ARAB REPUBLIC OF EGYPT

THE INSTITUTE OF NATIONAL PLANNING

Memo No (566)

Lectures on Production Theory and Techniques

by

ABDUL QAYUM

Economic Expert

1965

جمهورية مصر العربية – طريق صلاح سالم – مدينة نصر – القاهرة – مكتبب ريد رقم ١١٧٦٥

A.R.E Salah Salem St. Nasr City , Cairo P.O.Box : 11765

(566)

United Arab Republic

INSTITUTE OF NATIONAL PLANNING

Lectures on Production Theory and Techniques

BY

ABDUL QAYUM

Economic Expert

INP

Cairo

1965

CONTENTS

PREFACE

This is the first draft of my lectures on 'Production Theory and Techniques' prepared on the suggestion of Professor Bent Hansen with whom I had the good fortune to teach the so-called Advanced Economic Theory. As Prof. Hansen so rightly pointed out, Production Theory forms the backbone of all Growth Theory and the Growth Models differ from each other according to the Production Function they use or imply. (There cannot be any growth theory without production theory). Thus the Production Theory occupies the same or even more crucial position in the Post World War II 'Growth Era' as the Consumption Theory did in the Inter-War 'Business Cycles Era'. Unfortunately the pure economists cannot be expected to be at their best in Production Theory as they, along with statisticians, were in Consumption Theory. Any really worthwhile invention or discovery in Production Theory is more likely to come from engineers and physicists rather than from economists or econometricians. And it is for nought that the former group has been often more successful in programming and planning than the latter in the recent years.

I am extemely grateful to Prof. Hansen for his continuous ecouragement and suggestions. He has kindly glanced through all the manuscripts and pointed out gross errors. Many smaller errors remain which I discovered later. Also in one or two cases the derivation is incomplete, I shall correct and complete them if and when I revise these lectures.

I am not giving any bibliography. For this, reference may be made to the following two excellent survey articles, the first of which I was able to avail of, but the second I could not, as it appeared too late.

1) A.A.Walters: "Production and Cost Functions: An Econometric survey" Econometrica 1963.

2) F.H.Hahn and R.C. Mathews: "The Theory of Economic Growth: "A Survey" The Economic Journal 1964

My thanks are due to Mrs.A.Habib for her kindly undertaking the typing of all the Masters and the appropriate insertions and deletions which were required quite often and also to Miss Faten Fouad for typing my hand written manuscripts. If some errors remain, they are due to my own negligence.

A. Qayum.

Dr. A. Qayum

MICRO-PHYSICAL PRODUCTION FUNCTION.

I

Production function is a technically extremal concept. It implies that either the maximum amount of output is produced with the given amounts of inputs or the given amount of output is produced with the minimum amount of each of the inputs, the other inputs being given. Symbolically we can express a production function as

$$(1) \qquad x = f(u, v, w, \ldots \ldots \ldots),$$

where x is the amount of output X and u, v, w, ... are the amounts of inputs U, V, W, ... etc. There are two crucial assumptions about the techniques and the inputs which play a great role in the development of the theory of production. One is the assumption that the techniques are continuously variable which means that the inputs can be combined in any proportion we desire in producing the output, i.e., f is a continuous function in u, v, w, ... In reality this is not always true. The techniques or processes of production are not continuously variable, that is, the inputs cannot be combined in any proportion desired, to produce a certain output. This means that in reality f is not continuous. It must, however, be stressed that there is a difference between the variability in the combination of the commodity inputs and variability in the combination of the basic factors of production, labour and capital. If we retrace the variability in combination of inputs in the production of final output to the variability in the combination of inputs in producing the inputs which will go to produce the final output and so backwards till we reach the stage when the original factors labour and capital are combined to produce inputs at the first stage, we can visualize easily that the variability in the combination of the basic factors of production, labour and capital is likely to be much greater than that in the combination of inputs in producing the final inputs. It is not suggested here that labour and capital are continuously variable or perfectly substitutable but it is maintained that the reality can be roughly approximated by assuming that a continuous variability exists in the combination of factors. Because of this and because of the fact that the bulk of the theory so far developed and almost all of the theory developed earlier is based on the assumption of continuous variability, we shall discuss the theory in a major part of this series of lectures on the assumption of continuous variability of factor combination. In the concluding lectures we shall discuss the case when the processes of production are not continuously variable, but discrete.

The second crucial assumption relating to the development of the theory is that the inputs are continuously divisible. This is again not a wholly realistic assumption. However, as regards the divisibility, the commodity inputs fare better than some of the factors such as built capital, entrepreneurship and organizational capacity. The assumption of divisibility as regards the commodity inputs is largely realistic, but very tough problems are faced in connection with the non-divisibility of the factors of production mentioned above. In the earlier discussions, we shall assume that the factors and inputs are continuously divisible, we shall take up the problem of non-divisibility and rigidity of factors in later lectures.

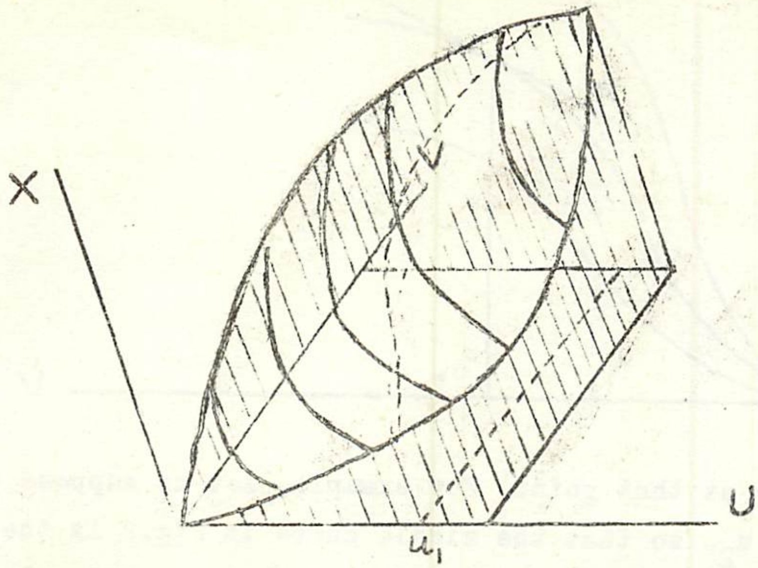
II

Our discussion will be much simplified if we include two factors U and V in the production function and this will also enable us to use graphic illustrations. Let the production function, then, be

(2) $x = f(u,v)$

Plotting for different values of u and v taken along the axes OU and OV and for x taken along OX vertically, we can construct a production surface, as shown below.

Fig.1.



The production surface can assume numerous forms. For the sake of simplicity we suppose that it takes the three dimentional form as shown in Fig.1. By using the method of plane sections, we can derive several useful results from (2). If we keep one of the factors, say U, constant at u, we can study the variation of X with respect to the other factor V. This is shown by the vertical section of the production surface by a plane parpendicular at u_1, u_2, etc., the resulting curves are given by

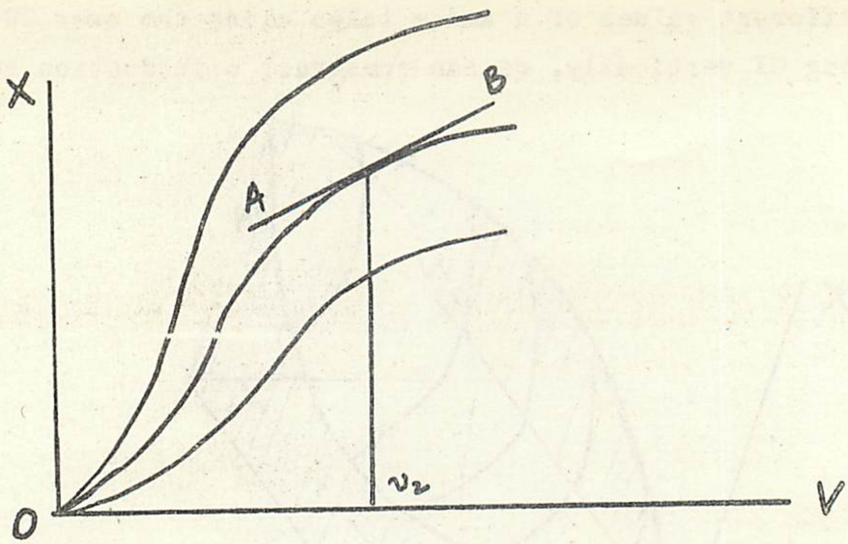
(3) $$x = f(u_1, v) = \phi_1(v)$$

(4) $$x = f(u_2, v) = \phi_2(v)$$

The curves for different values of u can be represented in the two dimensional plane v,x as in Fig.2. The curves in Fig.2 express output x in relation to variable v when u is held fixed in different quantities. Similarly we can hold v fixed in different quantities and demonstrate the variation of x in relation to u. In Fig.1, the dotted lines represent a vertical plane section at u_1 giving respectively the foot of the plane and the curve in which it cuts the surface. Curves like the dotted one in Fig.1 can be represented in two dimensional plane as done in Fig.2.

Slopes of the individual input-output curves shown in Fig.2 at points corresponding to different values of v give the physical marginal

Fig. 2.



productivity of v at that point. For example, let us suppose that the value of u is fixed at u_2, so that the middle curve in Fig. 2 is the relevant curve giving the variation in output x with the variable v. The slopes of tangents at different points of this curve give the marginal productivity of v at its corresponding values. In Fig. 2 the slope of AB gives the marginal productivity of v at its value equal to v_2. Similarly we can draw input-output curves relating to the variable u, with v held fixed at certain levels, by drawing a plane section perpendicular to the v axis at points corresponding to those levels. Symbolically the marginal productivity of v at v_2 when u is fixed at u_2 is

$$(5) \qquad \frac{\partial x}{\partial v} = \frac{d\phi_2(v)}{dv} = \phi_2'(v_2) \quad , \quad (u = u_2)$$

Similarly we can express the marginal productivity of u.

Another quantity which may be of some interest especially in analysis relating to product imputation, is production elasticity. Equations (3) or (4) express the variation of X with changes in V when U is fixed at u_1 or u_2. From these we can derive the elasticity of production with respect to v, the formula of elasticity of production is

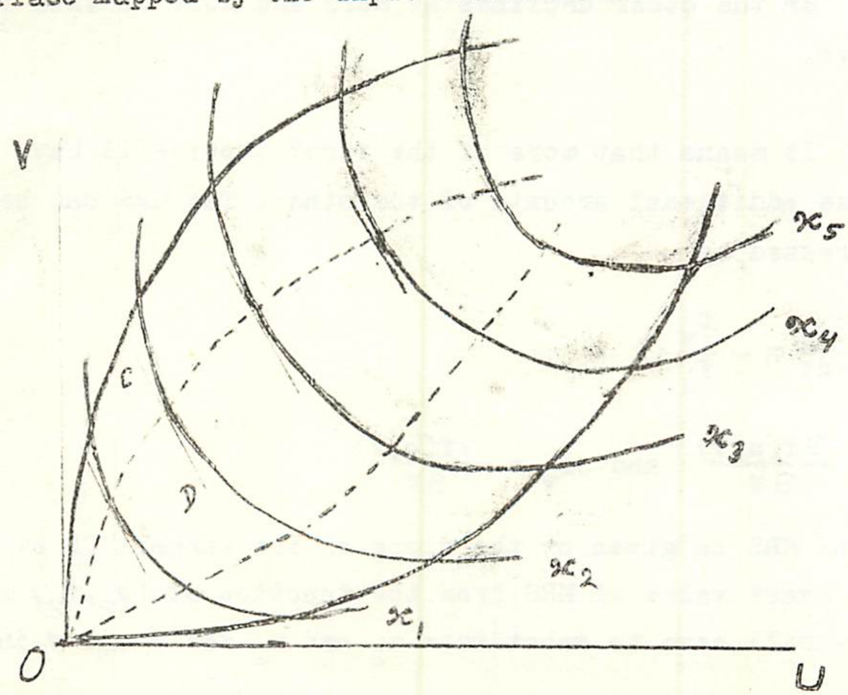
$$(6) \qquad e = \frac{dx}{x} \Big/ \frac{dv}{v} = \frac{v}{x} \cdot \frac{dx}{dv}$$

$e >/< 1$, according as increasing, constant or decreasing returns to factor V prevail in production. And similarly for factor U.

III

A second relationship of great importance is obtained through intersecting the production surface not by plane sections perpendicular to the coordinates representing the inputs but by a plane section perpendicular to the ordinate representing the output, i.e., to OX in Fig.1. The plane section will intersect the surface in a curve, each point of which will represent a combination of the two inputs yielding the same level of output, i.e. the level of output at which the perpendicular plane-section has been erected. By drawing plane-sections perpendicular to the vertical output axis at different level, we can have a set of curves, every point on each of which represents a different combination of inputs yielding the same level of output. In Fig.3 we draw the contours of these curves by projecting them on the horizontal surface mapped by the input coordinates.

Fig.3.



In Fig.3 we have drawn only 5 of the infinite number of the contour lines corresponding to each point on OX in Fig.1. As each point on a curve represents the same level or amount of output, they are called isoquants or product isoquants. For example, each point on curve x_2 represents different combinations of input u and v, but gives the same level of output x_2. Symbolically, the curve x_2 can be expressed by substituting x_2 for x in (2), i.e.

$$(7) \qquad x_2 = f(u,v)$$

Similarly we can express the function for other isoquants.

The slope of an isoquant at a certain point indicates the marginal rate of substitution (or MRS for short) of one input for the other, if the output is to be maintained at the specified level. In normal cases, when the isoquants are curved as they are shown in Fig.3, the rate of substitution of one factor for the other declines as more and more of this factor is replaced by the other.

It means that more of the first input will have to be given away for the same additional amounts of the other. The MRS can be derived from (3) and is expressed as

$$(8) \qquad \frac{du}{dv} = -\frac{f_v}{f_u}$$

where $f_u = \dfrac{\partial f(u,v)}{\partial u}$ and $f_v = \dfrac{\partial f(u,v)}{\partial v}$

In Fig.3 the MRS is given by the slope of the tangent CD at point (u_2,v_2). To get the exact value of MRS from the function at (u_2,v_2) relating to isoquant x, we will have to substitute u_2 and v_2 for u and v in (8).

We can use another type of plane sections of the production surface with advantage. These planes may be taken through the vertical output axis OX at different angles between OU and OV relating to different combinations of U and V. For instance, if a plane is taken through OX at an angle 30°

from OU, amd so 60° from OV, the ridge line that we get from the intersection of this plane and the production surface gives the growth of output when U and V are combined in proportion 2 to 1 at each magnitude of production.

Thus, if a point at the foot of the plane is (u_1, v_1), then any other point at the foot of this plane through OX can be expressed as $(\lambda u_1, \lambda v_1)$ and the production function corresponding to these points giving a proportionate change in inputs can be written as

$$x = f(\lambda u_1, \lambda v_1) = \Theta(\lambda) \tag{9}$$

(9) is a function of the variable λ only. We can get different functions like (9) by taking planes through OX and passing through points other than (u_1, v_1).

As indicated above, the slope of an individual isoquant varies from one point to the other. If we join the points on each of the isoquants at which the respective isoquants have the same slope, we will get a line wich passes through all the successively higher isoquants of the set at points where they have the same slope or they are equally 'inclined'. These lines may be called isoclines. They are shown in Fig.3 by the dotted lines. These isoclines also connect the points on the production surface where the marginal rates of substitution between the factors are equal, hence they trace (and also they are called) expansion paths. That is, they show the path of expansion on which the economy will move, if the factors at the margin are combined in such a way that their rates of substitution remain unaltered. In a competitive equilibrium, when the ratio of factor prices tends to equal their marginal rate of substitution, this means that if the prices of factors are given and remain constant, then the production in the economy will expand along the corresponding isocline or the expansion path.[1] The isoclines can be expressed by

---

1) When the prices of output and, factors are given, the cost can be mimised if $\frac{dC}{dx} = \frac{d(p_u u + p_v v + b)}{dx} = 0$ or $\frac{p_u}{p_v} = -\frac{d_v}{d_u}$.

(10) $$\frac{du}{dv} = k'$$

where k' is any constant, representing the rate of substitution between the factors.

