a locus of the tangency points between various isoquant and isocosts that shows the least-cost combinations of labour and capital a firm would choose as it expanded its output level. This assumes constant price of labour and capital and constant state of technology.

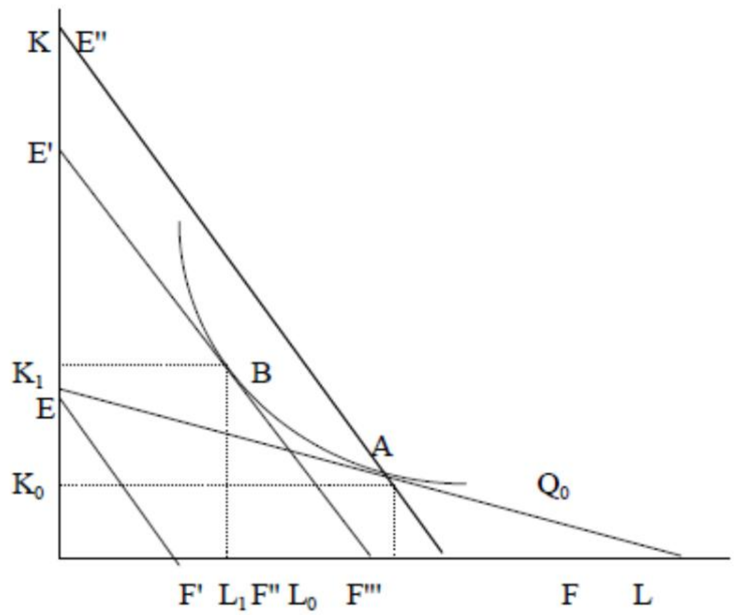
Input substitution

Economic efficiency depends on the relative factor prices. If the price of one factor changes, all else being kept constant, the profit-maximising firm will attempt to substitute away from the factor that has become relatively more expensive and in favour of the factor that has become relatively less expensive. Suppose the initial situation in Figure 5-9, is point A, where output level Q_0 is being produced economically efficiently by a combination of K_0 and L_0.

Now suppose that labour prices rise, for example, because of a new agreement with the labour union. This causes an increase in the cost of labour and the isocost line rotates down (clockwise) from EF to EF′. If the firm wishes to maintain its current expenditure, it cannot produce Q_0. Alternatively, if the firm wishes to maintain its output level at Q_0 to hold its market share, it will need to spend more money on the inputs.

In the long run, the firm will substitute capital for labour, that is, the firm will increase its plant size to K_1 and decrease its labour input to L_1, point B. The increased input price ratio, P_L/P_K has caused the production process to become relatively more capital intensive. Note that E'F", the new isocost line that is tangent to the isoquant Q_0 at point B, is parallel to EF'.

Figure 5-9



Returns to scale

Figure 5-10 illustrates three sets of isoquants for different production processes. In each panel, the points on the ray from the origin show the proportion by which output increases when both inputs are increased proportionately. The distance between the successive isoquants brought about by a proportionate increase in inputs is a reflection of how output responds to changes in inputs. For example, doubling all inputs means doubling the distance from the origin along the ray. We refer to changing all inputs by the same proportion as a change in *scale*. What is the effect on output of a change in scale?

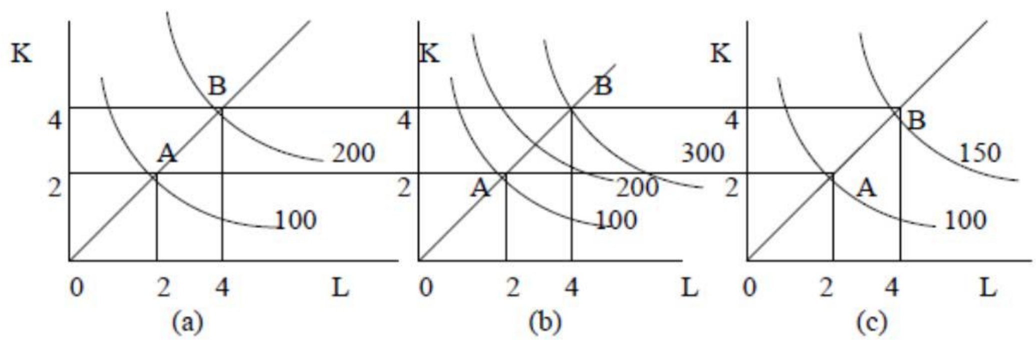
In panel (a), we find that doubling inputs leads to a doubling in output. That is, the move from point A to point B doubles the distance from the origin and also doubles output from 100 to 200. The property of output increasing proportionately to the scale of inputs is called *constant return to scale*, because the amount of output per bundle of inputs is always constant.

