(e) In view of these facts, is it desirable or undesirable, from a social point of view, for a producer to operate with average product below its maximum (to the left)? With marginal product at its maximum? Why is your answer true in each case?

SUGGESTED READINGS

1. CLARK, J. M. "Diminishing Returns," *Encyclopaedia of the Social Sciences,* Vol. V, pp. 144–46. New York: Macmillan Co., 1931.
2. KNIGHT, FRANK H. *Risk, Uncertainty, and Profit,* pp. 94–104. Boston: Houghton Mifflin Co., 1921.
3. MACHLUP, FRITZ. "On the Meaning of the Marginal Product," *Explorations in Economics,* pp. 250–63. New York: McGraw-Hill Book Co., 1936. Reprinted in AEA, *Readings in the Theory of Income Distribution,* pp. 158–74. Philadelphia: Blakiston Co., 1951.
4. PRESCOTT, J. A. "The Law of Diminishing Returns in Agricultural Experiment," *Economic Record,* IV (1928), pp. 85–89.
5. ROBBINS, LIONEL. "Production: Theory," *Encyclopaedia of the Social Sciences,* Vol. XII, pp. 462–67. New York: Macmillan Co., 1934.
6. HENDERSON, JAMES M. AND RICHARD E. QUANDT. *Microeconomic Theory: A Mathematical Approach,* pp. 43–47. New York: McGraw-Hill Book Co., Inc., 1958. [Elementary math necessary.]

Chapter 7

PRODUCTION AND OPTIMAL INPUT PROPORTIONS: TWO VARIABLE INPUTS

7.1 INTRODUCTION

The fundamental physical relationships of production were discussed in Chapter 6 under the assumption that there is only one variable input. The analysis is continued in this chapter for a more general case. Graphically, production is studied under the assumption that there are two variable inputs. One may regard these inputs either as cooperating with one or more fixed inputs or as the only two inputs. The latter situation, of course, is relevant only for the long run. In either case, however, the results of the two-input model are easily extended to cover multiple inputs.

7.1.a—Production Table

The land-labor example used in Chapter 6 may be expanded to introduce the theory of production with two variable inputs. In the illustration we considered an agricultural experiment in which three-acre tracts of land comprised the fixed input. Labor was the variable input, and we obtained eight sample observations corresponding to the cultivation of the three-acre tracts by one worker, two workers, and so on. In the present example the agricultural experiment is pushed further so as to obtain sixty-four sample observations. Land is, in a sense, still the fixed input; but now we suppose there are eight one-acre tracts, eight two-acre tracts, and so on up to eight eight-acre tracts. Each of the sets of eight constant-acre tracts is cultivated by one worker, two workers, etc., up to eight workers. Thus we have samples ranging from one worker on one acre to eight workers on eight acres. The hypothetical data are listed in Table 7.1.1.

The entries in the row corresponding to three-acre tracts of land are exactly the same as the entries in Table 6.2.2. Indeed, in every respect this table is just a "larger" example of the hypothetical experiment in Chapter 6.

In the spirit of Chapter 6, consider land as the fixed input. The entries in each row show the total outputs produced on the stipulated acreage when different numbers of workers cultivate the land. By successive subtractions along each row, the marginal product of labor is obtained. Next, by going to successively higher rows one sees that the total, average, and marginal products of labor increase as larger and larger tracts of land are used—that is, as the fixed input is expanded relative to the variable input.

TABLE 7.1.1

DATA FROM HYPOTHETICAL AGRICULTURAL EXPERIMENT*

Acres of Land per Tract	*Output in Bushels*							
8	9	46	69	92	109	124	136	144
7	13	46	69	91	108	123	134	140
6	16	42	66	88	106	120	128	132
5	15	37	60	80	100	113	120	121
4	13	30	54	72	85	93	95	95
3	10	24	39	52	61	66	66	64
2	6	12	17	21	24	26	25½	24½
1	3	6	8	9	10	10	9	7
	1	2	3	4	5	6	7	8
	Workers per Tract of Land							

* Notice that this production schedule does not represent a production function homogeneous of degree one

Up to a point! But just as too many workers per acre of land make cultivation too intensive, too many acres of land per worker make cultivation too extensive. Instead of viewing acres per tract as the fixed input, we can regard workers per tract as fixed and the number of acres per tract as variable. We then read up the columns rather than across the rows; but the same fundamental physical relationships are exhibited.

With one worker per tract, output increases as the size of the tract increases until six acres per tract is reached. Thereafter total output declines and the marginal product of land is negative. As the number of workers per tract is expanded, thus diminishing the land-labor ratio for each given acreage, total product expands continuously beyond four-acre tracts. Total product in these cases does not reach a maximum in the range shown in this example. But in each case the point of diminishing marginal returns is reached; thereafter output expands at a decreasing rate.

7.1.b—Input Substitution

Table 7.1.1 shows that the basic principles of physical production hold whether workers per tract are varied with acres per tract constant or whether acres per tract are varied with workers per tract constant. It also illustrates another very important physical relationship between inputs: the same amount of total output may be produced by different input combinations. For example, an output of sixty-six bushels can be produced by using six workers on three acres of land or by using three workers on six acres. Similarly, one hundred and twenty bushels can be produced either by seven workers on five acres or by six workers on six acres.

In this example no more than two different input combinations can be used to produce the same output. In a more general, continuous case, however, a given level of output can be produced by a wide variety of different input combinations. In other words, one input may be *substituted* for another in producing a specified volume of output. One of the important tasks of a businessman is to select the particular input combination that minimizes the cost of producing any given level of output. The chief purpose of this chapter is to show how this is done.

7.2 PRODUCTION SURFACE

Selection of the least-cost input combination requires knowledge of substitution possibilities and of relative input cost. For an individual producer, we assume the input prices are given by market forces of supply and demand. The input substitution is the center of our interest. To get at an explanation requires the use of a device much like that used in Part I to describe a consumer's preference surface. In the theory of consumer behavior we used equal-satisfaction contour lines, or indifference curves. Here we use equal-output contours, or *isoquants.*

7.2.a—Production Surface for Discrete Case

As an introduction, first look at the total production surface. Figure 7.2.1 is a graph of the discrete production function given in Table 7.1.1. The height of the rectangular blocks indicates the volume of output. By following the heights visually in either "horizontal" direction, one may see how the total product curve is shaped for a fixed amount of one input and variable amounts of the other. But as we have already observed in this example, substitution possibilities are very limited. In certain cases two different input combinations yield the same

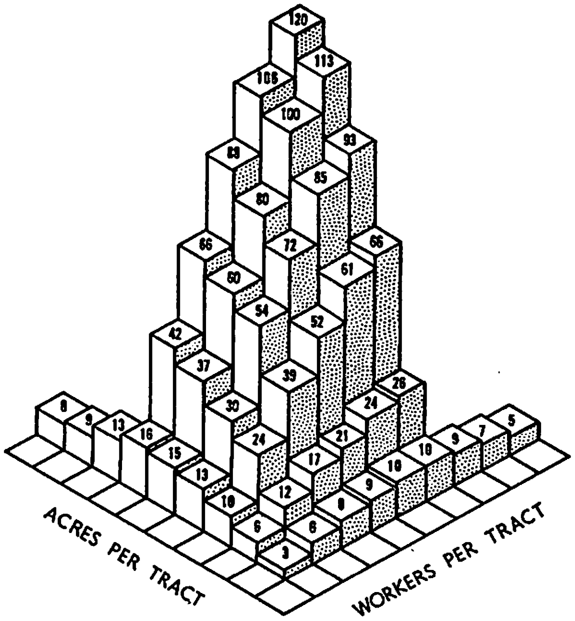


FIGURE 7.2.1

Physical Production Surface for Example in Table 7.1.1

output. However this example is *too* discrete to illustrate a wide range of production possibilities.

7.2.b—Production Surface for Continuous Case

For this purpose a *continuous* production function is required. Let us imagine a manufacturing process that requires two inputs—labor and capital—to produce a specific commodity. The production function for this good is continuous; it cannot, therefore, be shown conveniently in tabular form. However, either a mathematical or a graphical representation is suitable.[1] The production function for this particular example is shown in Figure 7.2.2, a three-dimensional diagram in which height measures quantity of output and the two "flat" or "horizontal" dimensions measure quantities of the two inputs.

---

[1] Let Q, K, and L represent the quantities of output, capital, and labor, respectively. The production function may be written $Q = f(K, L)$, where $\partial Q/\partial K$ and $\partial Q/\partial L$ are the marginal products of capital and labor, respectively.