A word may also be added to the range of the isoquants that are relevant to economic analysis. In normal circumstances, product can only be maintained if less of one input is used by increasing the use of the other. Alternatively, as the substitution of one input for the other proceeds, increasingly larger additions of the first input are needed to compensate for a given reduction of the other. In other words only that portion of an isoquant is interesting and relevant, along which the isoquant is convex to the origin.[1] In Fig.3 we have shown the range of the isoquants which are convex to the origin by thick lines. At all points between these two lines, each isoquant satisfies the following:

(11) $$\frac{du}{dv} < 0 \qquad \frac{d^2u}{dv^2} > 0 .$$

IV

We have already stated that the MRS measures the rate at which one resource, say U, is substituted for the other, viz. V, in the production of output X from the given combination of resources at an isoquant. The positive value of MRS can be denoted by r, so that

(12) $$r = -\frac{du}{dv} = \frac{f_v}{f_u}$$

Now one more interesting point to determine is how fast r changes in relation to the rate of change in the ratio of the resources at an isoquant. This will

1) It is just possible that the isoquants are convex to the origin all along their course.

give us what is called the elasticity of rate of substitution and can be denoted by σ, so that

$$(13) \qquad \sigma = \frac{d\frac{u}{v}}{\frac{u}{v}} \div \frac{dr}{r}$$

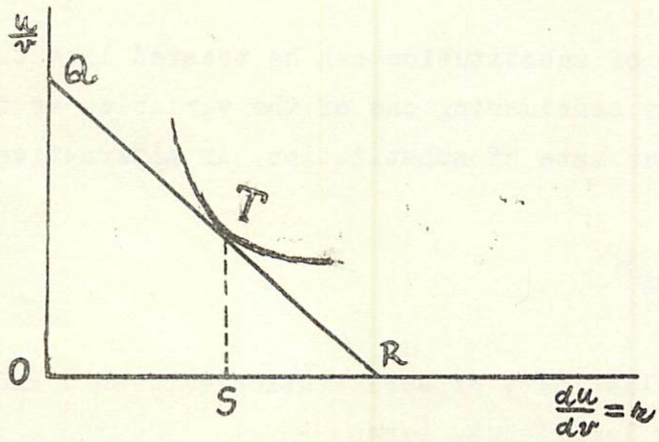
The value of σ can be expressed in an alternative form.[1)]

$$(14) \qquad \sigma = \frac{f_u f_v (uf_u + vf_v)}{-uv(f_{uu} f_v^2 - 2f_{uv} f_u f_v + f_{vv} f_u^2)}$$

(14) shows that the elasticity of substitution is symmetrical for the two resources. This can be readily seen because interchanging u and v in the right hand side of (14) does not change the expression. In other words, whether we consider substitution of U for V or V for U, we get the same value of the elasticity of substitution.

In diagrammatic terms if P is any point on the isoquant x_2, in Fig.3, then σ is the ratio of the relative change in the gradient of OP to the relative change in the gradient of the tangent at P to the isoquant, for a small movement of P along it. Alternatively we can illustrate σ by the following figure

Fig.4.



1) R.G.D. Allen, Mathematical Analysis for Economists (London 1960) p. 342. (14) can be derived immediately by substituting for $d\frac{u}{v}$ and dr in (13).

$d\frac{u}{v} = \frac{vdu - udv}{v^2}$ and $dr = \frac{\partial r}{\partial u} du + \frac{\partial r}{\partial v} dv$. But $du = -\frac{f_v}{f_u} dv = rdv$

$d\frac{u}{v} = -\frac{vr + u}{v^2} dv$ and $dr = -r(\frac{\partial r}{\partial u} - \frac{\partial r}{\partial v}) dv$.

In Fig.4 the ordinate represents the ratio of factors and the abscissa the rate of substitution between them. The curve represents the relationship between these two variables. At point T, the slope of the tangent QTR gives the marginal rate of change of $\frac{u}{v}$, the ratio of factors, with respect to $r = \frac{du}{dv}$, the marginal rate of substitution between them, so that

$$\frac{OQ}{OR} = \frac{d\frac{u}{v}}{dr}$$

Further we have $$\frac{TS}{OS} = \frac{\frac{u}{v}}{\frac{du}{dv}}$$

But $$\frac{TS}{OQ} = \frac{TR}{QR} \quad \text{and} \quad \frac{OS}{OR} = \frac{QT}{QR}$$

$$\frac{TS}{OS} = \frac{TR}{QR} OQ \times \frac{QR}{OR \cdot QT} = \frac{TR \cdot OQ}{OR \cdot QT}$$

$$\therefore \sigma = \frac{d\frac{u}{v}}{dr} \cdot \frac{r}{\frac{u}{v}}$$

$$= \frac{OQ}{OR} \cdot \frac{OS}{TS} = \frac{QT}{TR}$$

Thus the elasticity of substitution can be treated like the elasticity of any two variables by considering one of the variables as the ratio of the factors and the other rate of substitution. An alternative expression for σ may be

$$\frac{u}{v} = \left(\frac{du}{dv}\right)^{\sigma}$$

where σ gives the elasticity of substitution between U and V. Or we can express the same in logarithmic terms

$$\frac{d \log \frac{u}{v}}{d \log \frac{du}{dv}} = \sigma$$

which is the same as (13).

As shown by Allen , (13) can be also written in the form

(15) $$\sigma = \frac{r}{uv} \cdot \frac{vr + u}{r \cdot \frac{dr}{du} - \frac{dr}{dv}}$$

However, $r \cdot \frac{dr}{du} - \frac{dr}{dv} = \frac{d^2u}{dv^2}$, which is the curvature of the isoquant at (u,v) of the isoquant at a specific point. The magnitude of σ indicates the ease with which the output can be maintained by substituting one resource for the other. If $\frac{d^2u}{dv^2}$ is zero, which will be the case, if the isoquant assumes the form of a straight line in the neighbourhood of (u,v), σ will be infinite which means that the product can be maintained by substituting one resource for the other in a constant proportion. If $\frac{d^2u}{dv^2}$ is ∞ which will be the case if the isoquant has a right angle at (u,v) then $\sigma = 0$, indicating that U and V are needed in fixed proportion, and no substitution is possible, and an increase in one of the resources without increasing the other will leave the output unchanged.

V.

In production theory as in other branches of economic theory, the homogeneity of functions is of great significance. A production function in variables u and v is homogeneous of degree n if for any value of λ,

(16) $$f(\lambda u, \lambda v) = \lambda^n f(u,v) = \lambda^n x$$

If n = 0, f is a homogeneous function of degree zero, so that

$$f(\lambda u, \lambda v) = f(u,v) = x$$

This means that if f is a homogeneous function of degree zero, then the value of the function does not change, if the variables are changed in an equal proportion, say λ.

(17) $$x = \frac{au^m}{bv^m} = \frac{a(\lambda u)^m}{b(\lambda v)^m}$$

where m may assume any real value. In (17) the value of x remains unchanged when u and v are changed in the same proportion. Function homogeneous of zero degree are of great importance in the theory of demand and supply, though not so in production theory as such.

However, a very important case in production theory arises when $n = 1$ in (16), i.e. when the production function is homogeneous of degree one; for in that case

$$(18) \qquad f(\lambda u, \lambda v) = \lambda f(u,v) = \lambda x$$

(18) states that when f is homogeneous of degree one then the output x is changed in the same proportion in which the factors u and v are simultaneously changed. That is if we double u and v, i.e. when $\lambda = 2$, x will also be double and so on. This is the case of <u>constant returns to scale</u> which means that production or output is changed at the same scale at which all the factors are changed.

But if in (16) $n \neq 1$, then we cannot get constant returns to scale but we get varying returns to scale. If $n > 1$, we get increasing returns to scale, for in this case

$$(19) \qquad f(\lambda u, \lambda v) = \lambda^n f(u,v) = \lambda^n x > \lambda x \qquad (\lambda > 1).$$

It is evident from (19) that when $n > 1$, and $\lambda > 1$, then the resulting output increases more than the proportion in which the factors are increased. In the reverse case when $n < 1$, we get decreasing returns to scale, for then

$$(20) \qquad f(\lambda u, \lambda v) = \lambda^n f(u,v) = \lambda^n x < \lambda x.$$

It is obvious from (20) that when $n < 1$, $(\lambda > 1)$, then the resulting output increases less than the proportion in which the factors are increased.

We see then that when the production function is homogeneous we can easily determine if the function depicts increasing, constant or decreasing returns to scale. All that we have to notice is the degree of homogeneity i.e. the value of n. The function characterises,

increasing returns to scale	according as $n > 1$
constant returns to scale	" " $n = 1$
decreasing returns to scale	" " $n < 1$

The determination of the returns to scale is not so easy or clear-cut when the production function is not homogeneous in which case there may be increasing returns to scale in one interval and decreasing returns in another.

In production theory homogeneity of degree one is very important for another reason. The well-known Euler's theorem states that the following condition is satisfied by a homogeneous function.[1]

(21) $$u f_u + v f_v = n f(u,v) = nx$$

In the case when f is homogeneous of degree one we have

(22) $$u f_u + v f_v = f(u,v) = x$$

(22) states that the sum of the factors multiplied by their marginal products, i.e. $u f_u + v f_v$ equals the total output x. This property enables us to impute back the total output to the factors that produce it.[2] Under ideal circumstances, it should - according to the marginal productivity theory - be possible to reallocate the total output back to the factors employed in the production. The following are a few of the examples of homogeneous functions of degree one:

(23) $$x = a u^m v^{1-m} \pm b u^s v^{1-s} \qquad 1 \geq m \geq 0, \quad 1 \geq s \geq 0$$

1) For this and other properties of a homogeneous function see R.G.D. Allen op.cit. pp. 317-18.

2) The total output can always be imputed back to factors under long term competitive equilibrium, whether the production function is homogeneous of degree one or not, cf. P.A.Samuelsons Foundations of Economic Analysis 1947, p. 85-86.

(24) $$x = \frac{u^m v^m}{au^{m-1} + b v^{m-1}}$$; m any number

(25) $$x = (a u^m + b u^{m-1} v + \dots g u v^{m-1} + h v^m)^{\frac{1}{m}}$$; m any number.

The equations from (23) to (25) which represent a wide variety of homogeneous production functions of degree one and all other functions of this type can be expressed in the following form.[1)]

$$x = uf(\frac{v}{u}) .$$

Some well-known and commonly used production functions will be discussed later.

VI

We have touched upon varying returns to factors and varying returns to scale, but we have not studied them in juxtaposition. This is needed for two reasons. Firstly because of the similarity of the two expressions; especially excluding the last terms which are often omitted in discussions, the beginners are not infrequently confused about their exact meaning. It ought to be stated, however, that the former is also called the "law of variable proportions" and if this term were consistently used in the literature and discussions, the confusion might have been lesser. The other reason is that recently an interesting discussion has taken place as to whether an increasing returns to a factor is compatible with a constant returns to scale.[2)]

The law of varying returns to a factor states that as the input of a certain factor is increased, the inputs of other factors remaining constant, the resulting additions to output may increase, remain constant or decline according to whether the factor is experiencing increasing, constant or

---

1) Let $x = f(u, v)$ be a homogeneous function of degree one, then $x = u\, f(1, \frac{v}{u}) = u\, f(\frac{v}{u})$.

2) cf. American Economic Review, September 1964.

diminishing returns at its successive values and the given level of other factors. In other words, if the marginal productivity of a factor increases, remains constant or diminishes, they show increasing, constant or diminishing returns to that factor, in that interval, it being understood that the marginal productivity of a factor is derived on the assumption that the amounts of other factors are kept fixed. It is evident that as long as the marginal product of a factor remains positive, the total product will increase but at an increasing, constant or diminishing rate according as the law of increasing, constant or the diminishing returns are applicable to that factor in the region under consideration given the level of other factors. In our symbols, as we have already stated, $\frac{d^2x}{du^2}$ or $\frac{d^2x}{dv^2}$ $\gtreqless$ 0 according as the factor u or v is experiencing an increasing, constant or diminishing returns.[1)]

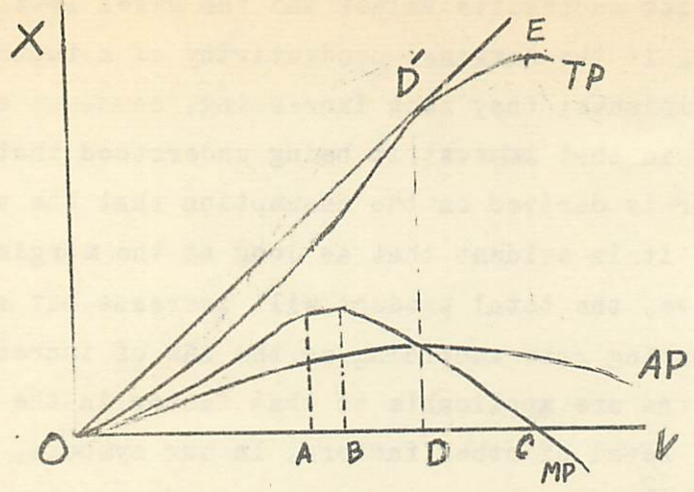
Returns to scale, on the other hand, describes the behaviour of the total product when all the factors are increased in the same proportion. As stated earlier, the output increases in a proportion more than, equal to or less than that in which the factors increase according as the production function depicts increasing, constant or diminishing returns to scale.

Returns to a factor can be illustrated with the help of the following diagram. In Fig.5, the total product curve per unit of u is plotted, the product per unit of u taken vertically and v per unit of u horizontally. In this illustration u is fixed at a level, say u_1, and v is the variable factor. Fig.5 shows the total curve, TP, average product curve, AP, and the marginal product curve, MP.

---

1) The law of variable proportions or the law of returns to factors has been defined by several text-book writers by relating the changes in one factor, other factors being held constant, to the movement in the total product; cf. R.H. Leftwich, The Price System and Resource Allocation 1960, p. 109. It (the law of diminishing returns) states that if input of one resource is increased by equal increments per unit of time, while the inputs of other resources are held constant, total product output will increase, but beyond some point the resulting output increases will become smaller and smaller.

Fig.5.



According to our definition increasing returns to factor v occur till point A, constant returns till point B, and diminishing returns after that till point C. After point C, no sensible producer would like to apply more of v, given the fixed level of u at u_1; for it will not lead to any addition to the total product. TP is concave upward for earlier units of v, for the given units of u at u_1 because small amounts of inputs are too inadequate and that output rises at an ever faster rate as the units of v are increased, and thus till point A, increasing returns to v are obtained. Then from A to B the output increases at a constant rate characterized by constant returns to the variable factor v. However, the TP keeps rising but at a diminishing rate till point D', where the AP attains its maximum value. This point can be easily spotted by drawing a tangent to TP passing through the origin. For on no point other than the point of tangency, i.e. D' will the slope of the line joining a point on TP to the origin which gives the average product, will be larger. After D, TP increases further, but at an ever slower rate till it reaches its maximum point at E where the MP becomes zero. Beyond this point no producer would like to go because any addition of the variable factor v will not only not bring about any additions of total output, but on the contrary will lead to the reductions in the total output.

Now we can return to the interesting discussion whether increasing returns to a factor are possible under conditions of constant returns to scale. We know that constant returns to scale are obtained when the production function is homogeneous of degree one. According to Euler's theorem then

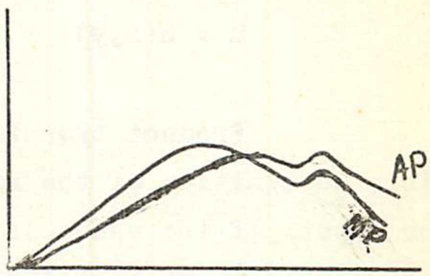
$$x = \frac{dx}{du} u + \frac{dx}{dv} v$$

(26) or $$\frac{x}{v} - \frac{dx}{dv} = \frac{dx}{du} \cdot \frac{u}{v}$$

As x is output and u and v are inputs, they are all non-negative and so $\frac{x}{v}$ and $\frac{u}{v}$ are non-negative. From (26) it is obvious that whenever $\frac{dx}{dv}$ is negative $\frac{dx}{du}$ is positive and $\frac{dx}{du}$ is negative only when $\frac{dx}{dv}$ is positive and greater than $\frac{x}{v}$. This means that whenever the marginal product of v is greater than its average product, the marginal product of u must be negative. The excess of marginal product of v over average product implies that the marginal product of v has been increasing over part of its course, i.e. it has been experiencing an increasing returns to itself.[1] Thus it can be shown that increasing returns to a factor are not incompatible with constant returns to scale, or more correctly, with linear homogeneous functions. This seems to contradict the assertion of some text-books that only diminishing returns to individual factors are compatible with constant returns to scale.[2] However, if we add the restriction that the marginal productivities of the factors do not become negative, then we can immediately see from (26) that $\frac{x}{v}$ must be greater than $\frac{dx}{dv}$ all through hence $\frac{dx}{dv}$ must be positive but diminishing from the very beginning. By changing places we can say the same for the other factor.

---

1) In some cases the AP curve may be above the MP curve even if the latter is rising. This may happen when MP curve rises after falling a certain distance, but not enough to cross the AP curve and then falls again, as shown in the opposite figure.



2) A. Stonier and D. Hague, A Text-Book of Economic Theory, New York 1961; p. 229. Erich Schneider, Einführung in die Wirtschaftstheorie, Vol.II, Tübingen 1958, p. 197.

A definite conclusion, however, depends on how far the above restriction is legitimate. Intuitively it seems that the restriction is in order.[1)] Further we have to be on the guard that the range within increasing returns to a factor are obtained, the total output is not negative or is not falling.

VII

So far we have considered that several factors or, as in our illustrative cases, two factors combine together to produce one output. Some production processes are such that one or more factors combine to produce more than one output. Examples of joint production where two or more outputs are produced in variable proportions are wool and mutton, grain and chaff etc. A process is characterized by joint production only when the quantities of two or more outputs are technically interdependent as in the two examples mentioned here. A firm may be producing two or more outputs, but if the factors going into the production of each one of these outputs are independent of the other, then the situation will not be one of joint production.