A production process exhibits constant returns to scale if increasing all inputs proportionately increases output in equal proportion.

Unlike the proportionately spaced isoquants in panel (a) of Figure 5-10, panel (b) has isoquants that bunch closer together and panel (c) has Isoquants that spread out as the distance from the origin along a ray increases. In panel (b), we find that doubling inputs leads to a more than doubling in output (tripling). That is, the move from point A to point B triples output and hence less than doubles the distance between 100 and 200 from the origin. The property of output increasing more than proportionately to the scale of inputs is called *increasing return to scale*, because the amount of output per bundle of inputs increases with the scale.

Panel (c) shows the case of a production function in which output increases less than proportionately to scale. That is, the move from point A to point B *more than* doubles the distance between 100 and 200 from the origin and also less than doubles output (50 per cent increase). The property of output increasing less proportionately to the scale of inputs is called *decreasing returns to scale,* because the amount of output per bundle of inputs falls as scale increases.

Figure 5-10

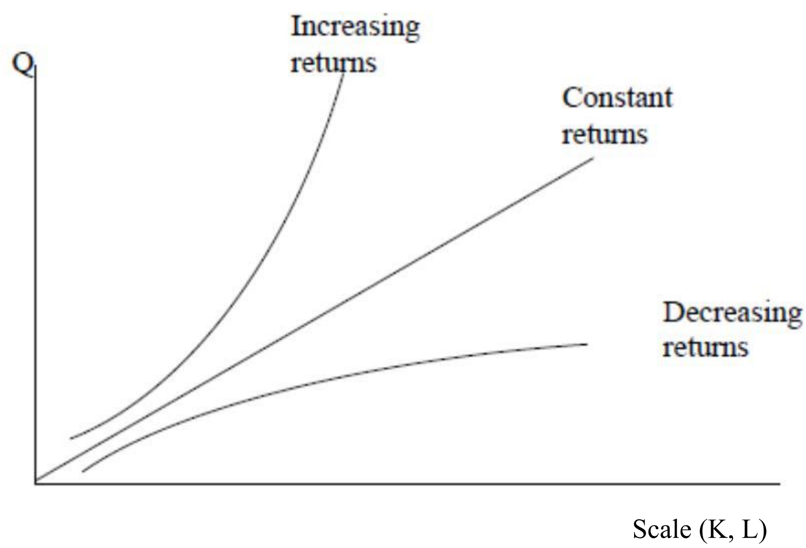


Why might some production processes exhibit constant returns to scale and others exhibit increasing or decreasing returns? One answer is that a larger scale of operation often allows workers and managers to specialise in different tasks. Henry Ford's assembly line, for example, could produce more cars than the same workers and equipment could do if production had to be done one car at a time in separate garages, with each worker designing, machining, assembling and finishing the cars. On the other hand, large scale can also introduce inefficiencies, often associated with breakdown of coordination and critical information flows.

From a policy perspective, increasing returns to scale seem to imply that large firms are desirable and should be encouraged, because they are able to produce more efficiently than small ones and thus should be able to sell their product at a lower price. Yet, as we will see later, large firms often have greater market power that would allow them to charge high prices, and so large firms are sometimes broken into smaller, competing firms to encourage lower prices. Alternatively, as in the case of regional electric utilities, large firms may be regulated by government bodies.

Another way to describe the effects of different returns to scale is to graph output against scale. Assuming that we have fixed a particular ratio of inputs (that is, we have limited ourselves to a particular ray from the origin), scale is measured on a graph like Figure 5-10 as the distance from the origin. Plotting scale versus output leads to Figure 5-11. The product-to-scale curve for decreasing returns to scale is concave with respect to the bottom of the graph. For increasing returns to scale, the curve is convex; and for constant returns to scale, it is a ray from the origin. The positioning of the curves in this Figure has no significance, as each type of curve will be positioned according to the equation that represents the expansion path.

Figure 5-11



Testing production function for returns to scale

If the production function is known, it can be analysed algebraically for returns to scale. Suppose we have a Cobb-Douglas production function $Q = AK^a L^b$. This is a general expression assuming no prior knowledge of returns to scale. Let us double the inputs of K and L

$$Q' = A(2K)^a (2L)^b$$

or

$$Q' = (2)^{a+b}(AK^a L^b) = (2)^{a+b}.Q$$

Hence, if (a + b) = 1, Q' = 2Q, which is the case of constant returns to scale. If (a + b) >1, Q' > 2Q, which is the case of increasing returns to scale, and if (a + b) <1, Q' < 2Q, which is the case of decreasing returns to scale.