The production surface is $OCQL$. Any point on this surface represents a particular quantity of output. Dropping perpendiculars from the point to the axes shows the quantities of inputs required. For example, P is a point on the surface, and PP' is the associated volume of output. Drawing perpendiculars to the axes, $OL_1 (=C_1P')$ units of labor and $OC_1 (=L_1P')$ units of capital are required to produce the amount PP' at this particular point.

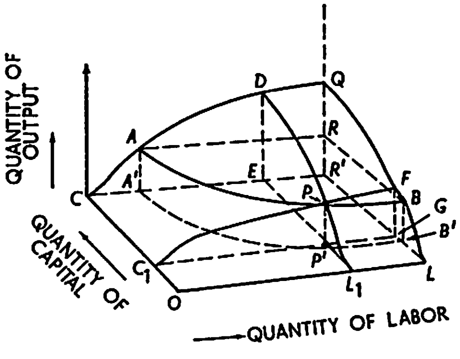


FIGURE 7.2.2

Physical Production Surface for a Continuous Production Function

The production surface may be viewed in a different manner. Hold the capital input constant at the amount OC_1. The total product curve for OC_1 units of capital and variable inputs of labor is C_1PF. At labor input OL_1, total output is PP'; and at labor input OL, total output is FG. The total product curve C_1PF rises rapidly for small quantities of labor input, reaches a point of maximum slope (the point of diminishing marginal physical returns to labor for the given capital input OC_1), and thereafter increases at a decreasing rate.

The same statement applies to a typical total product curve for a fixed labor input and variable capital usage. Hold the input of labor constant at OL_1 units. L_1PD is the curve of total output resulting from variable inputs of capital. For example, when OC_1 units of capital are used output is PP'; when OC units are employed output is DE.

7.2.c—Production Isoquants

Still using Figure 7.2.2, let us determine all the different input combinations capable of producing PP' units of output. To do this, we slice (or "intersect") the production surface $OCQL$ at the height

$PP' = AA' = BB'$. This slicing process generates the curve APB, a locus of points equidistant ($AA' = PP' = BB'$) from the C–L plane. By dropping perpendiculars from each point on the APB curve to the C–L plane, one obtains the input combinations associated with each point. In other words, the curve APB is projected onto the C–L plane, generating the curve $A'P'B'$. The latter is a locus of points each of which represents a combination of inputs capable of producing the stipulated quantity of output $PP' = AA' = BB' = RR'$. For examples, the following three combinations of capital and labor are points on the curve $A'P'B'$: OC, CA'; OC_1, OL_1; LB', OL.

The curve $A'P'B'$ is called an *isoquant.*[2]

Definition: an isoquant is a curve in input space showing all possible combinations of inputs physically capable of producing a given level of output. The entire three-dimensional production surface can be exactly depicted by a two-dimensional isoquant map, the quantity of output being represented by the distance of the isoquant from the origin.

A portion of an isoquant map, derived from a production surface such as $OCQL$ in Figure 7.2.2, is shown in Figure 7.2.3.[3] The two axes measure the quantities of inputs and the curves show the different input combinations that can be used to produce 100, 200, 300, and 400 units of output respectively. As is obvious, the further northeast a curve lies the greater is the output associated with it.

Consider first the isoquant for 100 units of output. Each point on this curve shows a capital-labor combination that can produce 100 units of output. For example, OC_1 units of capital and OL_1 units of labor may be used, or OC_3 units of capital and OL_3 units of labor, or any other input combination found by dropping perpendiculars to the axes from a point on the curve.

A ray from the origin, such as OAB or $OA'B'C'$, defines a constant capital-labor input ratio. In particular, the slope of the ray is the input ratio. For example, at points A and B, 100 and 200 units of output, respectively, are produced at the capital-labor ratio $OC_1/OL_1 = OC_2/OL_2$. Similarly, at points A', B', and C', 100, 200, and 300 units of output, respectively, are produced at the capital-labor ratio $OC_3/OL_3 = OC_4/OL_4 = OC_5/OL_5$.

---

[2] Let the production function be $Q = f(K, L)$ as in footnote 1. The different input combinations that can produce $\overline{Q}$ units of output can be found by solving $f(K, L) = \overline{Q}$ for K and L. This expression is the equation for the isoquant associated with $\overline{Q}$ units of output. The isoquant map is generated by allowing $\overline{Q}$ to vary over all possible output values.

[3] The excluded portion of the isoquant map is discussed in subsection 7.3.d.

Along the ray OAB, various levels of output are producible by the same input ratio; the magnitude of the inputs increases as one moves out the ray but the capital-labor ratio remains unchanged. This contrasts clearly with movements along an isoquant. In this case the level of output remains unchanged and the capital-labor ratio changes continuously.

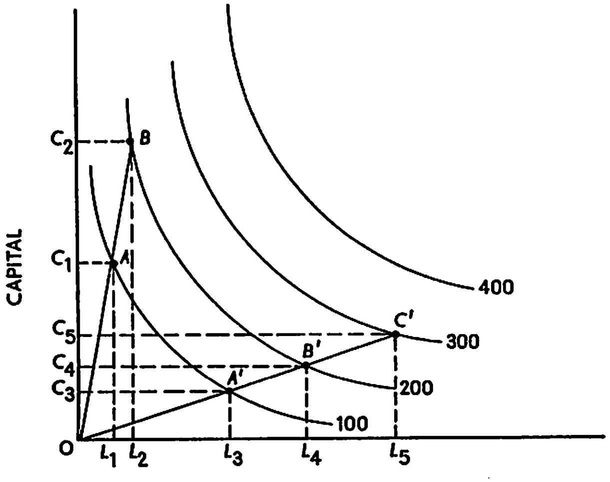


FIGURE 7.2.3

TYPICAL SET OF ISOQUANTS

These points may be summarized as follows.

Relationships: an isoquant represents different input combinations, or input ratios, that may be used to produce a specified level of output. For movements *along an isoquant,* the level of output remains constant and the input ratio changes continuously. A ray from the origin defines a specific, constant input ratio. For movements *along a ray,* the level of output changes continuously and the input ratio remains constant.

7.2.d—Fixed-Proportions Production Functions

Using the isoquant device, it is easy to illustrate the case of fixed-proportions production functions, briefly mentioned in Chapter 6. As you will recall, production is subject to fixed proportions when one, and only one, combination of inputs can produce a specified output. For example, consider the hypothetical production process illustrated in Figure 7.2.4. Two inputs, capital and labor, must be used in the fixed ratio 2:3. That is, two units of capital and three units of labor are required to produce 100 units of output. Thus four units of capital and six units of labor can produce 200 units of output; six units of capital and nine units of labor can produce 300 units, and so on.

The required capital-labor ratio is shown by the slope of the ray *OR* in Figure 7.2.4. Isoquants are constructed for 100, 200, and 300 units of output. Rather than taking the more conventional shape shown in Figure 7.2.3, the isoquants for fixed-proportions processes are right angles. This illustrates, for example, that if three units of labor and two units of capital are employed, 100 units of output are obtainable. However, if the quantity of capital is expanded, labor input held constant, no additional output can be obtained. Similarly, if capital input is held constant and labor expanded, output is unchanged. In other words, the marginal product of either labor or capital is zero if its usage is expanded while the other input is held constant. On the other hand, doubling inputs at the required ratio doubles output; trebling inputs at the required ratio trebles output, etc.

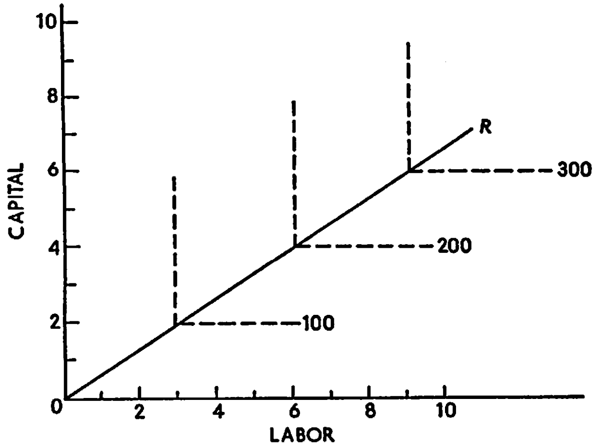


FIGURE 7.2.4

ISOQUANT MAP FOR FIXED-PROPORTIONS PRODUCTION FUNCTION

A rather realistic case is that in which many, but not an infinite number, of different fixed-proportions processes are available. For example, Table 7.2.1 contains hypothetical data regarding the production of a commodity for which five different fixed-proportions processes are available. The 100-output isoquants, together with the capital-labor ratio rays, are plotted in Figure 7.2.5.