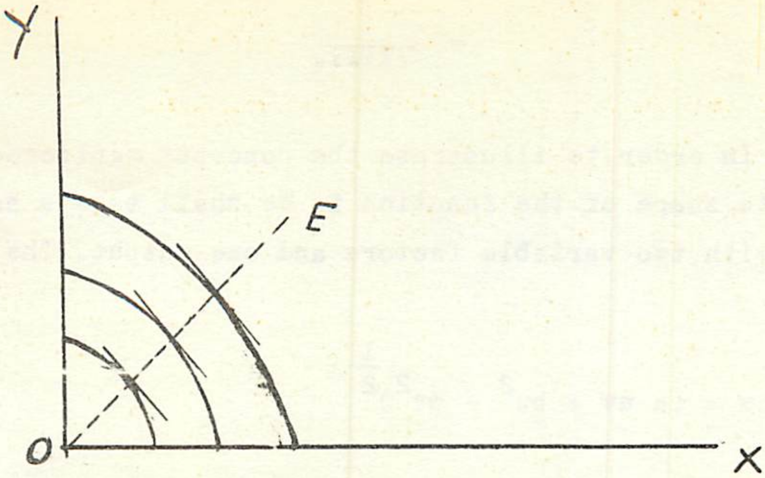
In order to maintain the simplicity of the exposition we shall consider the case of two commodities X and Y and one factor U. The production function can now be expressed as,

(27) $u = h(x, y).$

Product transformation curve joins the points giving the minimum possible combination of the amounts of the outputs with a given amount of factor input. If the value of u in (27) is fixed at u^o, we get a product transformation curve associated with that level of output; for higher values of u, we shall get higher transformation curves and vice versa as shown in Fig.6.

1) For various comments and discussions on this problem see American Economic Review, September 1964.

Fig.6.



The slope of the tangent at a point on a product transformation curve gives the rate at which more of one commodity can be obtained in place of the other, given the amount of factors or resources. The negative slope of a transformation curve is defined as the marginal rate of (product) transformation or MRT:

(28) $$MRT = -\frac{dx}{dy}$$

From (28) we can derive

(29) $$MRT = -\frac{dx}{dy} = \frac{h_y}{h_x}$$

(29) states that the marginal rate of transformation of x to y is the ratio of marginal factor resource requirement of y to that of x. Product transformation curves are concave to the origin, i.e.

$$\frac{dx}{dy} > 0$$

$$\frac{d^2x}{dy^2} < 0$$

If, on successive transformation curves, we join points where the marginal rates of product transformation are equal to a certain constant, we get what may be termed again as the expansion path under the condition that the two commodities are produced in such a proportion throughout that their MRT's are equal to the constant assumed. In Fig.6 OE shows such a path.

VIII.

In order to illustrate the concepts mentioned above, we can take a concrete shape of the function f. We shall take a homogeneous function of degree one with two variable factors and one output. The form of the function is

$$(30) \qquad x = (a\,uv + bu^2 + cv^2)^{\frac{1}{2}}$$

where a, b and c are constant technical coefficients. In order to find the variation of output X with variations in one factor, say U with V fixed at a certain level, say v_1, we just substitute v_1 for v in (30) to get

$$(31) \qquad x_1 = (a\,uv_1 + bu^2 + cv_1^2)^{\frac{1}{2}}$$

(31) gives the section of production surface formed by plotting the right hand side of (30) by a plane perpendicular to OV at v_1. Similarly we can get a function giving the variation of x with V, for a certain fixed value of U. Differentiating (31) with respect to u, we get the marginal productivity of u, so that

$$(32) \qquad \frac{dx_1}{du} = \frac{1}{2}(a\,uv_1 + bu^2 + cv_1^2)^{-\frac{1}{2}}(av_1 + 2bu)$$

Similarly we can get the marginal productivity of v. The production isoquant representing the level of output, say x°, can be obtained just by inserting x° for x in (30) so that

$$(33) \qquad x_o = (a\,uv + bu^2 + cv^2)^{\frac{1}{2}}$$

$$\text{or} \qquad u = \frac{av + \left((a^2 - 4cb)v^2 - ub(x^\circ)^2\right)^{\frac{1}{2}}}{2b}$$

The above gives one of the sets of isoquants associated with the output level x°. The whole set of isoquants is derived easily by putting different values of x in place of x°. The whole system of the isoquants is thus given by

(34) $$u = \frac{av \sqrt{(a^2 - 4bc)v^2 - ubx^{\frac{1}{2}}}}{2b}$$

The elasticity of production of x with respect to u is, substituting from (32) in (6)

(35) $$e = \frac{u}{x} \cdot \frac{dx}{du} = \frac{u}{x}\frac{1}{2}(av_1 + 2bu)(auv_1 + bu^2 + cv_1^2)^{-\frac{1}{2}}$$

$$= \frac{\frac{1}{2}u(av_1 + 2bu)}{(auv_1+bu^2+cv_1^2)^{\frac{3}{2}}}$$

and similarly we can get the elasticity of production of x with respect to v.

The marginal rate of substitution of U and V at a point (u,v) along an isoquant, say (31), is

(36) $$r = -\frac{du}{dv} = -\frac{(au + 2cv)}{(av + 2bu)}$$

An isocline equation can be arrived at quite simply by equating r with the constant substitution rate desired. Let the latter be k', so that the isocline which passes through the successive isoquants at points where their slopes or MRS are equal to k' is given by

(37) $$r = -\frac{du}{dv} = -\frac{au + 2cv}{av + 2bu} = k'$$

or $$u = -\frac{2c + ak'}{a + 2bk'} v$$

For different values of k' we can get a set of isoclines or expansion paths. It is interesting to note that when the production function is homogeneous of degree one, as the example taken up by us is, the formula for, σ, the elasticity of substitution is greatly simplified. For according to Euler's theorem we have,

$$x = uf_u + vf_v$$

Differentiating the above with respect to u and v respectively, we get for a given value of x,

$$f_{uu} = -\frac{v}{u} f_{uv} \qquad \text{and} \qquad f_{vv} = -\frac{u}{v} f_{uv}$$

Substituting the above in (14) we have

$$(14') \qquad \sigma = \frac{f_u f_v}{x f_{uv}} \qquad \text{or} \qquad \frac{\frac{dx}{du} \cdot \frac{dx}{dv}}{x \frac{d^2x}{dudv}}$$

Therefore the elasticity of substitution of our illustrative function is

$$\frac{(av + 2bu)(au + 2cv)}{-(av + 2bu)(au + 2cv) + 2ax^2} .$$

References:

1) J.M. Henderson and R.E. Quandt, Micro-economic Theory (McGraw-Hill) 1958, Chap. 3.
2) R.G.D. Allen , Mathematical Analysis for Economists (MacMillan) 1960, Chap. XIII.
3) S. Carlson , A Study on the Theory of Production (Kelly and Millan) 1956.
4) E. Schneider , Pricing and Equilibrium (MacMillan) 1962, Chap. II.
5) R.H. Leftwich, The Price System and Resource Allocation (Holt-Rivehart) 1960, Chap. VII and VIII.
6) P.A. Samuelson, Foundations of Economic Analysis (Harvard) 1948, Chap. 4.

oo O oo

(A.H.)

Lecture 2

MICRO-COST FUNCTION

I.

In the preceding lecture we expressed the factors or the inputs in a way as if their quantities were variable. This may be true if we consider the part of the services of the factors that are actually utilized in the process of production. Even then, there are conceptual difficulties involved, since there are certain fixed factors which can yield their services at a certain time up to a particular limit, no more, while under most circumstances, if the services which these factors are able to provide currently are not fully used, there is no guarantee that the currently available but unutilized services can be stored for future utilization. The size of these fixed factors have to be decided with a consideration for long term schedule of production (i.e.) taking into view the maximum possible amount of output that is desired to be produced during the lifetime of the fixed factors installed. Thus the fixed factors are variable in the very long run, i.e. in a period exceeding the life-time of the currently installed fixed factors. But in the short run they are fixed and do not vary with changes in output. The life of these factors is mostly and generally independent of the amount of output produced at a particular time, so long as the level of output is below the upper limit that the fixed factor inputs can produce.

For a firm, thus all costs incurred on account of fixed factors are invariable and fixed, it is the cost of other factors and inputs that vary in relation to output.

The entrepreneur's total cost of production can be given by a linear equation

(1) $$C = p_u u + p_v v + b$$

where C is total cost, p_u is the price of variable factor U and p_v the price of variable factor V, and b is cost of the fixed factors or inputs. It should be noted that b is not the total cost of the fixed inputs, but only the cost of the fixed input that can be imputed to the current volume of production. In theory, alternative ways are possible to evaluate the fixed cost and in practice, numerous methods are adopted by individual firms to earmark the fixed cost. Broad considerations that matter in the calculation of the fixed costs are the total purchase cost of the fixed inputs, the life of the inputs, the average interest charges on the cost of the inputs over their life span, risk of absolescence and the probable appreciation of the value of these inputs. Generally fixed cost pertaining to output of a production period should approximate total purchase cost plus total interest cost divided by the number of periods over which the fixed inputs last in working condition with allowance for absolescence and appreciation.

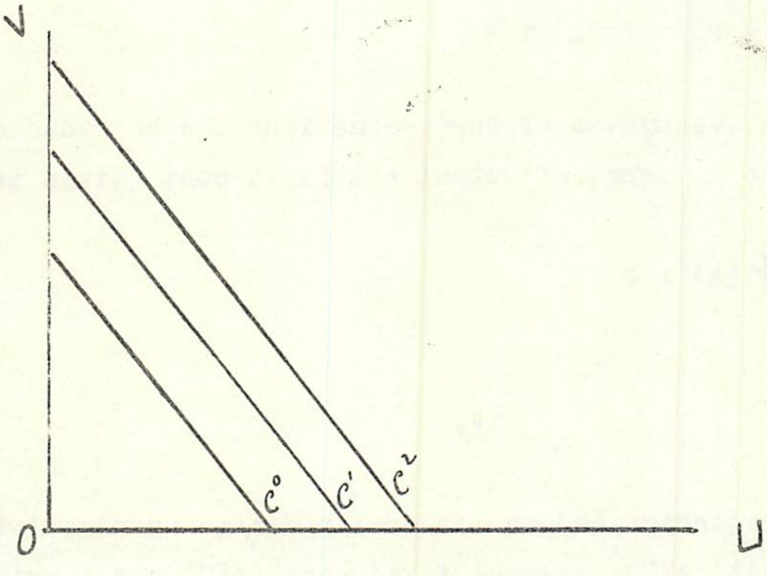
There are certain items of expenditure in the nature of fixed inputs such as training and research departments, recreation, housing and other facilities for workers and their children, construction of roads, rails and bridges etc. which bring benefits not only to the firm undertaking these expenditures, but to other firms in the locality and also in the country, not only in the current period but even in the distant future. The benefits of these expenditures are so diffusedover area, among firms of the same and other types and in future that it is very difficult if not impossible for a firm to make even a rough estimate of the cost of these items to be included in the cost function of the current production of the firm. In actual costing, the firms mostly follow some customary methods giving only broad consideration to the items mentioned above. It seems, however, that these items may cause a little less difficulty when we set to constructing macro production function for the whole country.

II.

Returning to equation (1), we can derive what is termed as isocost line. This is defined as the locus of the points giving various combination of inputs that may be purchased for a specified total cost. Let the specified total cost be C^o, then the corresponding isocost line given by (2) is a straight line in u, and v, given their prices, and the value of b. By specifying different values for C we can get a family of isocost lines as shown in Fig.1 $C^o = p_u u + p_v v + b$

(2) , given C^o is price p_u and p_v and b . By specifying different ... as shown in Fig.1.

Fig.1.



We can write equation (2) in the following form:

$$u = \frac{C^o - b}{p_u} - \frac{p_v}{p_u} v . \tag{3}$$

Differentiating (3) with respect to v, we can immediately see that the slope of the isocost lines is the reciprocal of the ratios of their corresponding prices with a negative sign. In Fig.1, we can get the intercept of the isocost lines with the U or V axis easily by putting v or u respectively equal to zero in (3). The greater the value of C^o, the greater the intercepts on the axes, and the higher the isocost lines.

III.

In lecture 1 we described physical production function, and we also derived what we call the isocline or the expansion path. We have just introduced the cost equation. From these three sets of equation we can derive the cost function. These three sets of information are sufficient for deriving a general cost function. The main idea underlying a cost function is that the entrepreneurs are assumed to economize as much as possible, given the prices of inputs and the output they produce. If they do so, it can be easily shown that they will move on the optimal expansion path.[1] The three equations required for the deriviation of the cost function are:

$$x = f(u,v)$$
$$0 = g(u,v)$$
$$C = (u\,p_u + v\,p_v) + b$$

The above system of three equations can be reduced to an equation expressing cost in terms of output and fixed cost (given the input prices)

(4) $$C = \psi(x) + b\ .$$

IV.

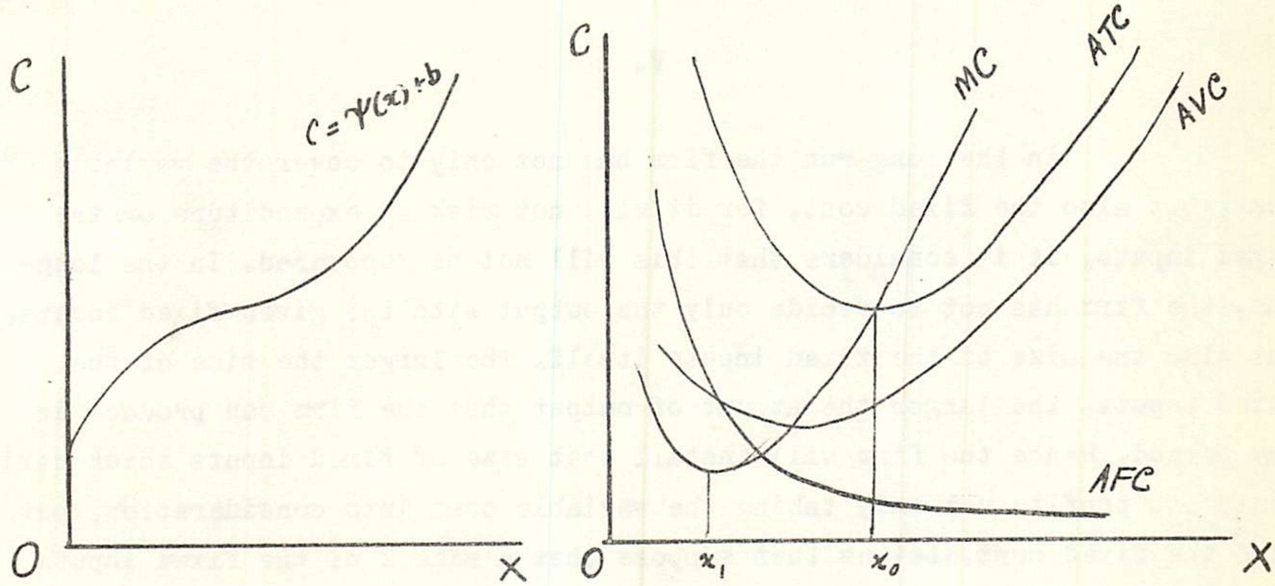
From equation (4) we can easily derive average total cost (ATC), average variable cost (AVC), average fixed cost (AFC) and marginal cost (MC).

(5) $$\begin{cases} ATC = \dfrac{\psi(x) + b}{x} \\ AVC = \dfrac{\psi(x)}{x} \\ AFC = \dfrac{b}{x} \\ MC = \psi'(x) \end{cases}$$

1) Cost is minimized when $\frac{dC}{dx} = \frac{d(p_u u + p_v v)}{dx}$, or $-\frac{du}{dv} = \frac{p_v}{p_u} = k'$.

Cost functions may assume various forms. Fig.2 gives the shapes of a cost function and other related functions mentioned above which are normally assumed by economists.

Fig.2.



Total cost curve is a cubic function, but ATC, AVC, and MC are second degree parabolic functions. MC reaches its minimum before AVC and AVC before ATC. AFC is a rectangualr hyperbola which can be easily seen from the third equation above. MC passes through the minimum points of both ATC and AVC curves. Differentiating ATC with respect to x, and equating to zero we have

$$(6)\quad \begin{cases} \dfrac{x\psi'(x) - \{\psi(x) + b\}}{x^2} = 0 \\[2ex] \dfrac{\psi(x) + b}{x} = \psi'(x) \end{cases}$$

which shows that the extreme lowest point on ATC (and similarly on AVC) curve is reached where ATC equals MC.

In order to show that in the short run the output of firm is affected by only the variable cost and not the fixed cost, we have to analyze its main motive, i.e. profit maximization. Given the price of the output, the profit of the firm, $p_x x - \psi(x) - b$, will be maximized if

$$(7)\qquad p_x = \psi'(x).$$

(7) implies that the entrepreneur has to equate his marginal cost to the price of the output to maximize his profit. As the marginal cost depends on the variable cost and not on the fixed cost. The entrepreneur may go on producing a certain output in the short run, even if he does not cover all or any part of the fixed cost so long as he covers the variable cost in full.

V.