Technological changes

So far, we have treated the firm's production function as fixed; that is, it remains stationary over time. But as knowledge in the economy evolves and as firms acquire know-how through experience and investment in research and development, a firm's production function will change. The notion of *technological progress* captures the idea that production functions can shift over time. In particular, technological progress refers to a situation in which a firm can achieve more output from a given combination of inputs, or equivalently, the same amount of output from fewer inputs.

We can classify technological progress into three categories: neutral technological progress, labour-saving technological progress, and capital-saving technological progress. Figure 5-12 illustrates *neutral technological progress*. In this case, an isoquant corresponding to a given level of output (10 units) shifts inward (indicating that lesser amounts of labour and capital are needed to produce a given output), but the isoquants shift so as to leave MRTS unchanged along any ray (e.g., OA) from the origin. Under neutral technological progress, in effect, the firm's entire map of isoquants is simply relabelled, each one now corresponding to a higher level of output, but the isoquants themselves retain the same shape.

Figure 5-12

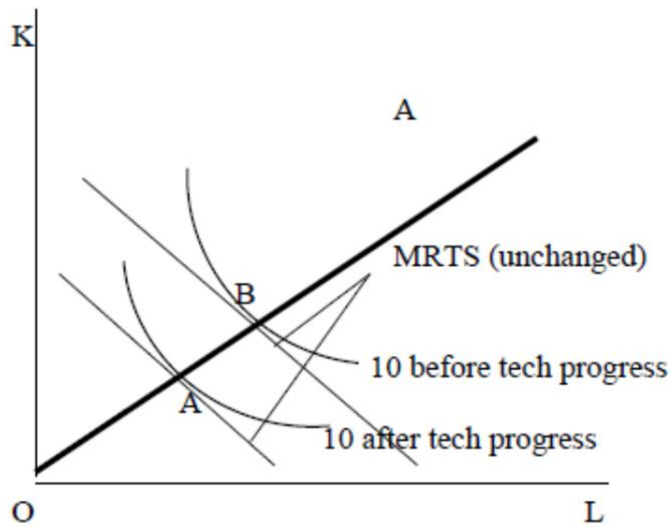


Figure 5-13 illustrates *labour-saving technological progress.* As before, the isoquant corresponding to a given level of output shifts inward, but now along any ray from the origin, the isoquant becomes flatter, indicating that the MRTS is now less than it was before. You should recall that MRTS = MP_L/MP_K, so the fact that the MRTS decreases implies that under this form of technological progress the marginal product of capital increases more rapidly than the marginal product of labour. This form of technological progress would arise when technical advances in capital equipment, robotics, or computers increase the marginal productivity of capital relative to the marginal productivity of labour.

Figure 5-13

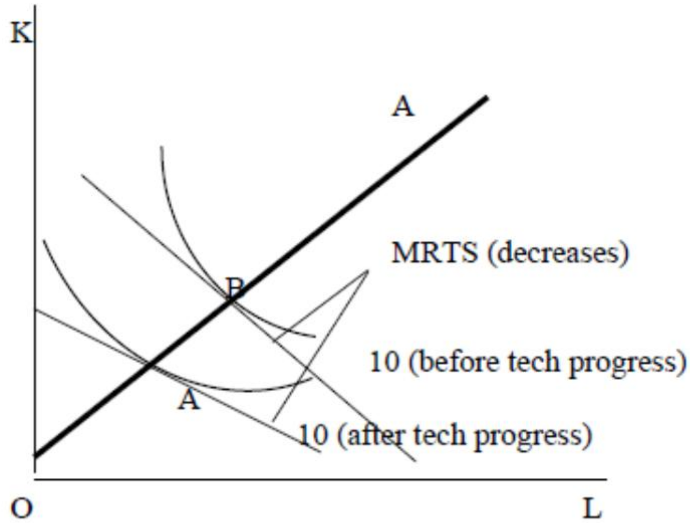
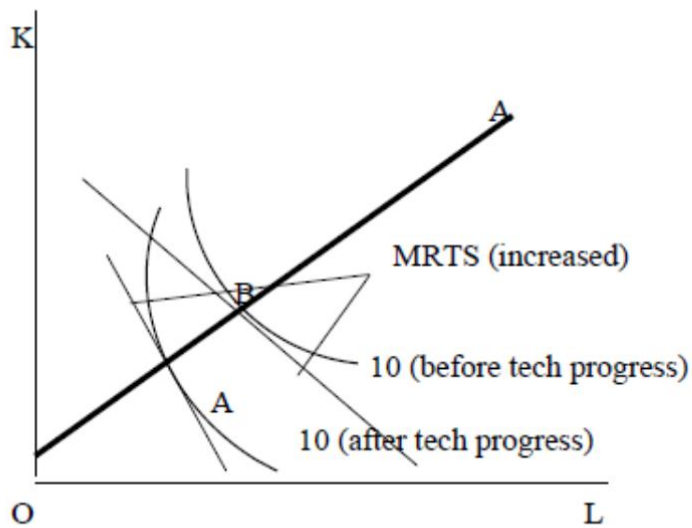


Figure 5-14 depicts *capital-saving technological progress*. Here, as an isoquant shifts inward, MRTS increases, indicating that the marginal product of labour is increasing more rapidly than the marginal product of capital. This form of technological progress would arise if, for example, the education or skill level of the firm's actual (and potential) work force rose, increasing the marginal productivity of labour relative to the marginal product of capital.