Heavily shaded straight lines have been drawn to connect the different possible input combinations. Each of the five points on this kinked line represents an input combination capable of producing 100 units of output. The kinked line *ABCDE* looks very much like the "normal" isoquant shown in Figure 7.2.3. It is different, however, in that no input combination lying on the arc between *A* and *B, B* and *C,* etc., is itself directly a feasible input combination. For example, it is not possible to produce 100 units of output by a process using 7.25 units of capital and 2.5 units of labor.

TABLE 7.2.1

PRODUCTION WHEN SEVERAL FIXED-PROPORTION PROCESSES ARE AVAILABLE

Ray	*Capital-Labor Ratio*	*Capital Input*	*Labor Input*	*Total Output*
OA	11:1	11	1	100
		22	2	200
OB	8:2	8	2	100
		16	4	200
OC	5:4	5	4	100
		10	8	200
OD	3:7	3	7	100
		6	14	200
OE	1:10	1	10	100
		2	20	200

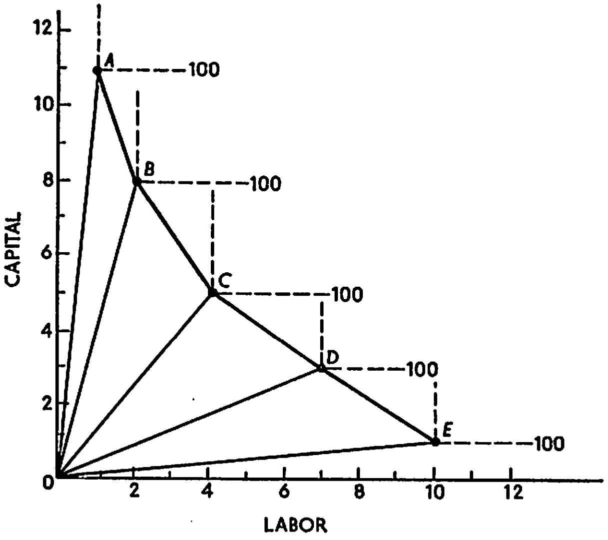


FIGURE 7.2.5

ISOQUANT MAP WHEN FIXED-PROPORTIONS PROCESSES ARE AVAILABLE

On the other hand, if input units are sufficiently divisible, any particular input ratio—represented by a point on the kinked line—can be achieved. All that is required is the proper combination of the two fixed-proportions processes with which it is most closely associated. For example, suppose a producer wished to obtain 100 units of output by using 7.25 units of capital and 2.5 units of labor. He could do so by producing 75 units of output by the process represented by the ray *OB* and 25 units by the process *OC*. To produce 75 units at the 8:2 ratio requires 6 units of capital and 1.5 units of labor. Producing 25 units at the 5:4 ratio requires 1.25 units of capital and 1 unit of labor. Thus 100 units of output can be produced at the desired ratio 7.25:2.5 by combining the two processes represented by the rays *OB* and *OC*.

Finally, suppose there are many fixed-proportions processes by which a given level of output can be produced. Instead of the five points in Figure 7.2.5 there would be many points. Similarly, there would be many straight-line facets of the type *AB*, *BC*, etc. As the number of processes increases, the kinked line looks more and more like a typical isoquant. Indeed, an isoquant depicting a variable-proportions production function is just the limiting case of fixed-proportions processes as the number of processes increases without bound.

This argument, in fact, constitutes one rationale for the use of smooth isoquants and variable-proportions production functions in economic theory. Many manufacturing processes may be characterized by fixed proportions; however, usually many different fixed-proportions processes are available. Constructing smooth isoquants rather than multiple-facet lines simplifies analysis while leading to relatively unimportant departures from real-world conditions. The chief difference is that with smoothly continuous isoquants, any desired capital-labor ratio can be attained (if it is feasible) by using one process, whereas when there are many fixed-proportions processes, a desired combination may require the proper mixture of two processes.

7.3 INPUT SUBSTITUTION

One of the chief features of production under conditions of variable proportions—or a large number of alternative fixed-proportions processes—is that different combinations of inputs can produce a given level of output. In other words, one input can be *substituted* for another in such a way as to maintain a constant level of output. Great theoretical and practical importance attaches to the *rate* at which one input must be substituted for another in order to keep output constant and to the proportionate change in the input ratio induced by a given proportionate change in the rate of substitution.

7.3.a—Marginal Rate of Technical Subsitution

Consider the representative isoquant I_1 in Figure 7.3.1. *P* and *R* are two of the many different input combinations that may be used to produce the I_1-level of output. If production occurs at *P*, OC_1 units of capital and OL_1 units of labor are required. OC_2 units of capital and OL_2 units of labor are required for production at *R*. Thus *P* is associated with the capital-labor ratio given by the slope of $OP = OC_1/OL_1$ and *R* with the capital-labor ratio given by the slope of $OR = OC_2/OL_2$.

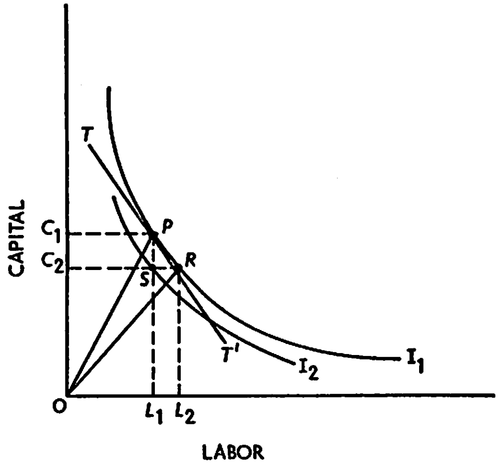


FIGURE 7.3.1

MARGINAL RATE OF TECHNICAL SUBSTITUTION

If there is a change from *P* to *R*, the same level of output is produced by using *more* labor and less capital—labor can be substituted for capital by moving from *P* to *R*, and *vice versa*. The rate at which labor can be substituted for capital over the arc *PR* is given by $-\frac{OC_1 - OC_2}{OL_1 - OL_2} = \frac{PS}{SR}$, where the minus sign is affixed so as to yield a positive number. Stated alternatively, the rate of substitution is the change in capital usage divided by the change in labor usage, or the slope of the curvilinear angle *PRS*.

As the distance from *P* to *R* diminishes the slope of the curvilinear segment *PR* approaches the slope of the tangent *TT'* at point *P*. In the limit, for a very tiny movement in the neighborhood of *P* the slope of the tangent at P measures the rate of substitution. In this case—for small movements along I_1—it is called the *marginal rate of technical substitution,* just as the slope of a consumer's indifference curve is called the marginal rate of substitution in consumption.

Next, suppose labor input is held constant at the OL_1 level while the input of capital is increased from OC_2 to OC_1. Output would increase from the I_2 level (say, Q_2) to the I_1 level (say, Q_1). The marginal product of capital is, of course, the increase in output per unit increase in input, or $\frac{Q_1 - Q_2}{OC_1 - OC_2}$. Since $OC_1 - OC_2 = PS$, the marginal product of capital is $\frac{Q_1 - Q_2}{PS}$.

Now return to the I_2 level and hold capital input constant at OC_2 while increasing labor input from OL_1 to OL_2, or by the amount SR. The marginal product of labor for this change is $\frac{Q_1 - Q_2}{SR}$. The ratio of the marginal product of labor to that of capital is $\frac{Q_1 - Q_2}{SR} \div \frac{Q_1 - Q_2}{PS} = \frac{PS}{SR}$, the rate of substitution of labor for capital. Thus in the limit, as the distance from P to R becomes very small, the marginal rate of technical substitution of labor for capital is equal to the ratio of the marginal product of labor to the marginal product of capital.

These results may be summarized as follows:

Relationships: the marginal rate of technical substitution measures the reduction in one input per unit increase in the other that is just sufficient to maintain a constant level of output. The marginal rate of technical substitution of input x for input y at a point on an isoquant is equal to the slope of the isoquant at that point. It is also equal to the ratio of the marginal product of input x to the marginal product of input y.[4]

7.3.b—Diminishing Marginal Rate of Technical Substitution

The marginal rate of technical substitution of labor for capital diminishes as more and more labor is substituted for capital. This proposition sounds plausible; and it is not difficult to explain.