In the long-run the firm has not only to cover the variable cost, but also the fixed cost, for it will not risk an expenditure or the fixed inputs, if it considers that this will not be recovered. In the long-run, the firm has not to decide only the output with the given fixed inputs, but also the size of the fixed inputs itself. The larger the size of the fixed inputs, the larger the amount of output that the firm can produce in any period. Hence the firm will install that size of fixed inputs which maximizes its profits not only taking the variable cost into consideration, but also the fixed cost. Let us then suppose that a size k of the fixed inputs will enable the firm to maximize its profits. Then the production function, the cost function and the expansion path will be

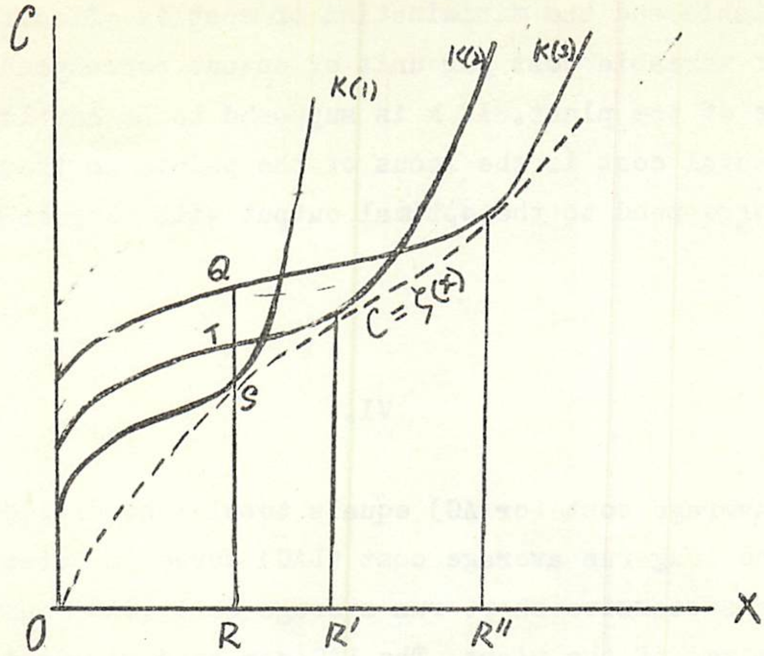
$$x = f(u, v, k)$$
$$C = p_u u + p_v v + \phi(k)$$
$$0 = g(u, v, k)$$

In the middle equation above, the fixed cost is an increasing or linear function of k, as that $\phi'(k) > 0$ or $\phi(k) = p_k \cdot k$ where $\phi'(k) = p_k$, the price per unit of fixed inputs. Eliminating u and v from above we can derive

$$C = F(x, k) + \phi(k) \qquad (8)$$

(8) describes the family of total cost curves, for different values of k. For each size of k we get a cost function, but this function will be relevant for short term analysis. In the long run we require a cost function which minimizes the cost of production of the firm whatever the level of output and the corresponding size of k. For example, as shown in Fig.3, the output OR

Fig. 3.



can be produced in any of the plants with fixed inputs of sizes K(1), k(2) and k(3). But his cost will be RS in the plant with fixed input of size k(1), RT with k(2), RQ with k(3). The plant with k(1) size of fixed input gives the minimum cost, amd OR" amount of output can be produced at the minimum cost by a plant with k(3) size of fixed inputs. As we have to find the function which produces any level of output at the minimum cost, we differentiate (8) with respect to k and equate it to zero, so that

$$(9) \qquad \frac{dC}{dk} = \frac{F(q,k)}{k} + \emptyset'(k) = 0$$

Eliminating k from (8) and (9) we get

$$(10) \qquad C = \zeta(x)$$

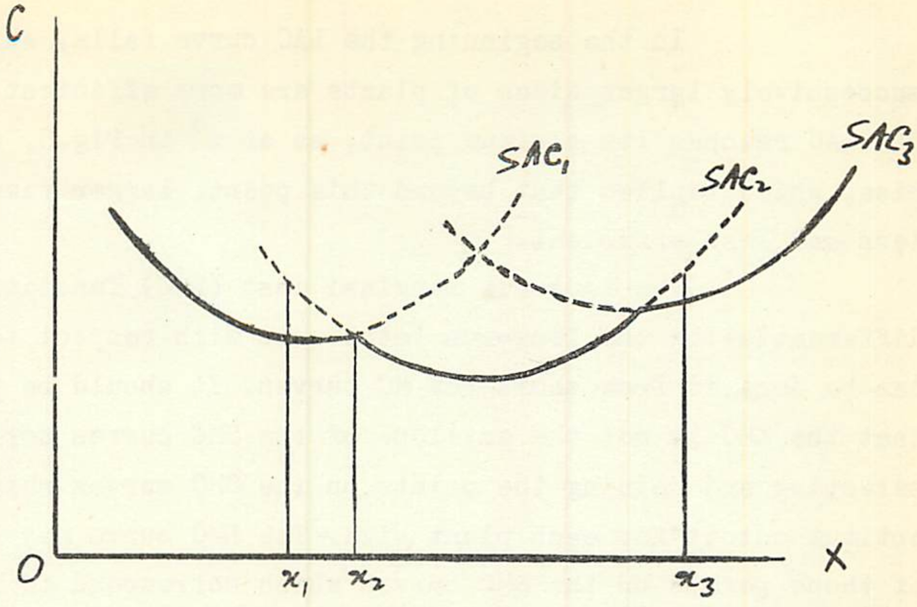
as the long-run cost function. Thus the long run cost function is the locus of the lowest points on short-run cost functions corresponding to different sizes of fixed inputs at each level of output. The long-run cost curve is the envelope of the short-run curves. It touches each but it intersects none. The main difference between the two is that in the case of the short-run cost curves the size of the plant or the magnitude of the fixed inputs is kept constant and the minimization of the cost is related to the variable inputs, but in case of the long-run cost functions, the size of the plant itself is treated as variable and the minimization of cost is effected through the minimization of variable cost per unit of output corresponding to each successive size of the plant. If k is supposed to be continuously variable, then long-run total cost is the locus of the points on the short run cost curves which correspond to the optimal output with respect to each size of the plant.

VI.

Average cost (or AC) equals total cost divided by output level. The shape of the long-run average cost (LAC) curve is determined by the position of the successive short-run average cost (SAC) curves corresponding to respective sizes of the plant. The LAC can be derived either by dividing the total cost at each level of output by the level of output or by correstructing an envelope of the short-run average cost curves. The LAC shows the movement of the average cost when the size of plant varies in such a manner as to yield the lowest possible cost for that output.

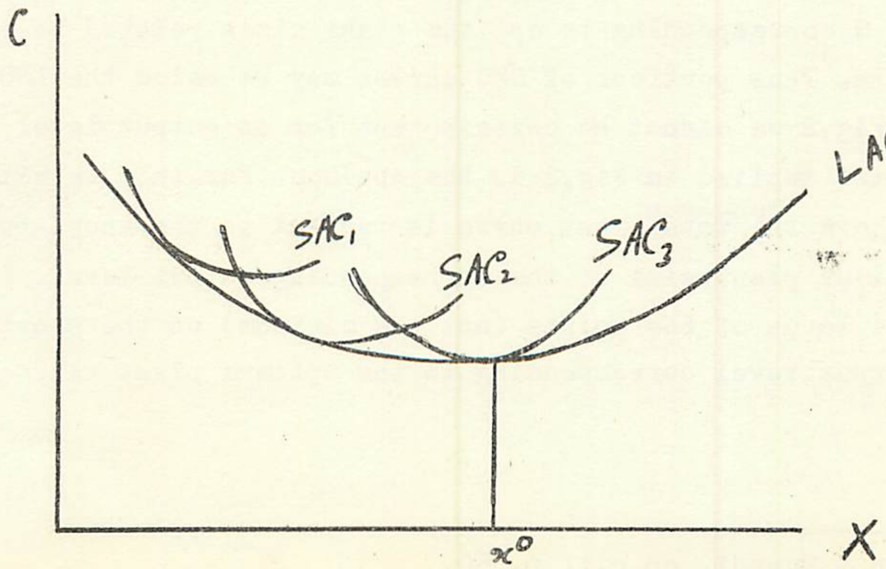
This can be shown by the following figure. In Fig.4 three short run average cost curves SAC_1, SAC_2, SAC_3 have been drawn corresponding to three plant sizes. In the short run the firm has only one of the plant sizes, but in the long-run it can build a plant of any size and can shift from one to the other. The firm will always want to produce an output at an average cost as low as possible. If the firm has to produce an output equal to x_1, it will build a plant of the size represented by SAC_1 rather than SAC_2 (or SAC_3) for in the second case the average cost per unit will be higher x_1B instead of x_1A. If the amount of output to be produced is x_2, then the plant sizes represented by SAC_1 and SAC_2 are equally efficient, for the cost in both the cases are the same. For an output equal to x_3, the firm will, for the reasons stated, will want to shift to a plant of the size represented by SAC_3. If only three sizes of the plant are possible to be built, then the LAC curve will be given by the line formed by tracing the lowest parts of all the successive SAC curves representing the different sizes of the plants. This is shown by the thick line in the diagram, the broken lines are irrelevant for the LAC.

Fig.4.



In reality, in the long run, a firm can build a plant of any size it wants, in fact, the number of possible scales is infinite. A series of infinite number SAC curves can be drawn as shown in Fig.5. The long run average cost, as before, can be obtained by tracing outer parts of the successive SAC curves. When the number of SAC curves is infinitely large, the LAC curve touches at each point a SAC curve, which happens to be lowest in relation to the output corresponding to that point. This is the same as saying that the long run average cost is an 'envelope curve' to the SAC curves.[1)]

Fig.5.



1) Leftwich, op.cit. p. 155.

In the beginning the LAC curve falls, which implies that successively larger sizes of plants are more efficient than the smaller ones. The LAC reaches its minimum point, as at x^o in Fig.5, and then it begins to rise, which implies that beyond this point, larger sizes of plants become less and less efficient.

The long-run marginal cost (LMC) function can be derived by differentiating the long-run total cost with respect to output level, or it can be derived from short-run MC curves. It should be pointed out, however, that the LMC is not the envelope of the SMC curves nor can it be derived by selecting and joining the points on the SMC curves which correspond to the optimum output for each plant size. The LMC curve may be defined as the locus of those points on the SMC curves which correspond to the optimum plant size for each output.[1] As can be seen from Fig.3 the long-run total cost curve is tangent to each short-run total cost curve at the output for which the latter represents the optimum plant size.

This can be explained by looking at the lowest points of MC and ATC curves in Fig.2. Obviously the curves in Fig.2 refer to a given plant size. This plant size is optimal for an output level equal to Ox_o, but the MC is minimum at output level Ox_1. This shows that the minimum point on MC curve related to the optimum plant size corresponding to a certain output level is not relevant to the LMC. The LMC curve will, thus, be locus of points like H rather than G corresponding to optimum plant sizes related to successive output levels. Thus portions of SMC curves may be below the LMC curves. However, from Fig.2 we cannot be certain that for an output level equal to Ox_o the plant size implied in Fig.2 is the optimum. For this we will have to turn to Fig.3, where the long-run total cost curve is tangent to the short-run curve representing optimum plant size at the corresponding output level. Thus the LMC curve is the locus of the points (not the minimum) on the short-run MC curves for each output level corresponding to the optimum plant size and <u>not vice versa</u>.

---

1) Henderson & Quandt, op.cit. p. 60.

VII.

In Fig.5 the long-run AC curve move downwards reach its minimum point and then rises up. The descending phase of the cúrve is characterized by internal economies of the producing unit resulting from larger scales of production. The well-known economies of large-scale production are due to advantages derived from the reduction in cost per unit of output by employing bigger but more efficient machines and equipments, the possibility of effecting greater specialization or division of work of men and machines, greater facilities and better terms in purchases and sales of larger inputs and outputs, utilization of by-products, possibilities of production of accessories and raw-material and other inputs, availability of easier and better financial arrangements, opportunities of establishing research departments, transport sections and legal wings etc. These advantages may not be available to a unit producing at a small scale, besause many of these items are indivisible and they will not pay their way if they are combined with a small scale production. It may be noted that all the advantages mentioned above are connected directly or indirectly with fixed inputs. Hence it is seen that the larger the size of plant upto a limit, the greater the economies of scale and the lower the average cost per unit of output.

The rising phase of the LAC curve is explained by the internal diseconomies which begin to be more effective after the lowest point has been reached and grow in effectiveness as the scale of production increases. The diseconomies, in turn, arise due to the increasing complexity of problems and lack of coordination. The main reason for the diseconomies is the fact that management is not perfectly variable. It is true that up to a certain extent the management group can be increased and specialization and division of labour can be introduced in the management as such, but beyond that problems of coordination become increasingly difficult, giving way to increasing bureaucracy and red tape.

Apart from internal economies and diseconomies a firm's total cost and so all other costs are affected by the output level and productive activities of other firms. External economies are realized, if the productive activity of any other firm or firms lower the cost of the firm and diseconomies are realized if it raises the cost. It is possible that apart from internal economies and diseconomies, the average cost curve of the firm descends in the early phases due to the external economies realized from the entry of other firms and diseconomies are realized in the later stages because either too many new firms have entered the field or fresh entry of new firms has stopped.

VIII.

We can illustrate the concepts described above by taking a concrete production function as we did in the preceding lecture.

(11) $$x = (auv + bu^2 + cv^2)^{1/2}$$

where a, b, and c are constants as before. Total cost function is

(12) $$C = p_u u + p_v v + b.$$

The isocline for a certain constant, k' is as given in (23) of the preceding lecture,

(13) $$u = \frac{2c + ak'}{a + 2bk'} v$$

Finding the values of u and v in terms of x from (10) and (13) and ignoring the negative values, we have

$$u = \theta x(a\theta + b\theta^2 + c)^{-1/2}$$
$$v = \theta (a\theta + b\theta^2 + c)^{-1/2}$$

where $\theta = \frac{2c + ak'}{a + 2bk'}$

Putting the values of u and v in (12), we get the short-run cost function

(14) $$C = p_u \theta x(a\theta + b\theta^2 + c)^{-1/2} + p_v x(a\theta + b\theta^2 + c)^{-1/2} + b'.$$

It is trifling to derive ATC and AFC; the marginal cost function, MC, can be derived by differentiating C with respect to x, so that

(15) $$MC = p_u \theta(a\theta + b\theta^2 + c)^{-1/2} + p_v(a\theta + b\theta^2 + c)^{-1/2}$$

which, in this case, is constant as θ, p_u and p_v are constant.

For the derivation of the long-run cost function, we shall have to introduce a production function which contains the size of the plant k as one of the variables. One simple form may be the following:

$$x = (akuv + bu^2 + cv^2)^{1/2} \tag{16}$$

The cost function will now be

$$C = p_u u + p_v v + p_k k \tag{17}$$

where p_k is the price of k per unit; we shall assume that the cost of fixed inputs varies in proportion to the size of the plant and is equal to the size multiplied by its (imputed?) price. The expansion path for a constant k' is the same as (12).

From (12) and (16) we can get the value of u and v in terms of x and k,

$$u = \theta x(ak\theta + b\theta^2 + c)^{-1/2}$$
$$v = x(ak\theta + b\theta^2 + c)^{-1/2}$$

where θ has the same value as the above. Putting the values of u and v in (17), we get the total cost function in terms of x and k, i.e. the output level and the plant size, so that

$$C = (p_v x + p_u \theta x)\ (a\theta k + b\theta^2 + c)^{-1/2} + p_k k. \tag{18}$$

In order to find the long-run cost function, we have to minimize C with respect to k. Differentiating C with respect to k for given values of x, and equating to zero, we have

$$\frac{dc}{dk} = -\tfrac{1}{2}(p_v x + p_u \theta x)\ (a\theta k + b\theta^2 + c)^{-\frac{3}{2}}\ a\theta + p_k = 0$$

$$\therefore \quad k = \frac{1}{a\theta}\left\{(2p_k)^{-\frac{2}{3}}\left[(xa\theta\ (p_v + \theta pu)\right]^{\frac{2}{3}} - (b\theta^2 + c)\right\} \tag{19}$$

Substituting (19) in (18) we get the long-run cost function. LAC function can be derived by dividing the long-run cost function by x and the LMC function by differentiating it with respect to x.

Reference:

J.M. Hendersen & R.E. Quandt , Micro-Economic Theory (McGraw-Hill) 1958, Chap.3.

R.H. Leftwich , The Price-System and Resource Allocation (Holt-Rivehart) 1960, Chaps. VII and VIII.

S. Carlson , A Study on the Theory of Production (Kelly and Millan), 1956.

E.O. Heady & J.L. Dillon , Agricultural Production Functions, (Iowa). 1961. Chap. 2.

oo 0 oo

(A.H.)

Lecture 3.

Aggregation of Micro-Production Function.

In the last two sections we have described production and cost functions in a way as if there were perfectly exact relationship between the quantities of inputs or factors and the magnitudes of output and cost. Unfortunately, this is not so and for several reasons: (1) qualities of inputs or factors are not perfectly uniform and the resulting output does not correspond exactly to the amounts of factors and also the quality of the latter is not uniform; (2) in a number of cases it is not possible even to determine the uniformity of the quality of factors, nor the commensurability of their different quantities, and in some cases it is not possible even to measure their amounts, (one has just to consider the skill, managerial and entrepreneurial ability of different grades of workers, to be convinced of these facts); (3) a number of inputs and factors are not divisible and quite frequently it happens that one or more of the factors are not fully utilized in the process of production, leading to absence of exact correspondence between the amounts of factors and the resulting output; (4) even if we assume that there exists an exact relationship between the amounts of factors and the consequent output, there may occur certain accidents or unforeseen developments which may disturb this relationship, e.g., vagaries of weather in agricultural production or breakdown in industrial production etc.; (5) it is not always possible to include all the elements that contribute to production of a producing unit. (For example, the climate of certain places is particularly suitable for the production of a certain good, but it is simply impossible to quantify climate and include it in the production function). Hence the production function of an individual producing unit i, can more properly be written as

$$x_i = f^i(u_i, v_i) + e_i \qquad (1)$$

where x_i is the quantity of the commodity produced by the i_{th} producer and u_i, v_i, are the amounts of factors used, e_i is the residual quantity that represents the disturbance due to the causes mentioned above.

For an individual producer, it is the individual micro production function (or the micro cost function) which is of main concern. An economist or social scientist, on the contrary, is interested, more often, in a production function at the macro-level, such as the total production of an economy or an industry with aggregate amounts of factors of production. Obviously these macro-production functions can be derived only through and on the basis of the micro-production functions by applying a systematic procedure of aggregation. As Dr. Bentzel[1] has pointed out, the problem of aggregation in production theory involves aggregation over goods, over factors of production and over firms, a triple aggregation. Further the problem of aggregation which arises in equilibrium analysis should be distinguished from that which arises in dynamic analysis. The following are the conclusions at which Bentzel arrives.