Figure 5-14



Module summary

Summary

Production of any goods requires input of at least two basic factors. In the short run, there has to be at least one fixed factor of production and one variable. This results in the law of variable proportions: diminishing and increasing. In the process of production, as the input of the variable factor (labour) increases, the marginal product of labour first increases (increasing returns), and then decreases (diminishing returns).

An isoquant shows the various combinations of two inputs that can be used to produce a specific level of output. Its slope, the 'marginal rate of technical substitution ($MRTS$),' is equal to the ratio of the marginal products of the two inputs. In order to minimise production costs or maximise profit, the firm must produce where an isoquant is tangent to an isocost.

Returns to scale is a long-range planning concept that enables management to determine the effect of increasing all inputs by the same proportion. 'Constant,' 'increasing,' and 'decreasing' returns to scale refer to the situation where output changes, respectively, by the same, by a larger, and by a smaller proportion than inputs. Increasing returns to scale arise because of specialisation and division of labour and from using specialised machinery. Decreasing returns to scale arise primarily because as the scale of operation increases, it becomes more and more difficult to manage the firm and coordinate its operations and divisions effectively. In the real world, most industries seem to exhibit near-constant returns to scale.

Assignment

Assignment

1. The manager of a plant calculated the cost at different output levels. The result is in the table below:

Units of Labour	Total Product (Units)
0	850
10	1,700
20	3,500
30	6,900
40	10,000
50	11,500
60	12,600
70	11,550
80	10,400

Find the average product and marginal product of labour for different levels of labour input.

2. A firm uses only labour to produce to output Q. The production function is:

$Q = 10 \times L^{1/2}$

If labour costs $ 25.00 per hour, then find the firm's total cost, average cost, and marginal cost functions.

Assessment

Assessment

1. Cost of labour is $100 per unit and capital is $200 per unit. If the marginal product of labour is 1,000 and marginal product of capital of is 2,500, then is it optimal to use the combination of 10 labour and 30 capital? If not, then how should the owner revise the combination?

2. A firm makes an optimal production decision based on its technology and prices of inputs. The inputs that are used are labour (L) and capital (K). Suppose labour cost is $20 per unit and capital cost is $10 per unit. What is the MRTS for production of 500 units? If prices of inputs remain the same, will the MRTS change for production of 1,000 units? If the firm employs 100 units of labour and 100 units of capital with the technology $K^{0.5} L^{0.5}$, then is the firm optimising its production decision?

Assessment answers

1. Wage (w) = \$100, Rent (r) = \$200

   MPL = 1,000, MPK = 2,500

   MPK/r = 2,500/200 = 12.5 > MPL/w = 1,000/100 = 10

   Therefore more capital (or less labour) needs to be employed until MPK/r = MPL/w.

2. In equilibrium (when optimal decisions are made) MRTS = Input Cost Ratio = 10/20 = 0.5. MRTS at an output level of 500 should not differ from MRTS at an output level of 1,000 if the input costs do not change.

   Technology = $K^{0.5}L^{0.5,}$ MPK = $0.5K^{-0.5}L^{0.5}$ and MPL = $0.5K^{0.5}L^{-0.5}$

   Therefore, MPK/MPL = L/K. When optimal decisions are made,
   L/K = 0.5, and
   L = 0.5K

   If 100 L and 100 K are used to produce any level of output, since L/K = 1 > 0.5.

References

References

Baye, M. (2002). *Managerial Economics and Business Strategy*. Irwin: McGraw Hill.

Besanko, D., & Braeutigam, R. (2002). *Microeconomics: An Integrated Approach*. New York: Wiley & Sons.

Douglas, J. E. (1992). *Managerial Economics: Analysis and Strategy*. Upper Saddle River, NJ: Prentice Hall.

Pindyck, R. S., & Rubinfeld, D. (2001). *Microeconomics, 5th Edition*. Upper Saddle River, NJ: Prentice Hall.

Salvatore, D. (1993). *Managerial Economics in a Global Economy*. Irwin: McGraw Hill.