As additional units of labor are added to a fixed amount of capital the marginal product of labor diminishes. Furthermore, as shown in Figure 6.2.4, if the amount of the "fixed" input is diminished the marginal product of labor diminishes. Thus two forces are working to diminish the marginal product of labor: (a) less of the "fixed" input causes a downward *shift* of the marginal product of labor curve; (b) more units of the "variable" input (labor) causes a downward movement *along* the marginal product of labor curve. Thus as labor is substituted for capital the marginal product of labor must decline.

For analogous reasons, the marginal product of capital rises. As shown in Figure 6.3.6, over stage II—the relevant range of production—the marginal product of capital rises as the marginal product of labor falls. A decrease in the capital-labor ratio is the cause of both. With the quantity of labor fixed, the marginal product of capital rises as fewer units of capital are used. But simultaneously there is an increase in labor input thereby shifting the marginal product of capital curve upward. The same two forces are present in this case: a movement along a marginal product curve and a shift in the location of the curve. In this situation, however, both forces work to increase the marginal product of capital. Thus as labor is substituted for capital the marginal product of capital increases.

As already defined, the marginal rate of technical substitution is the ratio of the marginal product of labor to the marginal product of capital. As labor is substituted for capital, the marginal product of labor declines and the marginal product of capital increases. Hence the marginal rate of technical substitution of labor for capital declines as

---

[4] These relationships can easily be expressed mathematically. The production function is that of footnote 1:

$$Q = f(K, L), \tag{7.3.1.}$$

where

$$\frac{\partial Q}{\partial K}, \frac{\partial Q}{\partial L} \tag{7.3.2}$$

are the marginal products of capital and labor, respectively.

Consider the total differential of the production function in equation (7.3.1):

$$dQ = \frac{\partial f}{\partial K} dK + \frac{\partial f}{\partial L} dL \cdot \tag{7.3.3}$$

For movements along an isoquant, output is constant, so $dQ = 0$. Substituting in (7.3.3):

$$\frac{\partial f}{\partial K} dK + \frac{\partial f}{\partial L} dL = 0 \tag{7.3.4}$$

is the equation for an isoquant. The marginal rate of technical substitution, by definition, is $-dK/dL$. From equation (7.3.4), therefore,

$$MRTS_{L \text{ for } C} = -\frac{dK}{dL} = \frac{\frac{\partial f}{\partial L}}{\frac{\partial f}{\partial K}} = \frac{MP_L}{MP_K}, \tag{7.3.5}$$

by expression (7.3.2).

labor is substituted for capital so as to maintain a constant level of output. This may be summarized as follows:

Relationship: as labor is substituted for capital along an isoquant (so that output is unchanged), the marginal rate of technical substitution declines.

The fact that the marginal rate of technical substitution falls as labor is substituted for capital means that isoquants must be concave from above (that is, in the neighborhood of a point of tangency, the isoquant must lie above the tangent line). This is illustrated in Figure 7.3.2.

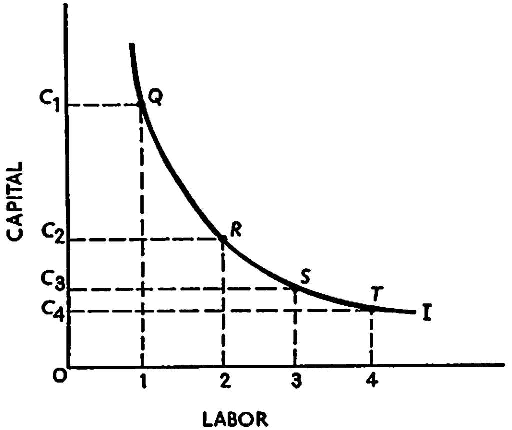


FIGURE 7.3.2

DIMINISHING MARGINAL RATE OF TECHNICAL SUBSTITUTION

Q, R, S, and *T* are four input combinations lying on the isoquant *I*. *Q* has the combination OC_1 of capital and one unit of labor; *R* has OC_2 units of capital and two units of labor; and so on. For the movement from *Q* to *R*, the marginal rate of technical substitution of labor for capital is, by formula,

$$-\frac{OC_1 - OC_2}{1 - 2} = OC_1 - OC_2 \cdot$$

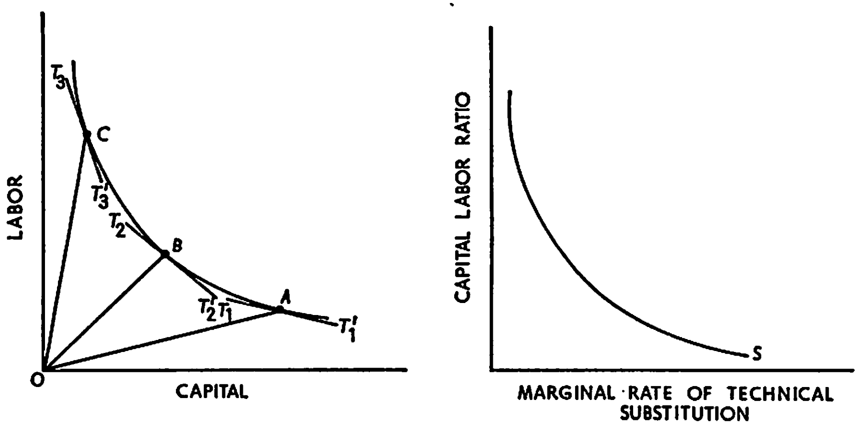
Similarly, for the movements from *R* to *S* and *S* to *T*, the marginal rates of technical substitution are $OC_2 - OC_3$ and $OC_3 - OC_4$, respectively.

Since the marginal rate of technical substitution of labor for capital diminishes as labor is substituted for capital, it is necessary that $OC_1 - OC_2 > OC_2 - OC_3 > OC_3 - OC_4$. Visually, the amount of capital replaced by successive units of labor will decline if, and only if, the isoquant is concave from above. Since the amount *must* decline, the isoquant must be concave from above.[5]

Relationship: isoquants must be concave from above at every point in order to satisfy the principle of diminishing marginal rate of technical substitution.

7.3.c—Elasticity of Substitution

We have now seen that the marginal rate of technical substitution declines as labor is substituted for capital along an isoquant. It is also interesting to know the percentage change in the capital-labor ratio induced by a given percentage change in the marginal rate of technical substitution. In other words, we wish to know the *elasticity of substitution.*



PANEL a

PANEL b

FIGURE 7.3.3

SUBSTITUTION CURVE

This concept is explained by means of Figure 7.3.3. Panel a of that figure shows a typical isoquant with labor input plotted on the vertical axis and capital input on the horizontal axis. For simplicity, only three different input combinations are shown, represented by the slopes of the rays *OA, OB,* and *OC*. Moving from point *A* to point *B*, and from point *B* to point *C*, the slope of the ray increases. This means that the labor-capital ratio increases or that the capital-labor ratio decreases. Furthermore, as we have seen, the marginal rate of technical substitution diminishes for movements to the right along an isoquant. Hence it increases for movements to the left and upward, as indicated by the

---

[5] This proposition is proved in footnote 8.

progressively steeper slopes of the tangents T_1T_1', T_2T_2', and T_3T_3'.

Now transfer to panel b, where the capital-labor ratio is plotted on the vertical axis and the marginal rate of technical substitution is plotted on the horizontal axis. Movements from A to B and from B to C in panel a are associated with a *decline* in the capital-labor ratio. At the same time the marginal rate of technical substitution increases. Hence there is an inverse or negative relationship between the marginal rate of technical substitution and the capital-labor ratio. This is shown in panel b by the negatively sloped curve labeled S—the substitution curve. The elasticity of substitution is the elasticity of this curve.

The elasticity of substitution is greater than, equal to, or less than unity according as a given percentage change in the marginal rate of technical substitution induces a greater, an equal, or a smaller percentage change in the capital-labor ratio in the opposite direction. In other words, the elasticity of substitution is the negative of the slope of the S-curve multiplied by the ratio of the marginal rate of technical substitution to the capital-labor ratio. Using "Δ" to indicate "the change in," the formula is[6]

$$\text{Elasticity of Substitution} = -\frac{\Delta(C/L)}{\Delta(MRTS)} \cdot \frac{(MRTS)}{(C/L)}.$$

7.3.d—Economic Region of Production

Many production functions lead to initial isoquant maps such as shown in Figure 7.2.3. Others, however, generate an isoquant map such as that shown in Figure 7.3.4. It is like the map in Figure 7.2.3 in that the isoquants do not intersect; the higher the isoquants the greater the level of output; and over a range of input values they are negatively sloped. The only difference lies in the fact that the isoquants in Figure 7.3.4 "bend back upon themselves," or have positively sloped segments.