In the analysis of the behaviour of a single firm one cannot, in the general case, lump together two or more factors into a group and apply the propositions of production theory to this group. This procedure is possible, however, in the case where the prices of the factors concerned vary proportionately. Similarly in a firm which produces several products an aggregation of products is not generally permissible. But in the special case where the prices of two or more products vary proportionately, these products can always be aggregated. And with proportional price variations, simultaneous aggregation of goods, factors and firms is always possible; each type of aggregation is possible by itself and they can be performed successively without conflicting with each other. We shall discuss below the simplest possible case of aggregation over firms.

The problem of aggregation is mainly statistical in nature. We can, at best, give an outline of this procedure based on one of the elementary cases treated by Prof. Theil.[2]

---

1) Ragnar Bentzel, "On the Aggregation of Production Functions", International Economic Papers, No. 8.

2) H. Theil, Linear Aggregation of Economic Relations, Amsterdam 1954. Other references on this problem are L.R. Klein, "Macro-economics and the Theory of Rational Behaviour", Econometrica, 1946; K. May, "The Aggregation Problem got a One-Industry Model", ibid; and A. Nataf, "Sur la Possibilite de Construction de certains Macromodeles", Econometrica, 1950.

The simplest case arises when we assume the production function to be of the following linear form

$$(2) \qquad x_i = \alpha_i + \beta_i u_i + \gamma_i v_i + e_i \qquad i = 1, 2, \ldots I.$$

In (2) α_i, β_i and γ_i are microparameters and the disturbance e_i is supposed to have zero means. Secondly, all macro-variables are assumed to be sums of the corresponding macro-variables:

$$x = \sum_{i=1}^{I} x_i$$

$$u = \sum_{i=1}^{I} u_i$$

$$v = \sum_{i=1}^{I} v_i$$

This assumption can be made more general by taking the macro-variables as not simple sums of the micro-variables, but weighted averages of the micro-variables. This may be preferable in cases when there are considerable differences between the sizes of individual producing units. For the sake of brevity and with a view to avoiding too much use of symbols, we shall not introduse weights here. Their introduction, however, does not raise any logical difficulty.

Thirdly, it is postulated that macro-variables satisfy a relation of the similar form as (1), so that

$$(4) \qquad x = \alpha + \beta u + \gamma v + e$$

The macro-variables α, β, and γ can be obtained by applying the classical least square method to the values of the macro-variables for a time series.

The main problem is to find the relationships between the macro-parameters α, β, and γ, the corresponding micro-parameters. In other words, to do this, we consider the micro-inputs u_i and v_i as linear functions of the macro-inputs of u and v which can be computed by the method of least squares for the time series, so that we have 2, I equations:

(5) $$\begin{cases} u_i = A_{u_i} + B_{u_i}u + C_{u_i}v + E_{u_i} \\ v_i = A_{v_i} + B_{v_i}u + C_{v_i}v + E_{v_i} \end{cases} \qquad i=1, \ldots I$$

where E's are residuals with zero means and uncorrelated with the values assumed by the macro-inputs for each period of the time series. Substituting (5) in (2) we get

(6) $$x = \alpha + \beta u + \gamma v.$$

where

(7) $$\begin{cases} \alpha = \sum_{i=1}^{I} \alpha_i + \sum_{i=1}^{I} (A_{u_i}\beta_i + A_{v_i}\gamma_i) \\ \beta = \sum_{i=1}^{I} (\beta_i B_{u_i} + \gamma_i B_{v_i}) \\ \gamma = \sum_{i=1}^{I} (\gamma_i C_{v_i} + \beta_i C_{u_i}) \end{cases}$$

From (7) it is obvious that the macro-parameters depend on the micro-inputs for the time series, because the coefficients, A's, B's and C's in (5) depend on them.

The macro-parameter α does not depend only on the "corresponding micro-parameters" α_i, but also on the other non-corresponding micro-parameters β's and γ's.

The macro-parameters β and γ do not depend on the micro-parameters α_i, but they depend not only on the corresponding parameters β_i and γ_i respectively, but on both.

The three statements made above can be immediately verified by looking at the equation in (5).

By showing that[1]

(8) $$\sum_{i=1}^{I} A_{u_i} = 0, \quad \sum_{i=1}^{I} A_{v_i} = 0$$

and

1) H. Theil, op.cit. Chap. IX, pp. 186-189.

$$(9)\quad \begin{cases} \sum_{i=1}^{I} B_{u_i}^{i} = 1, & \sum_{i=1}^{I} B_{v_i}^{i} = 1 \\ \sum_{i=1}^{I} C_{u_i}^{i} = 1, & \sum_{i=1}^{I} C_{v_i}^{i} = 1 \end{cases}$$

when the number of explanatory variables is the same in (5) as in (2), but all the 4 sums in (9) being equal to 0 if this does not hold, (7) can be reduced to the following:

$$(10)\quad \begin{aligned} \alpha &= I\bar{\alpha} + I\left\{\text{cov}\left(\beta_i^{i} A_{u_i}^{i}\right) + \text{cov}\left(\gamma_i^{i} A_{v_i}^{i}\right)\right\} \\ \beta &= \bar{\beta} + I\left\{\text{cov}\left(\beta_i^{i} B_{u_i}^{i}\right) + \text{cov}\left(\gamma_i^{i} B_{v_i}^{i}\right)\right\} \\ \gamma &= \bar{\gamma} + I\left\{\text{cov}\left(\beta_i^{i} C_{u_i}^{i}\right) + \text{cov}\left(\gamma_i^{i} C_{v_i}^{i}\right)\right\} \end{aligned}$$

From (10) it is seen that the macro-intercept α is the sum of the micro-incepts α_i and the macro-coefficients (or the macro slopes) β and γ are equal to the average value of the corresponding micro coefficients (or micro-slopes) β_i and γ_i, both apart from certain covariance corrections.

So the whole matter in the linear aggregation of micro-production functions just as for any other economic relation, hinges on the quantification of the covariance corrections. So far, it seems, no headway has been made in studying the limits or restrictions on these covariances.[1] And so long as our statistical knowledge about these covariances does not improve, the idea of deriving aggregate or macro-production function from micro function can hardly be rewarding.

The example of aggregation illustrated above is the simplest that can be used. It relates to units producing the same commodity with the same factors and their micro functions are linear in the micro-variables. Prof. Theil has discussed other cases when the micro-functions are polynomial in the micro-variables and also aggregation over-time. But the basic approach is the same as stated above. Even these results have been possible because the functions used are additively separable which must be the case for a suitable aggregation as was proved

1) A.A. Walters, "Production and Cost Functions: An Economic Survey", Econometrica, Jan.- April 1963.

sensible aggregation as was proved earlier by Nataf.[1]

When we consider non-additive production functions, aggregation not over the same commodities but over different commodities, with changing micro-parameters and micro-variables and varying changes in prices of goods and factors, efforts to establish relationships between the micro-parameters and macro-parameters appear even less satisfactory.

References:

L.R. Klein : 'Macro-economics and the Theory of Rational Behaviour,' Econometrica, 1964.

H. Theil : Linear Aggregation of Economic Relations, Amsterdam 1954.

Ragnar Bentzel, 'On the Aggregation of Production Functions' International Economic Papers, No. 8.

1) A Nataf, op.cit.

Lecture 4.

Some Well-Known Production Functions and their Properties.

I. Introduction, . II. Capital-output ratios, . III. Cobb-Douglas production function, . IV. SMAC production function, . V. Derivation of cost-function, . VI. Appendix, .

I.

As stated by Heady and Dillon[1], 'Justus von Liebig's "law of the minimum" was the first attempt to define the fundamental relationship between fertilizer and crop yields.[2] He stated that soil containing all nutrients necessary for plant growth except one is barren for all crops for which the lacking nutrient is indispensible.' Thus each nutrient that is needed for the production of a crop is a limitational factor to the others. 'The production surface would reduce to a simple "knife's edge" with a constant slope to the maximum per acre yield.' The purpose for mentioning this earliest attempt at describing a production function is not historical curiosity, but the fact that a tremendous amount of theoretical and empirical research work carried out during the last two or three decades from the simple Harrod-Domar type models to the Leontief's input-output models is based essentially on the assumptions of strict complimentarity which characterises von Liebig's definition. We shall discuss input-output models later, but it will be interesting to comment upon a special type of ratio, called the capital-output ratio, which is the backbone of the Harrod-Domar type of growth models and is very frequently used as a production function.

---

1) Heady and Dillon, Agricultural Production Functions, Iowa State University Press, 1961. For historical account of the development of production functions refer to Chapter 1 of this book.

2) von Liebig, Justus, Die Grundsätze der Agricultur-Chemi mit Rücksicht auf die in England angestellten untersuchungen. Friedrich Viewig und Sohn, Braunschweig, 1855.

II.

Capital-output ratio is the ratio of capital to output, $\frac{K}{P}$, where K is the stock of capital and P the output, produced by it. The ratio is related to a certain period usually a year. If $\frac{K}{P} = 6$, then the production function can be expressed as $P = \frac{1}{6} K$. The concept is not as simple as it appears nor is it easy to estimate. Measurement of any economic variables is difficult, it is much more so in case of capital. If there is one simple type of capital and the resulting output is one of homogenous quality, the ratio between the two can be easily found. In reality a firm combines capital items of vastly different types and grades to produce several joint and by-products of differring qualities. Hence in the evaluation of capital and output even in case of a single firm index number problems are involved. When we consider an industry as a whole or the economy as a whole, which are more relevant cases for planning and policy making purpose, these problems assume huge proportions. We shall not discuss the estimation problems here. But apart from these, there are several conceptual clarifications needed for computing the ratio more or less accurately.

The ratio is derived by dividing the value of capital stock needed by the value of output produced. For example, if residential buildings are included in capital, their rents should be included in the output. In certain cases it is not easy to do this; for instance, public expenditure on parks or highways, etc., can be quite properly included in the stock of capital, but there is no way of evaluating or imputing a value to the services produced by these items. On the other hand certain non-physical capital, such as the stock of knowledge, improvement in skill etc. has to be excluded simply because we do not know any way of evaluating it so far.

A normal proportion of inventories may be included in the stock of capital, but there may arise certain discrepancies as to the determination of 'the amount of capital needed itself.' If there is no such thing as 'capital exhaustion' then production can be carried on in three shifts, but the capital needed will correspondingly increase if production is executed in less than three shifts. Certainly, some idea about normal working hours of factories and firms will have to be fixed before 'capital needed' can be estimated.

The depreciation and replacement considerations raise additional but well-known and frequently discussed difficulties in the estimation of the stock of capital. A systematic method of evaluation will have to be evolved to estimate the value of the capital stock in successive years of its life-time till it is retired. Prof. Kuznets presents several estimates of capital-output ratio; gross capital output to gross national product, gross capital stock net of capital retirements to net national product, and net (of depreciation) capital stock to net national product.[1)]

Capital-output ratios can be further distinguished according as they are ratios of total capital stock to total output during a year or they are ratios of increments to capital stock to increases in the output during a year. The latter are called marginal capital output ratios, and the former average capital-output ratios. In the field of development policy, the marginal ratios are more useful, though in their computation the underutilization of new capital stock created must be accounted for. Unfortunately in most of the underdeveloped countries, it is the marginal capital-output ratios alone which can be computed with higher or smaller degree of reliability. But the problem of under-utilisation of capital is rather confusing and complex, for as we have stated, it is not known how many hours per day a capital item can be worked and certainly the degree of under-utilization cannot be determined without some appropriate idea about this fact. Even so, it will be extremely difficult to estimate the value of capital actually in any given year. This drawback is very serious for depression years. The series computed by Prof. Kuznets generally suffer from this defect.

As can be easily understood the average capital output ratios have been more stable than the marginal capital output ratios. Annual increases in capital stocks and resulting outputs form a very small proportion of the aggregate capital stock and annual outputs and hence when the former are added to the latter, the over-all average changes in much smaller degree as compared with the variations in the incremental ratios. And if there is consistent tendency in the latter, the former will also tend to move towards the level attained by the marginal ratios.

---

1) Cf. E.D. Domar, "The Capital-Output Ratio in the United States: Its variation and stability" in The Theory of Capital, edited by E.A. Lutz and D. Hague, (Macmillan) 1961.

A characteristic of capital-output ratios that has been discovered empirically is that they are fairly stable, with larger variations in the marginal ratios than in the average as is to be expected. The following two tables give an idea of these ratios in case of the United States.

The use of capital-output ratio in planning future output or growth does not imply that capital alone does the whole trick unaccompanied by labour and other factors. The other factors are, however, supposed to be present in quantities that might be needed. They may be in excess supply but they are assumed not to be in short supply. According to this approach capital is all important, other factors are neglected or they play, a secondary role, and thus the approach is, at best, partial. In advanced countries, there may occur shortage of ordinary labour and in the underdeveloped countries, skilled and managerial labour is in as short supply as capital.

It is evident that the reciprocal of the average capital-output ratio gives the average productivity of capital and the reciprocal of marginal capital-output ratio gives the marginal productivity of capital. As these ratios include only one factor of production, it will be idle to talk about the rate of sustitute, or the elasticity of rate of substitution or the derivation of the short or the long term cost functions, based on these ratios.

III.

The equation which is probably the most popular production function, was arrived at by Paul H. Douglas in association with C.W. Cobb[1] and is popularly known as the Cobb-Douglas production function. Actually it traces back to Wicksell[2] who stated in a footnote the function as

(1) $$P = a^{\alpha} b^{\beta} c^{\gamma}$$

---

1) C.W. Cobb and P.A. Douglas, The Theory of Wages, Macmillan Co., New York, 1934.

2) Knut Wicksell, Den "kritiska punkten" i lagen för jordbrukets aftagande produktivitet. Ekonomisk Tidskrift, 1916, pp. 585-292, cited by Heady, Dillon, op.cit. pp. 15-16.

Ratio of Capital Stock to Annual National Product, United States, 1869-1955, based on totals in 1929 prices.

	Date of Capital Stock	Period of National Product Flow	Ratio, Gross Capital Stock to Gross National product		Ratio, Gross Capital Stock Net of Retirements to Net National product		Ratio, Net Capital Stock to Net National product	
			Kuznets Concept (Variant III) (1)	Department of Commerce Concept (2)	Kuznets Concept (Variant III) (3)	Department of Commerce Concept (4)	Kuznets Concept (Variant III) (5)	Department of Commerce Concept (6)
	Total Stock and Product (including Military)							
1.	1869 and 1879	1869-1878	5.2	5.3	4.6	4.6	3.5	3.5
2.	1879 and 1889	1879-1888	4.5	4.5	3.9	3.8	2.9	2.9
3.	1889 and 1899	1889-1898	5.3	5.2	4.6	4.5	3.5	3.4
4.	1899 and 1909	1899-1908	5.4	5.3	4.6	4.5	3.4	3.4
5.	1909 and 1919	1909-1918	6.0	5.9	5.2	5.0	3.7	3.6
6.	1919 and 1929	1919-1928	6.2	6.0	5.2	5.0	3.6	3.5
7.	1929 and 1939	1929-1938	7.6	7.3	6.0	5.7	4.1	3.9
8.	1939 and 1949	1939-1948	6.5	5.4	4.6	3.8	3.1	2.5
9.	1949 and 1955	1949-1955	6.0	5.4	4.0	3.6	2.8	2.5

Source: Simon Kuznets, Capital in the American Economy: Its Formation and Financing, National Bureau of Economic Research (mimeographed), Table III-5, pp. III-35-36. The figures for 1946-1955 were kindly supplied by Daniel Creamer.
Cf. E.D. Domar, op.cit. p. 101.

Ratio of Changes in Capital (Capital Formation) to Changes in Annual National Product, United States, 1869-1955, based on total in 1929 prices.

Intervals over which changes are compared (dates are for end of year, unless otherwise noted)	Ratio, Gross Capital Stock to Gross National Product		Ratio, Gross Capital Stock Net of Capital Retirements to Net National Product		Ratio Net Capital Stock to Net National Product	
	Kuznets Concept (Variant III) (1)	Department of Commerce Concept (2)	Kuznets Concept (Variant III) (3)	Department of Commerce Concept (4)	Kuznets Concept (Variant III) (5)	Department of Commerce Concept (6)
Total (including Military)						
1. 1973-1883	3.6	3.6	3.0	3.0	2.3	2.3
2. 1883-1893	7.5	7.2	6.7	6.4	5.1	4.9
3. 1893-1903	5.6	5.4	4.6	4.5	3.4	3.2
4. 1903-1913	7.7	7.4	6.8	6.4	4.6	4.3
5. 1913-1923	6.7	6.3	5.3	5.0	3.4	3.2
6. 1923-1933	30.2	25.6	20.1	16.2	11.8	9.6
7. 1933-1943	4.1	2.8	1.8	1.1	1.1	0.7
8. 1943-1953	4.8	5.3	2.1	2.4	1.5	1.8

Source: Simon Kuznets, Capital in the American Economy: Its Formation and Financing, National Bureau of Economic Research (mimeographed), Table III-6, pp% III-41-42. Cf. E',D. Domar, op.cit. p. 102.

and said that α, β, and γ should sum to 1.0. In early 1927 Douglas computed indexes for American manufacturing of the numbers of workers employed from years 1899 to 1922 as well as the indexes of the amount of fixed capital in manufacturing deflated to dollars of approximately constant purchasing power, and then plotting these on a log scale together with the day index of physical production of manufacturing and observed that the product curve lay consistently between the two curves for the factors of production and tended to be approximately a quarter of the relative distance between the curve of the index for labour, which showed the least increase over the period and that of the index for capital which showed the most.[1] At the suggestion of C.W. Cobb, the sum of the exponents was made equal to unity in the formula

(2) $$x = aL^{\alpha} K^{1-\alpha}$$

The value of a and α was found by the method of least squares and as was expected because of the relative distance of the product curve from those of the two factors, the value of α was found to be .75. Using this value of α the theoretical value of the product was estimated for each year from (2) and it was found that divergencies between the actual and theoretical product were not great giving due allowance for the imperfect nature of the indexes of capital and labour and the use of index of capital measuring the quantities which were <u>available for</u>, rather than their relative <u>degree of use</u>.