The parallel dashed lines in Figure 7.3.4 indicate the points at which the isoquant bends back upon itself. The lines OC and OL join these points and form, as we will see, the boundaries for the economic region of production (or the stage II region).

Suppose the quantity represented by isoquant I_4 is to be produced. Producing this amount requires a *minimum* of OC_4 units of capital inasmuch as any smaller amount would not permit one to attain the I_4 level of output. With OC_4 units of capital, OL_4 units of labor must be used. Beyond this level of input, additional units of labor in combination with OC_4 units of capital would yield a smaller level of output. To maintain the I_4 level of output with a greater labor input would require a greater input of capital as well—a palpably uneconomic use of resources.

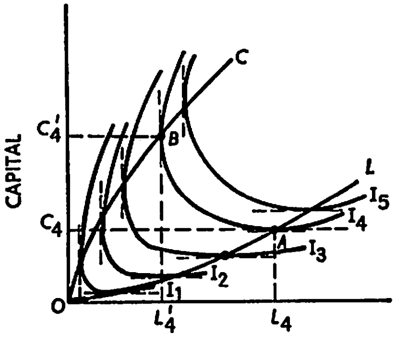


FIGURE 7.3.4

Full Isoquant Map and the Relevant Range of Production

Since an expansion of labor input beyond OL_4, in face of the constant capital input OC_4, reduces total output, point A on I_4 represents the intensive margin for labor. Its marginal product is zero, and hence the marginal rate of technical substitution of labor for capital is zero. This is shown by the horizontal tangent at point A. At this point labor has been substituted for capital to the maximum extent consistent with the level of output I_4.

Similarly, producing at the I_4 level requires a certain minimum input of labor, OL'_4 in Figure 7.3.4. The I_4 level cannot be attained without at least this much labor; and with this minimum amount additions to capital input beyond OC'_4 would reduce rather than

---

[6] Let x represent the capital-labor ratio and s the marginal rate of technical substitution. Then the elasticity of substitution (σ) is

$$\sigma = \frac{dx}{ds}\frac{s}{x}.$$

If the production function $Q = F(K, L)$ is linearly homogeneous in K and L, one may write : $Q = F(K, L) = LF(K/L, 1) = Lf(K/L)$. Dividing through and using the x-notation: $Q/L = f(x)$. It may be shown that in this notation the marginal product of capital is $f'(x)$ and the marginal product of labor is $f(x) - xf'(x)$. Hence the marginal rate of technical substitution is

$$s = \frac{f(x) - xf'(x)}{f'(x)}.$$

Taking the derivative of s with respect to x, inverting, and multiplying by the ratio of the marginal rate of technical substitution to the capital-labor ratio, one finds that the elasticity of substitution for linearly homogeneous production functions is

$$\sigma = -\frac{f'(x)[f(x) - xf'(x)]}{xf(x)f''(x)}.$$

augment output. Thus the marginal product of capital is zero at point B and negative for quantities in excess of OC'_4 units (in combination with OL'_4 units of labor). Since the marginal product of capital is zero, the marginal rate of technical substitution of capital for labor is zero at this point; capital is used to its intensive margin.

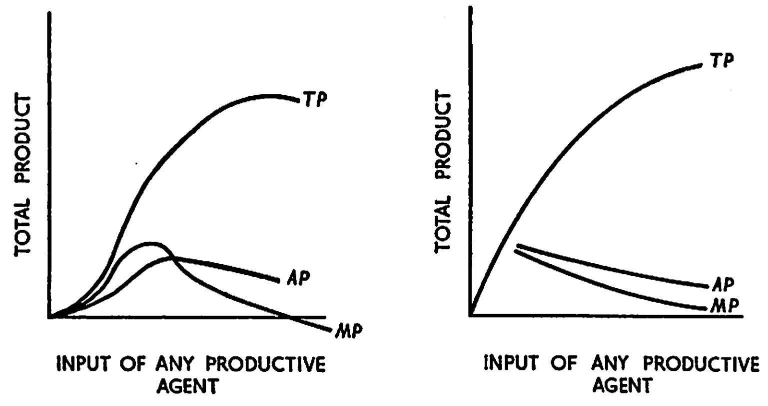
Now as we saw in section 6.3.e, the intensive margin with respect to one input is the extensive margin with respect to the other. Hence with OL_4 units of labor in use the average product of capital rises until OC_4 units are in use. Point A, therefore, separates stages I and II for capital and stages II and III for labor. In like manner, with OC'_4 units of capital the average product of labor rises until OL'_4 units are employed. Point B represents the boundary between stages I and II for labor and between stages II and III for capital.

By connecting the points of zero marginal labor product, the line OL is formed. Similarly, OC is the locus of points for which the marginal product of capital is zero. Production must take place within this range. Hence the "ridge" lines OL and OC separate the economic from the uneconomic regions of production. To summarize:

Relationships: if the production function is such that intensive and extensive margins for each input exist, the total isoquant map is like the one in Figure 7.3.4. Only those portions of the isoquants lying between the ridge lines (the loci of zero marginal products) are relevant to production. These economic portions of the isoquants are uniquely associated with stage II production of each input.

Stage I production for any input conforms to the region of rising average product; and if average product rises marginal product must exceed average product. Since a stage I area must be present to generate the isoquant map shown in Figure 7.3.4, the "normal" set of product curves—as shown in Figure 6.3.3—must be associated with the production function giving rise to this isoquant map. This "normal" set of product curves is reproduced in panel a, Figure 7.3.5.

Some production functions, however, generate isoquant maps such as that in Figure 7.2.3. There is neither a stage I nor a stage III range for either input. The entire production function represents stage II, or the economic region. Marginal and average products fall continuously, but neither reaches zero because there is not a maximum point on the total product curve. Such a production function is shown in panel b, Figure 7.3.5. The average and marginal product curves begin some distance from the origin. This is a mere convenience. They are both defined for infinitesimally small amounts of input; but at input levels less than unity average and marginal products exceed total product.



PANEL a

PANEL b

FIGURE 7.3.5

PRODUCT CURVES FOR DIFFERENT TYPES OF ISOQUANT MAPS

The importance of production functions giving rise to the product curves of panel b is an empirical question. For expository purposes, production functions giving rise to panel a are generally used. In empirical, statistical, and econometric applications, however, a broad class of production functions such as shown in panel *b* are most often used. The distinction, in fact, is relevant only in theory because observed production relationships are always those of stage II.[7]

---

[7] Two widely analyzed production functions of the sort shown in Figure 7.2.3 and panel b, Figure 7.3.5, are the Cobb-Douglas function and the Arrow-Chenery-Minhas-Solow function. Let q represent output and x_1 and x_2 denote two inputs. The Cobb-Douglas function has the form $q = ax_1^b x_2^c$, $0 < b, c < 1$, where a, b, and c are constants.

Exercise: Prove that marginal and average products always diminish and that the elasticity of substitution is precisely unity.

The Arrow-Chenery-Minhas-Solow function is

$$q = a\left[bx_1^{-c} + (1 - b)x_2^{-c}\right]^{-\frac{1}{c}},$$

where a, b, and c are again constants.

Exercise: Prove that marginal and average products always diminish and that the elasticity of substitution is $\frac{1}{1 + c}$.

The two production functions cited above are, in many respects, special cases, especially the Cobb-Douglas function. Two "cautions" are appropriate at this point.

First, the Cobb-Douglas function is very frequently written as a function homogeneous of degree one—that is, with b plus c equal 1. Unwary students, not exposed to other linearly homogeneous production functions, sometimes observe that the elasticity of substitution is unity (in the Cobb-Douglas case) and conclude that unitary elasticity of substitution and linear homogeneity are somehow related. They are not. The Cobb-Douglas function yields unitary elasticity of substitution irrespective of its degree of homogeneity—that is, whether $b + c = 1$ or not. In general, production functions homogeneous of degree one may be associated with *any* nonnegative value of the elasticity of substitution (including zero in the case of a Leontief, fixed-proportions function).

The Arrow-Chenery-Minhas-Solow function is a less special case in which the

7.3.e—Classification of Technological Progress

So far we have operated under the tacit assumption that a production function is both given and unchanging over the period of analysis; and our case has been strictly static. Technological progress does occur; and it is of some interest to classify the nature of technological change.

Many years ago Hicks defined technological progress as capital-using, neutral, or labor-using according as the marginal rate of technical substitution of labor for capital diminishes, remains unchanged, or increases at the originally prevailing capital-labor ratio. In other words, if technological change increases the marginal product of capital more than the marginal product of labor (at a given capital-labor ratio), progress is capital-using, because a producer now has an incentive to use more capital relative to labor because its (capital's) marginal product has increased relative to that of labor. The same type of statement holds, *mutatis mutandis,* for neutral and for labor-using technological progress.

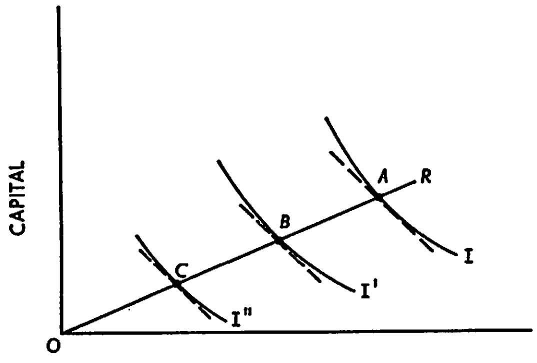
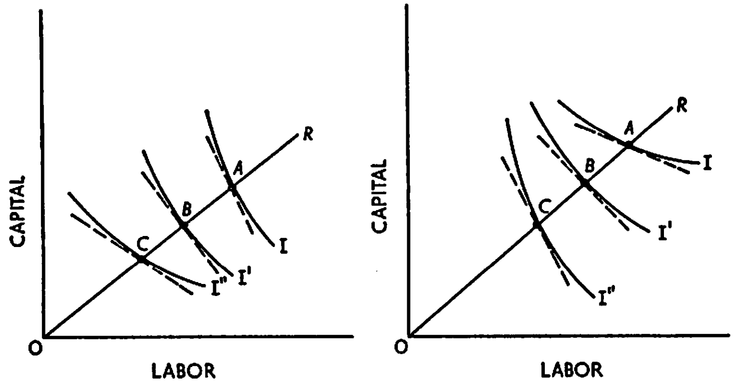


FIGURE 7.3.6

NEUTRAL TECHNOLOGICAL PROGRESS

Basically, technological progress consists of any change (graphically, shift) of the production function that either permits the same level of output to be produced with less input or enables the former level of inputs to produce a greater level of output.

Technological progress is shown graphically in Figures 7.3.6 and 7.3.7. The figures are constructed with uniform notation. The level of output is *I*, and the various isoquants (*I*, *I′*, and *I″*) show the combinations of inputs capable of producing this given level of output. *OR* is the ray whose slope gives a constant capital-labor ratio. The points *A*, *B*, and *C* show the points of production at the given capital-labor ratio as technological progress occurs.



PANEL a--CAPITAL-USING TECHNOLOGICAL PROGRESS

PANEL b--LABOR-USING TECHNOLOGICAL PROGRESS

FIGURE 7.3.7

BIASED TECHNOLOGICAL PROGRESS

Technological progress is shown graphically by a shift of an isoquant in the direction of the origin. In Figure 7.3.6, the three isoquants—I, I′, and I″—all represent the same level of output. As technological progress takes place *I′* shows that the given level of output can be produced by smaller quantities of inputs than at *I*. Similarly, as technological progress continues *I″* shows that still smaller input combinations can produce the same level of output.

Figure 7.3.6 illustrates neutral technological progress. Recalling the definition, technological progress is neutral if at a constant capital-labor ratio the marginal rate of technical substitution of labor for

---

elasticity of substitution is a *constant,* but not (necessarily) unity. But the elasticity of substitution may very well be a variable depending, for example, upon the input ratio. For a more detailed discussion, see Ronald W. Shephard, *Cost and Production Functions* (Princeton: Princeton University Press, 1953).

The second point concerns the shape of the full isoquant map in cases of linearly homogeneous production functions. As noted above, the Cobb-Douglas and the Arrow-Chenery-Minhas-Solow functions do not give rise to "uneconomic regions." But one should not conclude that functions homogeneous of degree one *cannot* display uneconomic regions. The isoquant map associated with such functions does not "look" exactly like the map in Figure 7.3.4. At the point of zero marginal product there is a point of inflection, the isoquant thereafter bending back to the left (rather than reaching a "maximum" and continuing to the right). For an excellent discussion that requires no mathematical background, see George Borts and E. J. Mishan, "Exploring the 'Uneconomic Region' of the Production Function," *Review of Economic Studies,* XXIX (1962), pp. 300–312.

capital is unchanged. The constant capital-labor ratio ray OR intersects the three isoquants at points A, B, and C, respectively. At these points, the slope of the isoquant—or the marginal rate of technical substitution—is the same. Hence it represents a shifting production function characterized by neutral technological progress.

Panels a and b, Figure 7.3.7, illustrate capital-using and labor-using technological progress, respectively. Capital-using technological progress occurs when, at a constant capital-labor ratio, the marginal product of capital increases relative to the marginal product of labor. In other words, since the marginal rate of technical substitution of labor for capital is the ratio of the marginal product of labor to that of capital, capital-using technological progress occurs when the marginal rate of technical substitution declines along a constant capital-labor ray. As one moves from A to B to C in panel a, the slope of the isoquant diminishes, representing a decline of the marginal rate of technical substitution. Hence this panel depicts a shifting production function characterized by capital-using technological progress.

By the same line of reasoning, panel b illustrates labor-using technological progress because the marginal rate of technical substitution increases as one moves from A to B to C.

7.4 OPTIMAL COMBINATION OF RESOURCES

So far the theory of production has been analyzed from the standpoint of an individual entrepreneur. However, nothing has been said regarding the *optimal* way in which he should combine resources. Any desired level of output can normally be produced by any of a number of different combinations of inputs. Our task now is to determine the specific combination a producer will select.

7.4.a—Input Prices and Isocosts

Inputs, just as outputs, bear specific market prices. In determining his *operating* input combination a producer must pay heed to relative input prices if he is to minimize the cost of producing a given output or maximize output for a given level of cost.

Input prices are determined, just as the prices of goods, by supply and demand in the market. For producers who are not monopsonists or oligopsonists, input prices are given by the market and his rates of purchase do not change them. Let us now concentrate upon a producer who is a perfect competitor in the input market, even though he may be a monopolist or an oligopolist in his output market. (Consideration of monopsony and oligopsony is deferred to Chapter 14.)

Let us continue to assume that the two inputs are labor and capital, although the analysis applies equally well to any two productive agents. Denote the quantity of capital and of labor by K and L, respectively, and their unit prices by r and w. The total cost C of using any volume of K and L is $C = rK + wL$, the sum of the cost of K units of capital at r per unit and of L units of labor at w per unit.

To take a more specific example, suppose capital costs \$1,000 per unit ($r = \$1{,}000$) and labor receives a wage of \$2,500 per man year ($w = \$2{,}500$). If a total of \$15,000 is to be spent for inputs, the equation above shows that the following combinations are possible: $\$15{,}000 = \$1{,}000\ K + \$2{,}500\ L$, or $K = 15 - 2.5\ L$. Similarly, if \$20,000 is to be spent on inputs, one can purchase the following combinations: $K = 20 - 2.5\ L$. More generally, if the fixed amount $\bar{C}$ is to be spent, the producer can choose among the combinations given by

$$K = \frac{\bar{C}}{r} - \frac{w}{r} L \,.$$

This is illustrated in Figure 7.4.1. If 15,000 is spent for inputs and no labor is purchased, fifteen units of capital may be bought. More generally, if $\bar{C}$ is to be spent and r is the unit cost, $\bar{C}/r$ units of capital may be purchased. This is the vertical-axis *intercept* of the line. If one unit of labor is purchased at \$2,500, two and five-tenths units of capital must be sacrificed; if two units of labor are bought, five units of capital must be sacrificed; and so on. Thus as the purchase of labor is increased the purchase of capital must be diminished. For each additional unit of labor, w/r units of capital must be foregone. In Figure 7.4.1, $w/r = 2.5$. Attaching a negative sign, this is the *slope* of the solid lines.

The solid lines in Figure 7.4.1 are called *isocost curves* because they show the various combinations of inputs that may be purchased for a stipulated amount of expenditure. In summary:

Relationship: At fixed input prices r and w for capital and labor, respectively, a fixed outlay $\bar{C}$ will purchase any combination of capital and labor given by the following linear equation:

$$K = \frac{\bar{C}}{r} - \frac{w}{r} L \,.$$

This is the equation for an isocost curve, whose intercept ($\bar{C}/r$) is the amount of capital that may be purchased if no labor is bought and whose slope is the negative of the input-price ratio (w/r).