Later the authors introduced two new features in their investigations. David Durand, in an article published in 1937, urged that the restriction that the indices of labour and capital sum to unity in the function be abandoned. He argued that 'the use of α and $1-\alpha$ in the function, assumed the existence of an economic law which it should be one of the tasks of science to test, namely, the assumption of true constant returns'. The authors adopted Durand's suggestion and decided that they should try to find the values in terms of the formula:

(3) $$x = aL^{\alpha} K^{\beta}.$$

---

1) Cf. Presidential address delivered at the Sixtieth Annual Meeting of the American Economic Association, Chicago, Ill., December 29, 1947.

In (3) if the exponents of capital and labour were independently determined, it would then be possible for the sum of the exponents to be wither greater or less than unity and hence to show true increasing, decreasing, or constant returns to scale.

Hitherto the authors had 'dealt only with time studies and had found the values of their exponents from index numbers of labour, capital and product within a given economy, with each year serving as a separate observation.' Now they broadened their field of investigation and made cross-section analysis between industries in a given economy for specific years. 'In these studies, differences between industries in the quantities of their net value product were presumed to be a function of the total number of employees and of the total quantities of fixed and working capital with each industry serving as a separate observation'. The production function based on cross-section studies is a somewhat different production function from that which is based on time series. Since the product is expressed in value items, its value in the individual industries is not the result of changes in the increments to the total physical product but also of changes in the exchange value, or the relative price per unit of the products of an industry. 'The net values turned out by the respective industries will, therefore, be affected in these cases not only by the quantities produced but also by the respective demand curves for the products'. The production function derived from cross-section studies, therefore, may be criticised on the ground that it 'does not measure production at all and is in no sense a test of marginal productivity theory'. Prof. Douglas,[1)] however, meets this criticizm by stating that' 'the marginal productivity theory has always implicitly dealt in terms of values as well as of physical quantities since it assumes that the supplies of labour and capital in each of the various industries are regulated by the principle that the respective marginal labourers will produce equal amounts of value as will the marginal units of capital'. In other words factor allocation in a market economy takes place more according to the marginal value productivity theory than that in any other way.

---

1) P.H. Douglas, ibid.

The important properties of the Cobb-Douglas production function are:

(1) If one of the factors L and K assumes a zero value, the total production becomes zero.

(2) α and β are the elasticities of production with respect to labour and capital, since

$$\frac{L}{x}\frac{dx}{dL} = \alpha \quad \text{and} \quad \frac{K}{x}\frac{dx}{dK} = \beta .$$

this means that the ratio between marginal and average production is constant and the two change always in the same proportion.

(3) The function is homogeneous of degree $\alpha + \beta$. If $\alpha + \beta > 1$, there are increasing returns to scale, if $\alpha + \beta < 1$, decreasing returns to scale and $\alpha + \beta = 1$ indicates constant returns to scale.

(4) Marginal productivity of labour and capital declines as their respective amounts increase, if $\alpha < 1$ and $\beta < 1$. i.e. $\frac{\partial x}{\partial L} > 0$, $\frac{\partial x}{\partial K} > 0$ and $\frac{\partial^2 x}{\partial L^2} < 0$ and $\frac{\partial^2 n}{\partial K^2} < 0$.

(5) Marginal productivity of labour increases when the amount of capital is increased and vice-versa i.e.

$$\frac{\partial^2 n}{\partial L \partial K} > 0.$$

(6) The marginal rate of substitution between capital and labour is $\frac{\alpha K}{\beta L}$.

(7) The elasticity of substitution between the factors is unity.

Cobb-Douglas' production function has been criticised on several rounds. Some of these criticisms are related to (1) lack of adequate data, (2) failure to use quantities that are exactly relevant, e.g. failure to use the values of quantities of capital actually utilized instead of total capital available, (3) the use of quantities of factors without differentiating according to quality e.g. in case of labour, and (4) other shortcomings in the measurement of the variables and the statistical estimation of the parameters etc. These defects of Cobb-Douglas function are more or less common to all quantitative investigations in almost all economic fields. There are, however, a couple of criticisms which seem to be more serious. One has been put forward by H. Menderhausen[1] who found a high degree of multicollinearity betwmen the variables in the time series data. 'He presented a three-dimensional figure to show that the product quantities lie practically in a straight line over the input plane rather than forming a surface dispersed over the resource plane'. What this criticism amounts to is that the function, instead of establishing any fundamental production law, describes the fact that the logarithms of the three variables happen to change at constant rates with time so that it is difficult to say whether any causal relationship exists, between the variables as such. Evidently this criticism is not applicable to cross-section studies.

A second line of criticism worth mentioning here has come from Reader[2] and Smith[3]. Reader points out the lack of comformity between the Cobb-Douglas production function and the production function postulated in theory. His main point is that in theory the physical production function shows the functional relationship between the input quantities and the output of a firm. The functions derived by Douglas are the loci of input-output quantities of all firms used in the particular study'. In reply to this criticism Bronfenbrenner[4] has pointed out that 'under competition, the slope of the production function under equilibrium should be the same between industries and firms'. Smith has raised several pertinent points, two of which can be mentioned here. Theoretically, the relevant input is the annual use of capital, but in fitting in function ordinarily the capital investment is used. If the two bear a constant relation, it will not be very objectionable, for the elasticities will be the same. Otherwise, there will occur a discrepancy.

---

1) Horst Menderhausen, On the Significance of Professor Douglas' Production Function, Econometrica, 1938.

2) M.W. Reader, An Alternative Interpretation of the Cobb-Douglas Production Function, Econometrica, 1943.

3) V.E. Smith, The Statistical Production Function, Econometrica, 1945.

4) M. Bronfenbrenner, "Production Functions: Cobb-Douglas Interfirm, Intrafirm", Econometrica, 1944.

Secondly, firms in a cross-section study may employ different techniques, particularly due to fixed plants inherited from the past, and the long run production functions so derived may represent "mongrels" or "hybrids".

It seems fair to state that a general production function of the Cobb-Douglas type describes at best the central over-all tendency in the movement of factors and the resulting output; it is unjustified to try to find a prototype in the actual world, except perhaps in the ideal economy with perfect present and future knowledge about everything with perfect mobility of factors and with perfect divisibility etc.

IV.

Ever since Cobb-Douglas function was discovered in 1927, it has received a remarkable degree of attention from research workers, economists, statisticians, econometricians and mathematical economists. Besides the basic properties of the function, one reason for its successful career has been that there was no serious alternative to this function. Recently a new production function has been arrived at by Arrow, Chenery, Minhas and Solow and for brevity it has been labled as SMAC. The basic change introduced by this function is to allow the elasticity of substitution to be constant at a value other than one as in case of Cobb-Douglas or zero as in case of facter combination in fixed proportions. The function[1)] is

$$(4) \qquad x = \gamma(\delta K^{-\rho} + (1-\delta)L^{-\rho})^{\frac{v}{\rho}}$$

(4) is a homogeneous function, it is linear when $v = 1$. It has an efficiency parameter γ which of course can be tacked on to any production function. The distribution parameter δ $(0 \leq \delta \leq 1)$ determines the division of factor income. The substitution parameter, ρ, is the simple function of the elasticity of substitution, thus, $\sigma = \frac{1}{1+\rho}$. When σ is infinite, $\rho = 1$; and when $\sigma = 0$, $\rho = \infty$.

The properties of SMAC function as compared to the Cobb-Douglas are:

(a) SMAC is a more general function in the sense that by choosing appropriate values for ρ, it can be specialised to the fixed factor proportion of the input-output type or the Cobb-Douglas production function. When $\rho \rightarrow \infty$, it represents the former and when $\rho \rightarrow 0$ it represents the latter.[1)]

(b) Cobb-Douglas is more general in two senses: elasticity of production of labour and capital may be different in the Cobb-Douglas when, as is obvious, $\alpha \neq \beta$. This possibility has not yet been introduced in the SMAC.[2)] Secondly, the Cobb-Douglas can be generalized to any number of factors as has been done quite frequently in empirical research, it has yet to be seen whether SMAC can be extended to more than two factors.

(c) The production is not reduced to zero when one of the factors is at zero level in the SMAC, as is the case in the Cobb-Douglas. This is, however, not a serious defect of the latter, as we can hardly visualize a situation where one of the factors, labour and capital, alone is employed all alone in production. In all cases the two factors are combined together whatever their proportions might be.

(d) The function is homogeneous of degree v. If $v > 1$, there are increasing returns to scale, if $v < 1$, decreasing returns to scale, if $v = 1$, constant returns to scale.

(e) When $v = 1$, the marginal product of capital is $v^{-\rho}(\frac{x}{K})^{1+\rho}$ and that of labour is $(1 - \delta)V^{-\rho}(\frac{x}{L})^{1+\rho}$. The marginal product of capital increases at an increasing, zero or decreasing rate according as $x\delta^{\frac{1}{\rho}} \gtrless K$. The marginal product of labour increases at an increasing, zero or decreasing rate according as $x(1 - \delta)^{\frac{1}{\rho}} \gtrless L$. In this respect the Cobb-Douglas conforms more readily to the theoretical requirement of convexity than the SMAC.

(f) The marginal rate of substitution of the factor in the SMAC is $\frac{1 - \delta}{\delta}(\frac{k}{L})^{1+\rho}$ which is now a function of the factor and the elasticity of substitution and income division factor. In Cobb-Douglas it is a function of factor ratio and the ratio of production elasticities.

(g) The elasticity of substitution of the factors in the SMAC is $\sigma = \frac{1}{1 + \rho}$. It is unity in case of Cobb-Douglas. This is the chief merit of SMAC that it releases the restriction on the σ which can now differ from unity.

We have briefly discussed the properties of the SMAC function. As for its criticisms, it is almost open to the same criticisms as the Cobb-Douglas, as regards the measurement of output and factors and the applicability of statistical methods. We have pointed out in what senses Cobb-Douglas is more general and in what the SMAC. Cobb-Douglas turns out to be a special case of the SMAC. It remains, however, to be seen whether the special case typifies economic reality and conforms to the requirements of economic theory more than the other cases. In a competitive economy, given the income distribution, the factors are used in proportion to their marginal productivities, and this conforms to the elasticity of substitution being unity than otherwise. About a non-competitive economy little is known, but in order that SMAC may be accredited with some superiority over Cobb-Douglas, it will have to be tested that σ remains constant and that it is substitutionally different from unity.

V.

It is not possible to derive cost function from the capital-output ratios as such. For we do not get any idea of how much labour (or any other factor) is combined with capital. Hence it is impossible to find out the marginal productivity of labour. Though we know the capital per unit of output in the over-all or the marginal sense from the capital-output ratios, it is not possible to derive the marginal productivity of capital from these ratios.[1] Even if we consider Leontief's type, fixed proportion factor combination, and assume that capital and labour are used in a constant proportion $\frac{K}{L}$, we cannot derive the marginal productivity of these factors. The derivation of marginal productivity of factors by itself does not enable us to estimate the cost function. But if we make the further and the omnipotent assumption of perfect competition, these marginal productivities would equal their repsective prices and then the estimation of cost of production would be possible. As the capital-output ratios or the production systems combining factors in fixed proportions do not furnish absolute or relative prices of

1) In the literature, marginal productivity of capital has been sometimes confused with return on capital or output-capital ratio, which is unfortunate.

factors endogenously under any condition, we cannot compute cost functions therefrom.

In case of Cobb-Douglas production function, under assumptions of perfect competition, the price of labour

$$W = \frac{\partial x}{\partial L} \cdot P = a\alpha L^{\alpha-1} K^{\beta} \cdot P$$

and the price of capital

$$R = \frac{\partial x}{\partial K} \cdot P = a\beta L^{\alpha} K^{\beta-1} \cdot P$$

where P is the price of output of the production function.

(1) $$x = a L^{\alpha} K^{\beta}$$

(2) $$\therefore \quad \frac{W}{R} = \frac{\alpha K}{\beta L}$$

Solving for L and K from (1) and (2)

$$L = \frac{\alpha R}{\beta W}\left[\left(\frac{x}{a}\right)^{\frac{1}{\alpha}} \cdot \frac{\beta W}{\alpha R}\right]^{\frac{\alpha}{\alpha+\beta}}$$

$$K = \left[\left(\frac{x}{a}\right)^{\frac{1}{\alpha}} \cdot \frac{\beta W}{\alpha R}\right]^{\frac{\alpha}{\alpha+\beta}}$$

The cost function is

(3) $$C = L \cdot W + KR$$

$$= R \cdot \left(\frac{\alpha+\beta}{\beta}\right)\left[\left(\frac{\beta W}{\alpha R}\right)^{\alpha} \frac{x}{a}\right]^{\frac{1}{\alpha+\beta}}$$

substituting the values of L and K in (3).

The supply function can be obtained by equating $\frac{dC}{dx}$ to P, and rearranging so that

(4) $$x = \left[a\left(\frac{\alpha}{W}\right)^{\alpha} \cdot \left(\frac{\beta}{R}\right)^{\beta} \cdot P^{\alpha+\beta}\right]^{\frac{1}{1-\alpha-\beta}}$$

In the case of SMAC, the cost function can be derived similarly by suitable substitutions. In the case of perfect competition in factor markets, we have

$$W = \frac{\partial x}{\partial L} = -\frac{v}{\rho}\left[\delta K^{-\rho} + (1-\delta)L^{-\rho}\right]^{-\frac{1}{\rho}-1} . (1-\delta)(-\rho)L^{-\rho-1}$$

$$R = \frac{\partial x}{\partial K} = -\frac{v}{\rho}\left[\delta K^{-\rho} + (1-\delta)L^{-\rho}\right]^{-\frac{1}{\rho}-1} . \delta(-\rho)K^{-\rho-1}$$

$$\therefore \quad \frac{W}{R} = \frac{1-\delta}{\delta}\left(\frac{L}{K}\right)^{-\rho-1}$$

$$\therefore \quad L = \left[\frac{W\delta}{R(1-\delta)}\right]^{\frac{1}{-\rho-1}} ; K$$

(5)
$$= \left[\frac{R(1-\delta)}{W\delta}\right]^{\frac{1}{\rho+1}} . K,$$

substituting (5) in the expression for cost function

$$C = WL + KR$$

we have

(6)
$$C = \left[W\left(\frac{R(1-\delta)}{W\delta}\right)^{\frac{1}{\rho+1}} + R\right] K$$

(6) gives the long run cost curve with capital as the independent variable. If we substitute L in the SMAC function, we have

(7)
$$K = \frac{x}{v\left[\delta + (1-\delta)\left(\frac{R(1-\delta)}{W\delta}\right)^{-\frac{\rho}{\rho+1}}\right]^{-\frac{1}{\rho}}}$$

Putting the value of K from (7) in (6) we can get the cost function in terms of output. The cost function is a linear function in output as is to be expected from a linear homogeneous function. The supply function can be obtained as in case of the Cobb-Douglas function.

References:

Paul H. Douglas, The Theory of Wages, (New York) 1957.

" " " , "Are There Laws of Production", Presidential address delivered at the Sixtieth Annual Meeting of the American Economic Association, Chicago, Illinois, Dec. 29, 1947. Reprinted in The Theory of Wages.

E.O. Heady & J.L. Dillon , Agricultural Production Function (Iowa) 1961, Chap. 3.

F. Lutz & D. Haque (Ed.) , The Theory of Capital, (Macmillan) 1961.

A.A. Walters , "Production and Cost Functions: An Econometric Survey", Econometrica, 1963.

oo 0 oo

Lecture 5.

Derivation of Production Functions

I.

As in other fields, two types of data can be used for the derivation of production functions; time series or cross section. The time series samples may consist of periodic observations of outputs and inputs for a single firm, an industry or the economy as a whole. For an economy the appropriate measurement of output would be in terms of gross national product measured in constant prices. For an industry the conventional and easily available index of industrial production may be used. Labour inputs can be measured in terms of labour hours employed. As for capital stock in the industry or in the economy, it has to be measured in terms of index of its value obtained from series on net investment and initial assets. In case of a particular firm related to a time series analysis, the variables may be measured in terms of physical quantities.

Cross-section samples consist of output and input statistics for individual firms in an industry or individual industries in an economy or individual regions in an economy or a larger part of the world at a particular time or during a given period. In cross section studies, of course, indexes of variables have not to be used.

In an intra-firm study of an industry, physical amounts of output or outputs of each firm can be used if data about their quantities are available. In the absence of such data, gross value of production of each firm can serve the purpose in view of the fact that within an industry firm to firm variations in value of output are good indicators of physical output at a given point of time or during the production period. As regards fuel, raw materials and other intermediate products, they can be measured either in physical units or value terms; in most cases value data will be more easily available and it can be more conveniently used. Measures of labour can be expressed in man-hours and that of capital in value terms.

In an inter-induatrial study, physical data can hardly be used since output and inputs vary very greatly in their composition from industry to industry. In such studies, value data have to be used even though they are not as good indicators of movements in physical output as they are in case of inter-firm studies.

The data about an individual firm can be obtained directly from the reports of the firm. The data for an industry to be used in cross-section inter-industrial studies are generally published in compilations like census of manufacturing. Such publications contain such data about industrial groups as gross output, number of employees or manhours worked, wage bill and asset values. The research worker has to derive from these rough statistics the data he wants to use in constructing the production function in a way which conforms as closely as possible to the definition of the variables which has been chosen or agreed to in advance in relation to the purpose for which the function is likely to be used.

II.

After a brief account of the type of data that will be needed and will be available, we turn to the statistical techniques that can be used in deriving the production function we wish to estimate. Two basic approaches to the estimation of a production function are possible: (1) simultaneous equations approach and (2) single equation approach. In the first approach the production function is treated as one of the several relations that describe the economic phenomena to be studied. In the second, the production function is considered as an independent relationship, between the output and the inputs, uninfluenced by other relationships that characterize the economy in which the productive activity is conducted. While the first is a more comprehensive approach, it requires more sophisticated statistical methods and is not easy as regards computations. The second approach requires well-known statistical methods and the corresponding production function can be more easily computed. As the second approach has been generally adopted in deriving the production functions, we shall confine ourselves to this only. It can, however, be pointed out that if we can include all the factors that influence the output in a systematic way, and add a disturbance factor to represent all other influences such as those of social milieu and other imponderables, the single equation approach may give nearly as good result as the simultaneous equations approach. For the sake of simplicity, we shall outline a method for deriving a linear production function only. This is firstly because the two production functions that we have discussed can be transformed into linear equations and secondly because that the derivation of non-linear equations proceeds on similar lines.

III.

Let us suppose output is a function of three variables which represent three factor inputs. I number of observations are made of the outputs and the corresponding inputs. Let X_i (i = 1,....,I) represent the values of outputs for succesive observations and U_i (i = 1,...,I) V_i (i = 1,....,I) and W_i (i = 1, ...,I) be the values of corresponding inputs observed. In a cross-section study i's will represent individual firms or individual industries at a particular point of time or during an accounting period such as a year, so that X_i will denote the output of the i_{th} firm or industry which uses U_i, V_i and W_i of factors U, V and W respectively. In a time series study, i's may represent successive points of time or periods of production such as a year. We can record these observations in a tabular form

Observation Number	Output	Input U	Input V	Input W
1	X_1	U_1	V_1	W_1
2	X_2	U_2	V_2	W_2
...	...	...	...	...
i	X_i	U_i	V_i	W_i
...	...	...	...	...
I	X_I	U_I	V_I	W_I

In our example there are four variables. The number of observations must be greater than or equal to four, i.e. $I \geq 4$. Our problem is to find a linear function in which U, V and W appear as explantory (or independent) variables which explain or predict the (dependent) variable X to be explained or predicted. It is to be understood that the observations that have been made and recorded are but a small sample out of a large number of observations that are possible.

Let us then postulate a linear relation between the variables as follows:

$$X_i = \beta_o + \beta_u U_i + \beta_v V_i + \beta_w W_i + E_i \quad (1)$$

(1) states that X is linearly dependent on U, V and W, that β_o, β_u, β_v, and β_w are the coefficients to be determined, that β_o is the intercept of the line and is zero when the line passes through the origin, β_u, β_v, and β_w are respectively the grandients of the line with respect to U, V and W, and give the increase in X due to one unit increase in each of the latter respectively, and that E is the disturbance, residual or the error between the actual value of X observed and that computed from the relation (1), so that

$$X_i - X_i^c = E_i \quad (2)$$

where X_i^c is the value of X_i computed from (1) leaving out the disturbance E_i. In statistical or econometric theory much depends on the assumptions about disturbance or residual term. We have explained earlier[1] that this disturbance is caused by several reasons whose effects cannot be recorded in the causal factors included in equation (1). The usual assumptions about E are:

a) Mean value of E_i ($i=1,\ldots I$) is zero
b) Co-variance of E_i and E_j is zero for $i \neq j$ and σ_E^2 for $i=j$.

The smaller the values of E_i and lesser the deviations the more the chance that the estimated values of β_o, β_u, β_v and β_w will give such values of X_i^c which will differ from X_i as little as possible. In other words, if we minimize the standard deviation of E_i and find the values of β's that bring about this minimisation, we can get a linear equation which fits our observations as best as possible. The mean of E_i's being zero, their squared standard deviation or their variance[1] is (summation being all through from 1 to I)

---

1) See page

1) If we can minimize the variance, the standard deviation will be minimized. It is better to do the former to avoid square roots. In fact the two alternatives are the same.

(3) $$\Sigma E_i^2 = \Sigma (X_i - X_i^c)^2 = \Sigma (X_i - \beta_o - \beta_u U_i - \beta_v V_i - \beta_w W_i)^2$$

As already stated, we want to find such values of β_o, β_u, β_v, and β_w that minimize (3). For this we differentiate (3) with respect to each of these parameters and equate to zero, to get

(4) $$\begin{cases} \Sigma X_i = I\beta_o + \beta_u \Sigma U_i + \beta_v \Sigma V_i + \beta_w \Sigma W_i \\ \Sigma X_i U_i = \beta_o \Sigma U_i + \beta_u \Sigma U_i^2 + \beta_v \Sigma U_i V_i + \beta_w \Sigma U_i W_i \\ \Sigma X_i V_i = \beta_o \Sigma V_i + \beta_u \Sigma U_i V_i + \beta_v \Sigma V_i^2 + \beta_w \Sigma V_i W_i \\ \Sigma X_i W_i = \beta_o \Sigma W_i + \beta_u \Sigma U_i W_i + \beta_v \Sigma V_i W_i + \beta_w \Sigma W_i^2 \end{cases}$$

Equations in (4) are called normal equations. By solving the four equations above for β_o, β_u, β_v, and β_w we get the value of these parameters that minimize the difference between the actual observed values of output and the computed values of the outputs. Let the values of the parameters found after solution be b_o, b_u, b_v, and b_w respectively, then the linear production function thatwe wanted to derive is

(5) $$X = b_o + b_u U + b_v V + b_w W$$

It should be noticed that the above method of deriving the function through the principle of least squares is centred upon X, the dependent variable. Such a procedure implies that the estimates of the values of the parameters would have been different, if instead of minimizing the squares of the differences of the observed and postulated values of X, we would have minimized the corresponding squares related to one of U, V and W.[1] Thus the above estimated regression equation is an irreversable relationship in terms of causal influences. It seems, however, more appropriate to work out the regression equation centering upon X rather than on any of U, V and W except when we are especially interested in one of them. In any case given the values of b's, X

---

1) This is one of the criticisms of the estimations of Cobb-Douglas production function.

and all but one of U, V, W we can predict the level at which the remaining variable was fixed in producing the given level of X.

The solution of the system of equation can be simplified in the following way. The first of the equations can be written as

$$\bar{X} = \beta_o + \beta_u \bar{U} + \beta_v \bar{V} + \beta_w \bar{W}$$

so that (6) $b_o = \beta_o = \bar{X} - \beta_u \bar{U} - \beta_v \bar{V} - \beta_w \bar{W}$

Substituting (6) in (3) we get

$$\sum E_i^2 = \sum\{(X_i - \bar{X}) - \beta_u(U_i - \bar{U}) - \beta_v(V_i - \bar{V}) - \beta_w(W_i - \bar{W})\}^2$$

$$= \sum(x_i - \beta_u u_i - \beta_v v_i - \beta_w w_i)^2$$

where $x_i = X_i - \bar{X}$, $u_i = U_i - \bar{U}$, $v_i = V_i - \bar{V}$ and $w_i = W_i - \bar{W}$

The last 3 equations of (4) leaving out the first one can be written now in the following way

$$(7) \quad \begin{cases} \sum(x_i u_i) = \beta_u \sum(u_i^2) + \beta_v \sum(u_i v_i) + \beta_w \sum(u_i w_i) \\ \sum(x_i v_i) = \beta_u \sum(u_i v_i) + \beta_v \sum(v_i^2) + \beta_w \sum(v_i w_i) \\ \sum(x_i w_i) = \beta_u \sum(u_i w_i) + \beta_v \sum(v_i w_i) + \beta_w \sum(w_i^2) \end{cases}$$

(7) can be written in matrix form in the following way, let

$$A = \begin{pmatrix} \sum u_i^2 & \sum u_i v_i & \sum u_i w_i \\ \sum u_i v_i & \sum v_i^2 & \sum v_i w_i \\ \sum u_i w_i & \sum v_i w_i & \sum w_i^2 \end{pmatrix}$$

$$G' = (\sum x_i u_i \quad \sum x_i v_i \quad \sum x_i w_i$$

$$B' = (\beta_u \quad \beta_v \quad \beta_w)$$

we can now write the equation (7) as $AB = G$ or

(8) $$B = A^{-1} G$$

(8) gives the values of b_u, b_v and b_w, corresponding to the values of β_u, β_v and β_w.[1)]

IV.

The statisticians or econometricians are not generally satisfied by just getting the values of the parameters, they go a step further to ascertain the reliability of their estimates. The usual techniques for doing this are the tests of significance and the calculation of confidence limits. The best that is, perhaps, possible to do is to determine whether the variables whose coefficients or parameters are derived do affect the output, i.e., these parameters are significantly different from zero. The formula for the well-known 'student' test is

(9) $$t = \frac{bi}{\sqrt{var\ (b_i)}} \qquad i = u, v, w.$$

The var (b_i) can be obtained from the inverted matrix A i.e. A^{-1}. If we denote the elements of A^{-1} by C_{ij} $(i, j = u, v, w)$, then

(10) $$var\ (b_i) = C_{ii}\ \sigma^2$$

where Cii is a diagonal element of the matrix A^{-1} and σ^2 is the variance of the error term.[2)]

Tables of t tests are available. They give values of t corresponding to degrees of freedom (which is defined as the number of observations minus the number of variables; in our case it will be I-4) at 5%, 1% and .1%.

---

1) Some easy methods known as do-little methods have been developed to solve equations like (8) which involve inversion of matrices ; cf. Anderson and Bancroft, Statistical Theory in Research, (McGraw Hill 1952), pp. 192-197 and 197-199.

2) If the true variance of error is not known, variance of the sample error, s^2, may be used, $\sum E_i^2$ in the present case, $E_i = X_i - X_i^c$.

we can now write the equation (7) as $AB = G$ or

(8) $$B = A^{-1} G$$

(8) gives the values of b_u, b_v and b_w, corresponding to the values of β_u, β_v and β_w.[1)]

IV.

The statisticians or econometricians are not generally satisfied by just getting the values of the parameters, they go a step further to ascertain the reliability of their estimates. The usual techniques for doing this are the tests of significance and the calculation of confidence limits. The best that is, perhaps, possible to do is to determine whether the variables whose coefficients or parameters are derived do affect the output, i.e., these parameters are significantly different from zero. The formula for the well-known 'student' test is

(9) $$t = \frac{bi}{\sqrt{var\ (b_i)}} \qquad i = u, v, w.$$

The var (b_i) can be obtained from the inverted matrix A i.e. A^{-1}. If we denote the elements of A^{-1} by C_{ij} $(i, j = u, v, w)$, then

(10) $$var\ (b_i) = C\ ii\ \sigma^2$$

where Cii is a diagonal element of the matrix A^{-1} and σ^2 is the variance of the error term.[2)]

Tables of t tests are available. They give values of t corresponding to degrees of freedom (which is defined as the number of observations minus the number of variables; in our case it will be I-4) at 5%, 1% and .1%.

---

1) Some easy methods known as do-little methods have been developed to solve equations like (8) which involve inversion of matrices; cf. Anderson and Bancroft, Statistical Theory in Research, (McGraw Hill 1952), pp. 192-197 and 197-199.

2) If the true variance of error is not known, variance of the sample error, s^2, may be used, $\sum E_i^2$ in the present case, $E_i = X_i - X_i^c$.

significance levels. If the value of t derived from (9) is larger than the tabled value of t at a certain level of significance corresponding to the degrees of freedom we have, then b_i is significantly different from zero at that significance level. If b_i is significantly different from zero, then the confidence limits can be ascertained by

(11) $$b_i + t_{\alpha} \text{ Cii.s}$$

where α in t_{α} denotes the degree of confidence limit required. For example, for 95% confidence limit α = .05. This means that in 95% cases the value of b_i will lie between the limits given by (11).

We now denote what is known as the coefficient of multiple correlation as

(12) $$R^2 = \frac{\sum(X^c - \bar{X})^2}{\sum(X - \bar{X})} = \frac{b_u\sum xu + b_v\sum xv + b_w\sum xw}{\sum x^2}$$

R^2 measures the proportion of variability of the dependent variable X explained by the explanatory variables U, V and W in the regression.

V.

The brief discussion of the method of estimating the production function does imply that no statistical estimation can ever be perfect. However, there are some serious difficulties which must be faced before any reliability can be placed on the causal relationship between the independent variables and the dependent variables. Out of these, two can be briefly stated below.

The first one is multicollinearity which has been mentioned earlier. It is, in broad terms, the tendency of many economic series to move together in the same trend over time. It is an expression of common cause running through many economic variables. In cross-section data, the possibility of multicollinearity is small, but in time-series data, this may happen. In order to illustrate, let us consider the problem of estimating the production function.

(13) $$X = \beta_o + \beta_u U + \beta_v V + t$$

when U and V are perfectly correlated, we have

$$(14) \qquad r_{uv} = r_{vu} = 1$$

where r denotes the coefficient of correlation. The normal equations can be expressed in terms of the partial correlation coefficients

$$(15) \qquad \begin{aligned} \beta_u r_{uu} + \beta_v r_{uv} &= \beta r_{ux} \\ \beta_u r_{vu} + \beta_v r_{vv} &= \beta r_{vx} \end{aligned}$$

In (15) if $r_{uu} = r_{vv} = 1 = r_{uv} = r_{vu}$, then the two normal equations are identical. In other words, we have one equation to determine two unknowns β_u and β_v; hence the values of b_u and b_v are indeterminate. It can also happen that U and V may be perfectly correlated, but due to errors in measurement or observation of U and V, the partial correlation between the observed values is different from 1, in that case the two equations in (15) will not be identical and we can get the values of b_u and b_v. But these values will, however, be meaningless as they are based on not the accuracy of data but its errors.

One way of getting over multicollinearity is to drop, one of the two variables which are highly correlated from the estimation of the production function.

Another flaw in the data to be used may be autocorrelation. It means correlation between successive items in time-series of observations. Autocorrelation in the observation of one or two variables does not invalidate the estimates of the regression coefficient obtained via least squares, but if a variable U is autocorrelated, the variance of its regression coefficient b_u will be affected and hence the confidence limites and significance tests cannot be directly applied. Some more complicated methods will be needed.[1)]

1) H. Wold, Demand Analysis, John Wiley & Sons, 1953, p. 44 and 209-12.

References:

L.R. Klein , An Introduction to Econometrics, (Prentice-Hall) 1962, Chap. 3.

E.O. Heady & J.L. Dillon , Agricultural Production Functions, (Iowa) 1961, Chaps. 4 & 5.

P.H. Karmel , Applied Statistics for Economists, (Pitman) 1957, Chap. IX.

E.S. Keeping , Introduction to Statistical Inference, (Norstad) 1962, Chap. 12.

oo 0 oo

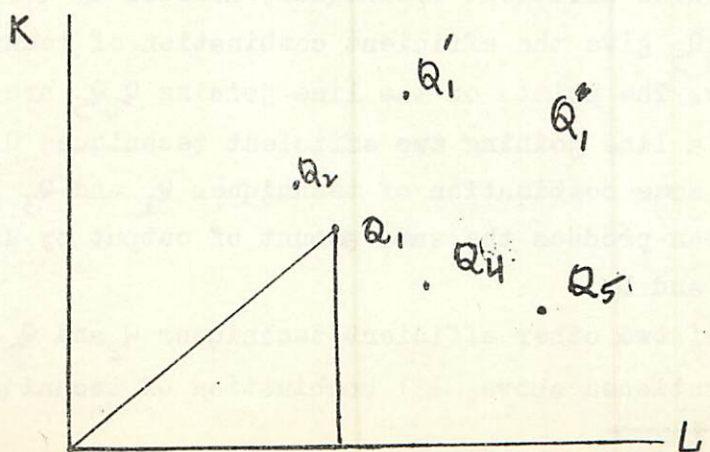
Lecture 6.

Representation of Techniques of Production.

I.

Technique means 'manner of artistic execution, the part of artistic work that is reducible to formula, mechanical skill in art'.[1)] In economics the word technique does not connote the artistic aspect as much as the mechanical one. In fact, technique of production is related to the method of production giving the relative amounts of different factors needed for producing a certain commodity. Strictly speaking, by technique of production in economics is meant the proportion in which different factors and inputs are combined to produce an output. To take a simple example, if there are two factors of production L and K and there is only one technique of production which can produce an output X by using these factors in a certain proportion, it can be represented by a point Q_1 in the two dimensional plane as in Fig.1. Thus, if there is only one technique of producing the commodity X, and Q_1 is the point which represents this, it means that labour and capital L and K can be combined in only one

Fig.1.