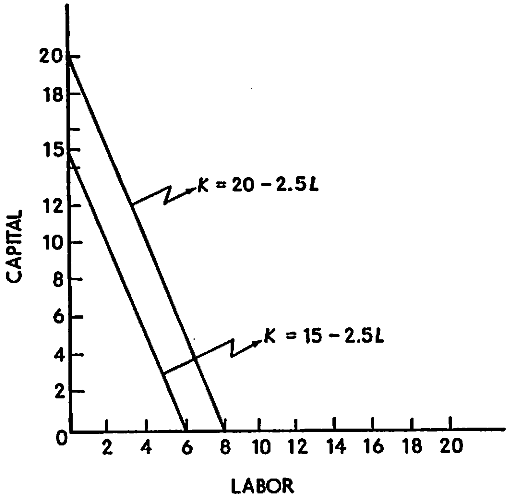


FIGURE 7.4.1

Isocost Curves for $r = \$1{,}000$ and $w = \$2{,}500$

7.4.b—Maximizing Output for a Given Cost

Suppose at given input prices r and w, a producer can spend only $\bar{C}$ on production. Subject to this input cost, he only operates efficiently if he maximizes the output attainable. To do this he must select the proper input combination. That is, among all input combinations he can purchase for the fixed amount $\bar{C}$, he must select the one that results in the greatest level of output.

Let the given level of cost $\bar{C}$ be represented by the isocost curve KL in Figure 7.4.2. The slope of KL is therefore equal to the (*negative*) ratio of the price per unit of labor to the price per unit of capital. I_1, I_2, and I_3 are isoquants representing various levels of output. First, observe that the I_3-level of output is not obtainable because the *available* input combinations are limited to those lying on or beneath the isocost curve KL.

Next, the producer could operate at points such as R and S. At these two points, the input combinations *required* to produce the I_1-level of output are *available* for the given cost represented by the isocost KL. In this case, however, output can be increased without incurring additional cost by the selection of a more appropriate input combination. Indeed, output can be expanded until the I_2-level is reached—the level at which an isoquant is just tangent to the specified isocost curve. A greater output is not obtainable for the given level of expenditure; a lesser output is inefficient because production can be expanded at no additional cost. Hence the input combination represented by the slope of the ray OQ is optimal because it is the combination that maximizes output for the given level of cost.

After studying the theory of consumer behavior, this proposition should be more or less obvious. However, a sound reason is behind it. For a moment suppose the entrepreneur contemplated producing at point R. The marginal rate of technical substitution of labor for capital —given by the slope of the tangent TT'—is relatively high. Suppose it is 3:1, meaning that one unit of labor can replace three units of capital at that point. The relative input price, given by the slope of KL, is much less, say 1:1. In this case, one unit of labor costs the same as one unit of capital but it can replace three units of capital in production. The producer would obviously be better off if he substituted labor for capital. The opposite argument holds for point S, where the marginal rate of technical substitution is less than the input-price ratio.

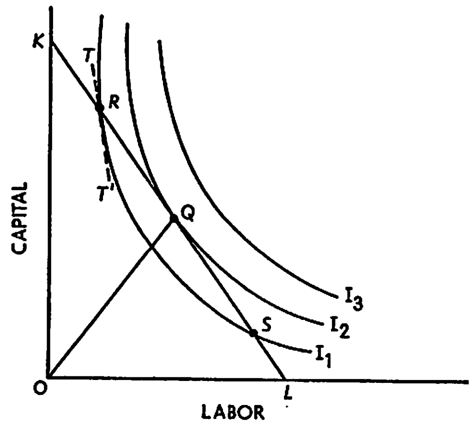


FIGURE 7.4.2

Optimal Input Combination to Maximize Output Subject to a Given Cost

Following this argument, the producer reaches equilibrium (maximizes output for a given level of cost) only when the marginal rate of technical substitution of labor for capital is equal to the ratio of the price of labor to the price of capital. The market input-price ratio tells the producer the rate at which he can substitute one input for another in purchasing. The marginal rate of technical substitution tells him the rate at which he can substitute in production. So long as the two are not equal, a producer can achieve either a greater output or a lower cost by moving in the direction of equality.[8]

Principle: to maximize output subject to a given total cost and given input prices, the producer must purchase inputs in such quantities as to equate the marginal rate of technical substitution of labor for capital with the input-price ratio (the price of labor to the price of capital). Thus

$$MRTS_{L \text{ for } K} = \frac{MP_L}{MP_K} = \frac{w}{r}.$$

---

[8] Let the production function be $Q = f(K, L)$. Total cost is $C = rK + wL$. Maximizing output subject to a *given* cost and given input prices is a simple exercise in the Lagrange technique for constrained extrema. Introduce the multiplier λ and construct the Lagrange function

$$f(K, L) - \lambda(rK + wL - \bar{C}). \tag{7.4.1}$$

Taking the first partial derivatives, one obtains

$$\begin{aligned} &\frac{\partial f}{\partial L} - \lambda w = 0 \cdot \\ &\frac{\partial f}{\partial K} - \lambda r = 0 \cdot \end{aligned} \tag{7.4.2}$$

Transferring the second term to the right-hand side, one obtains

$$\begin{aligned} &\frac{\partial f}{\partial K} = \lambda w , \\ &\frac{\partial f}{\partial K} = \lambda r , \end{aligned} \tag{7.4.3}$$

or

$$\frac{\frac{\partial f}{\partial L}}{w} = \frac{\frac{\partial f}{\partial K}}{r} = \lambda \cdot \tag{7.4.4}$$

This is one way in which the principle may be stated: in equilibrium, the marginal product per dollars worth of input must be the same for each input.

Rearranging equation (7.4.4):

$$\frac{\frac{\partial f}{\partial L}}{\frac{\partial f}{\partial K}} = MRST_{L \text{ for } K} = \frac{w}{r}, \tag{7.4.5}$$

the principle as stated in the text.

The second-order conditions for a constrained maximum require that the bordered determinant

$$\begin{vmatrix} 0 & f_K & f_L \\ f_K & f_{KK} & f_{KL} \\ f_L & f_{LK} & f_{LL} \end{vmatrix} \tag{7.4.6}$$

be negative definite, where

$$f_K = \frac{\partial f}{\partial K}, f_{KL} = \frac{\partial^2 f}{\partial K \partial L}, \text{ etc.}$$

Expanding, this requires that

$$f_L^2 f_{KK} - 2 f_K f_L f_{KL} + f_K^2 f_{LL} < 0. \tag{7.4.7}$$

Condition (7.4.7) means that the isoquants *must* be concave from above, as previously indicated.

7.4.c—Minimizing Cost Subject to a Given Output

As an alternative to maximizing output for a given cost, an entrepreneur may seek to minimize the cost of producing a stipulated level of output. The problem is solved graphically in Figure 7.4.3. The isoquant I represents the stipulated level of output, while C_1, C_2, and C_3 are isocost curves with the same slope (input-price ratio).

First notice that the level of cost represented by C_1 is not feasible because the I-level of output is not physically producible by any input combination available for this outlay. Next, the I-level could be produced, for example, by the input combinations represented by the points R and S, both at the cost level C_3. But by moving either from R to Q or from S to Q, the entrepreneur can obtain the same output at lower cost.

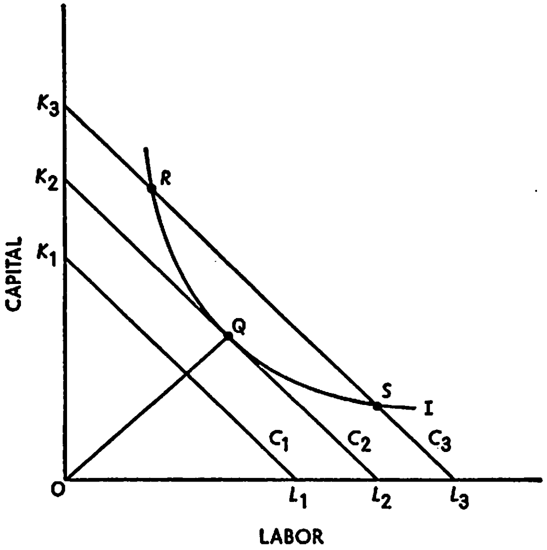


FIGURE 7.4.3

OPTIMAL INPUT COMBINATION TO MINIMIZE COST SUBJECT TO A GIVEN LEVEL OF OUTPUT

By the very same arguments used in subsection 7.4.b, a position of equilibrium is attained only at point Q where the isoquant is just tangent to an isocost curve. Thus in equilibrium the marginal rate of technical substitution of labor for capital must equal the ratio of the price of labor to the price of capital. The previous principle may thus be elaborated.[9]

---

[9] Using the notation of footnote 8, we now wish to minimize $rK + wL$ subject to producing $\overline{Q} = f(K, L)$ units of output. Again, introduce the Lagrange multiplier λ and

Principle: in order either to maximize output subject to a given cost or to minimize cost subject to a given output, the enterpreneur must employ inputs in such amounts as to equate the marginal rate of technical substitution and the input-price ratio.

7.4.d—Effect of Changes in Input Price upon Input Usage

From Part I, you know that a change in the price of a good has two theoretically discernible effects: the substitution effect and the income effect. The substitution effect is always negative and the income effect normally reinforces it. Much the same type of effects can be isolated for changes in input price. In this case, both substitution and output effects are always negative.

First, consider Figure 7.4.4. This graph illustrates increases in the price of labor inputs, the price of the capital input remaining constant. The original price ratio (w/r) is given by the slope of the line KL_1. As this input-price line shifts leftward to KL_2 and KL_3, the price of labor increases because the same total expenditure on labor will at first purchase OL_1 units, then OL_2 units, and finally only OL_3 units.

The substitution and output effects are shown in Figure 7.4.5. The original point of equilibrium is Q. The level of production is indicated by the isoquant I_1, the input-price ratio by the slope of the input-price line KL_1; and Ok_1 units of capital and Ol_1 units of labor are used. We now assume a rise in the price of labor, the price of capital remaining constant. This shifts the input-price line to KL_2. If the producer attempts to maximize output for a given cost, the equilibrium changes from Q to S, the level of output falling to that indicated by the isoquant I_2. In the ultimate equilibrium position, Ok_3 units of capital and Ol_3 units of labor are used. The total effect upon labor usage is, therefore, a decrease from Ol_1 to Ol_3, or a reduction of l_1l_3 units of labor.

The total effect may be decomposed into two components. The change in labor usage attributable *exclusively* to the change in relative input prices is called the *substitution effect.* To determine this effect graphically, construct the fictitious input-price line $K'L'$. This line is constructed so there is an equilibrium at the *old* output level and the *new* input prices. In other words, the rise in input prices has been compensated by an increase in expenditure sufficient to maintain the old level of output. A fictitious equilibrium is reached at point R, and the movement from Q to R represents the substitution effect, the change in input usage attributable only to the change in relative input prices, the level of output remaining constant. In input units, the substitution effect reduces labor input from Ol_1 to Ol_2, or by the amount l_1l_2. Capital is substituted for labor, increasing capital usage from Ok_1 to Ok_2, or by k_1k_2.

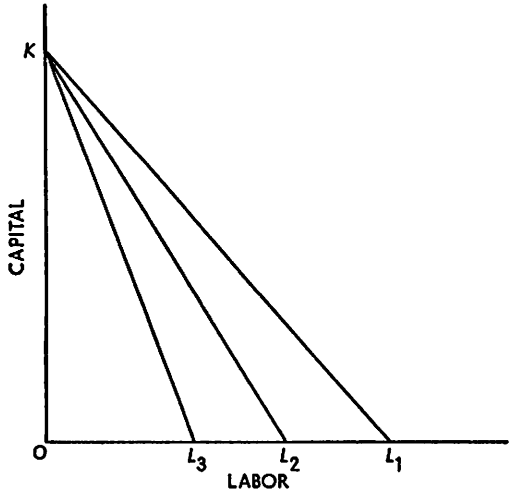


FIGURE 7.4.4

SHIFTING ISOCOST CURVES TO SHOW AN INCREASE IN THE PRICE OF LABOR

When an input price increases, however, there must be a decrease in output if the level of expenditure does not increase. The *output* effect is represented by the shift from the fictitious equilibrium point R on I_1 to the ultimate equilibrium point S on I_2. The output effect leads to a reduction in labor input from Ol_2, to Ol_3, or by the amount l_2l_3. Capital usage is also reduced by the output effect, from Ok_2, to Ok_3, or by k_2k_3.

---

construct the function

$$(7.4.8)\qquad rK + wL - \lambda[f(K, L) - \bar{Q}].$$

Setting the first partial derivatives equal to zero, one obtains

$$(7.4.9)\qquad \begin{aligned} w - \lambda\frac{\partial f}{\partial L} &= 0\,, \\ r - \lambda\frac{\partial f}{\partial K} &= 0\,. \end{aligned}$$

Transferring the second term on the left-hand side of each equation to the right-hand side and dividing the first equation by the second to eliminate λ, one obtains

$$(7.4.10)\qquad \frac{w}{r} = \frac{\dfrac{\partial f}{\partial L}}{\dfrac{\partial f}{\partial K}} = MRTS_{L \text{ for } K}\,,$$

precisely the result obtained in the text and in footnote equation (7.4.5).

The total effect upon labor usage of a rise in the price of labor is simply the sum of the two effects:

$$\underset{(Total\ effect)}{l_1l_3} = \underset{(Substitution\ effect)}{l_1l_2} + \underset{(Output\ effect)}{l_2l_3}$$

Summarizing, we have the following:

Relationship: the total effect of a change in the price of an input upon the usage of this input may be decomposed into two components. The substitution effect shows the change in input usage attributable exclusively to the change in relative input prices, output being held constant. This effect is always negative in that a rise in input price leads to a reduction, and a fall in input price to an increase, in the usage of the input. The output effect shows the change in input usage attributable exclusively to a change in the level of output, input prices remaining constant. This effect is also always negative because an increase in cost leads to a reduction in both output and input, while a decrease in cost results in augmented output and input.

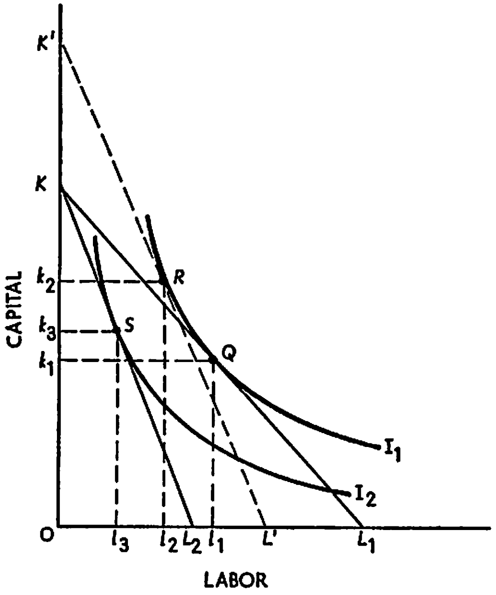


FIGURE 7.4.5

OUTPUT AND SUBSTITUTION EFFECTS OF A RISE IN THE PRICE OF LABOR

7.5 CONCLUSION

Chapters 6 and 7 contain an explanation of the theory of production and of the optimal combination of inputs when input prices are known. We turn next to the theory of cost, which relies upon the physical laws of production and upon the prices an entrepreneur must pay for his inputs.

SUGGESTED READINGS

1. Borts, George H. and Mishan, E. J. "Exploring the 'Uneconomic Region' of the Production Function," *Review of Economic Studies,* XXIX (1962), pp. 300–12.
2. Cassels, John M. "On the Law of Variable Proportions," *Explorations in Economics,* pp. 223–36. New York: McGraw-Hill Book Co., 1936.
3. Cobb, C. W. and Douglas, P. H. "A Theory of Production," *American Economic Review* (supplement), XVIII (1928), pp. 139–65.
4. Hicks, John R. *The Theory of Wages.* London: Macmillan & Co., Ltd., 1932.
5. Hicks, John R. *Value and Capital,* pp. 78–98. 2d ed. Oxford: Oxford University Press, 1946.
6. Lerner, A. P. "The Diagrammatical Representation of Elasticity of Substitution," *Review of Economic Studies,* I (1933–34), pp. 68–71.
7. Machlup, Fritz. "The Commonsense of the Elasticity of Substitution," *Review of Economic Studies,* II (1934–35), pp. 202–13.
8. Robinson, Joan. "The Classification of Inventions," *Review of Economic Studies,* V (1937–38), pp. 139–42.
9. Henderson, James M. and Quandt, Richard E. *Microeconomic Theory: A Mathematical Approach,* pp. 47–53. New York: McGraw-Hill Book Co., Inc., 1958. [Elementary math necessary.]
10. Samuelson, Paul A. *Foundations of Economic Analysis,* pp. 57–76. Cambridge: Harvard University Press, 1947. [Advanced math necessary.]