---

proportion in producing X given by $\frac{OM_1}{M_1Q_1}$ or the cotangent of the angle Q_1OM_1. It may be pointed out, however, there may be abailable a number (or an infinite number) of techniques which can produce the same quantity of X but using one or both the factors in greater quantities than is required by technique represented by Q_1 such as Q_1' or Q_1'', etc. But these techniques will not be adopted by any sensible manager of a firm. These techniques are really inferior techniques. For more of X can be produced by adopting technique Q_1 and using the amounts of factors needed by Q_1' and Q_1'', etc. to produce the initial amount of X. These techniques have, therefore, not to be considered since they are inefficient, even though, they might combine the two factors in the same proportion as Q_1 probably does.

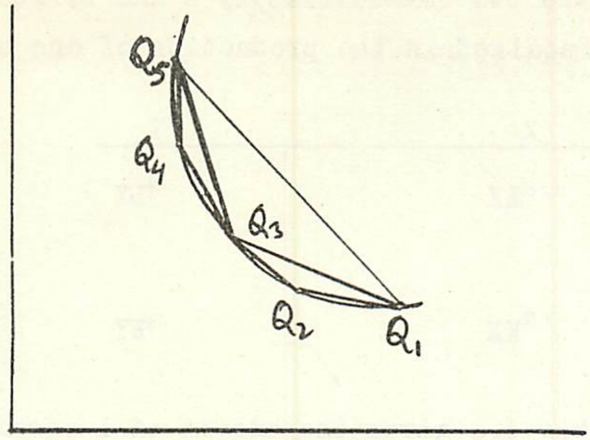
But there may be some other techniques which are not inferior or inefficient but efficient in the sense that though they may require one of the factors in greater quantity, they require the other in a smaller quantity. For example techniques Q_2, Q_3, Q_4 etc. are such that they require the two factors in different proportions in such a way that if the amount of one factor is greater than that in case of Q_1, the amount of the other is lesser. All these techniques are non-inferior and efficient.

Let us consider techniques Q_1 and Q_3 in Fig.2 for a moment. If the factors are perfectly divisible, i.e. they can be partly used in techniques Q_1 and partly in Q_3, irrespective of the proportion going to Q_1 or Q_3, then it can be easily seen that the amount of X can be produced by any combination of techniques Q_1 and Q_3, and these really infinite number of combinations of the two techniques are given by the points on the line joining Q_1 and Q_3. In Fig.2 we can begin with three efficient techniques, denoted by Q_1, Q_3 and Q_5. The lines joining Q_1Q_3 and Q_3Q_5 give the efficient combination of techniques Q_1 and Q_3 and Q_3 and Q_5 respectively. The points on the line joining Q_1Q_5 are not efficient, though they are on a line joining two efficient techniques Q_1 and Q_5, for the simple reason that some combination of techniques Q_1 and Q_3 lying on Q_1Q_3 and that of Q_3 and Q_5 can produce the same amount of output by using less of one or both the factors K and L.

Now if two other efficient techniques Q_2 and Q_4 are discovered[1)] then by the fact mentioned above, all combination of techniques Q_1 and Q_3

1) Some such techniques may be discovered that make the existing efficient techniques inefficient. Then the older techniques will have to be dropped out of consideration.

Fig. 2.



lying on Q_1Q_3 will become inefficient, and they will be replaced by the combinations of Q_1 and Q_2 and those of Q_2 and Q_4 lying on Q_1Q_2 and Q_2Q_3. The same will happen in case of Q_4.

If there exist or we discover a sufficiently larger number of efficient techniques Q_1, Q_2, Q_3 ...., combining the two factors L and K in varying proportions, we shall obtain a curve joining these points denoting efficient techniques. This curve actually traces a production function.[1] An alternative definition of a production function can be given as a function which represents the locus of the points denoting efficient techniques in a systematic way. We add the words 'in a systematic way' to indicate the fact that if a production function is viewed in a way described above, it will be obvious that it cannot but be convex to the origin.

Cobb-Douglas and SMAC function represent an infinite number of techniques that can combine K and L in any imaginable proportion. The capital output ratio by itself does not specify any particular ratio between K and L. A fixed proportion factor combination of the Leontief type gives only one technique denoted by only one point in Figs.1 and 2.

In the above, we have represented the techniques by taking example of a given amount of commodity to be produced. We can represent techniques in

1) If the number of efficient techniques is infinite, we will get a smooth curve representing a continuous production function.

another way by taking the case of two factors[1] as given and the production of two commodities whose amounts are not given. Let the amounts of factors be denoted by L and K and the two commodities by X and Y. The following table gives the amounts of K and L required in the production of one unit of each of X and K.

	X	Y
L	a_{LX}	a_{LY}
K	a_{KX}	a_{KY}

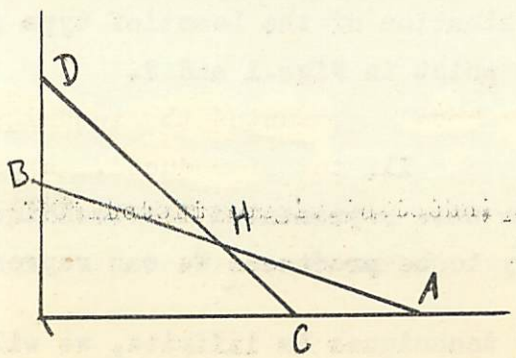
In the table a_{LX} gives the amount of L required in the production of one unit of X and similarly for other a's. Thus the elements in either of the columns in the table give the amounts of factors needed in producing one unit of the repsective commodities, thus they represent a technique because these elements give the proportion in which the factors will be combined to produce the output.

The quantities X and Y of the two commodities producible by using L in full have to satisfy the condition.

(1) $$a_{LX}X + a_{LY}Y = L.$$

By tracing the straight line given by (1) on a two dimensional plane with coordinates X and Y as in Fig.3, we get all the attainable points representing various combinations of X and Y. The triangle formed by joining the points AB and the axes contains all the points giving the amounts of X and Y that can be produced by all or less of L, while the points lying on the hypothenuse maximum combination of X and Y.

Fig.3.



1) We must have at least two factors to give a technique which may be of some

Similarly the triangle formed by the axes and the straight line CD traced by

(2) $a_{KX}X + a_{KY}Y = K$

contains all the points that are feasible with the given amount K, and the points lying on CD give the maximum combinations of X and Y producible with K. As can be seen the two factors are both fully utilized at point H; the coordinates of H give the maximum quantities of X and Y produced by using L and K fully, if there is only one technique to produce X and only one for Y given by the columns of the table above.

Let us now consider the case of more than one technique to produce the commodities. Obviously if there is a technique which requires more of the two factors to produce a given quantity of a commodity than that required by the original technique, it is inferior or inefficient. Reversely, if a combination of techniques to produce X and Y are such that the given amount of a factor, say L, can produce one or both of X and Y in lesser quantities, this combination will certainly be inferior. In terms of Fig.4, the line associated with the combination of the inferior technique corresponding to L will lie below AB all along without intersecting it. In the figure A'B' combines inferior techniques. But if two alternative techniques to produce X and Y respectively are discovered or introduced, such that by combining them more of both the commodities can be produced with given amount L, then these two techniques become efficient making the earlier or the original techniques inefficient, as far as the utilisation of L is concerned. In the figure C"D" combines superior techniques in respect of L.

If a combination of techniques is inferior in respect of the utilisation of one factor, it is not necessary that it should be inferior in respect of the utilisation of the other factor K, it can be even superior. But however superior it may be in respect of one factor, if it is inferior in respect of the other, the utilisation of the given amounts of the two factors L and K will yield lower amounts of the two commodities X and Y, if the combination of techniques is such that it is inferior in respect of, say, L and superior in respect of, say, K. This is because the factor in respect of which it is inferior becomes a bottleneck. This can be seen from Fig.4 where AB and CD represent the full utilisation curves related to L and K respectively when the original techniques are combined. A'B' and C"D" represent the corresponding curves when an alternative set of techniques are introduced. A'B' represents an inferior combination in respect of L, but C"D" represents a superior one. However, whatever the positions of A'B' and C"D", so long as A'B' is inferior in respect of L and C"D" is superior in respect of K, the feasible amounts of X and Y that can be produced with given L and K is given by the points which must be contained in the triangle formed by A'B' and the axes. Hence an alternative set of techniques can be efficient only if their combination in respect of any of the factors is not inferior in the manner described here.

Now consider combinations of techniques in respect of L. If the lines representing them intersect each other, none of them will be inferior, they will all be efficient, since if one combination can produce less of X, it can produce more of Y and vice versa. In Fig.4 a number of lines representing the combination of efficient known or available techniques in respect of L and K for producing X and Y are drawn. AB's for L and CD's for K.

Fig.4.

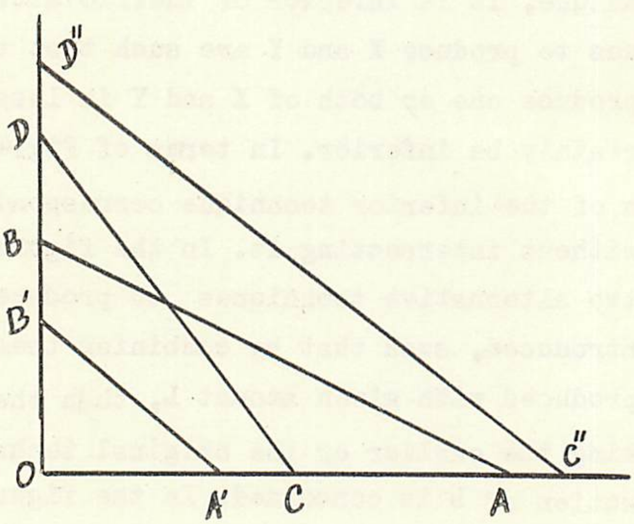
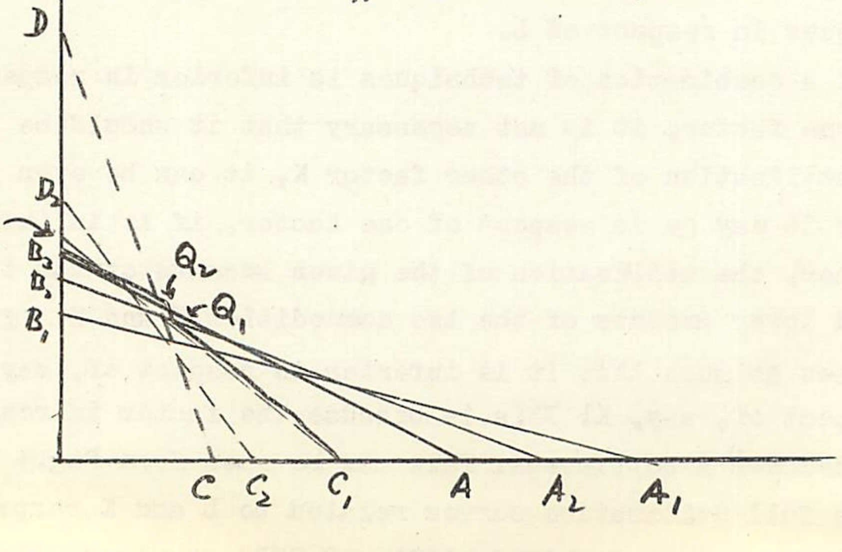
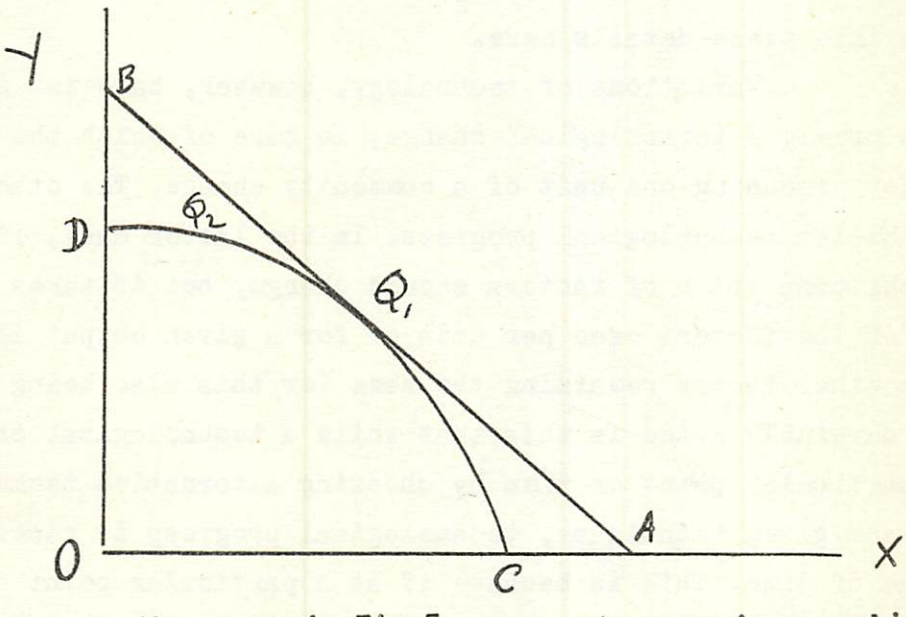


Fig.5.



The area covered by the polygon $OC_1\ Q_1\ Q_2\ B$ gives the feasible region so that the quantities of X and Y denoted by each point of this region can be produced by the given amounts L and K and the available efficient techniques for producing the two commodities X and Y. The thick line $C_1\ Q_1\ Q_2\ B$ is concave in the neighbourhood of Q_1 and convex in the neighbourhood of Q_2 to the origin. But if we have a sufficiently large number of alternative efficient techniques in respect of each of the factors for producing the two commodities, we can visualize that the line $C_1\ Q_1\ Q_2\ B$ will assume the shape of the smooth curve shown in Fig.5, concave to the origin all along its course. This curve has been called by different names such as, production possibilities curve, production frontier curve, opportunity cost curve, product transformation curve, etc. In fact, it is the same curve which we met towards the end of the first lecture.

Fig.6.



Each point on the curve in Fig.5 represents a maximum combination of the commodities X and Y which can be produced with given amounts L and K.

III.

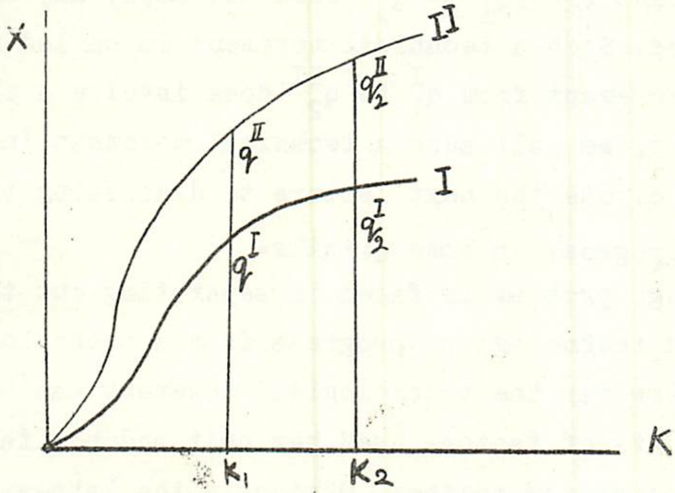
In recent years, in the context of the developing countries the choice of techniques, or which is the same things the proportion of factors to be used in production, has assumed great importance. In the literature a number of terms have been coined to describe a certain type of techniques. For example terms like capital light or capital heavy, capital shallowing or capital deepening, labour intensive or capital intensive, labour biased or capital biased, capital saving or labour saving, capital economizing or labour economizing, higher labour-capital ratio or lower labour-capital ratio and so on, have become common currency. And all these terms are quite interchangable. In the first case they mean relatively more of labour or less of capital than the existing techniques and in the other case the reverse.

There has been an endless discussion as to whether in a labour surplus and dapital scarce country, the techniques to be used should be labour intensive or capital intensive and all these discussions have been inconclusive, primarily because the writers or the participants in the discussion are not given, and they fail to get, a set pattern of consumption over time, so that results of the discussions are biased in favour of the former type of techniques if the present consumption is relatively given more emphasis than that in the future and latter type of techniques are favoured in the opposite case when the consumptions in the future are given relatively more importance. But we shall not go into these details here.

Variations of technology, however, have two important aspects. One is merely a technological change, in case of which the proportion of factors for producing one unit of a commodity change. The other is what is generally called technological progress. In the latter case, it is not necessary that the proportion of factors should change, but it takes place when one (or both) of the factors used per unit or for a given output is reduced, the amount of the other factor remaining the same (or this also being reduced). One thing to be carefully noted is this that while a technological change can be effected at a particular point of time by choosing alternative techniques out of several known and given techniques, technological progress is necessarily related to passage of time. This is because if at a particular point of time such a technique exists which uses one or both factors in smaller quantities, amount of the other factor being the same or also being reduced, all other techniques will become inferior. Further a technilogical progress may or may not involve a technological change. If the relative amounts of the factors remain the same after a technological progress, technological change does not take place but if the ratio of factors differs, then a technological change is also involved in technological progress.

The difference between a technological change and progress can be expressed in terms of isoquants and transformation curves. A technical change is said to take place when a certain quantity of output is produced by different combinations of factors, but moving along the same isoquant. A technical progress is realised when the same amount of factors can produce more output, i.e., when the isoquant curve moves out towards the north-east direction. A clearer illustration can be given by making use of the total product curve of one of the factors, say K, the amounts of other factors held constant.[1)]

Fig.7.



In Fig.7, curve I represents the initial TP with respect to K for a given amount of L, and curve II represents the same after the technical progress has taken place. Now if we change the amount of K with that of L remaining fixed, the proportion of factor changes. As K increases we can see total product increases, but if it remains on curve I all through, this involves just a series of technical changes. For if we have an infinite number of (non-inferior) techniques, we can increase the output by increasing K, and keeping L fixed up to a maximum limit.

In a technical progress, however, the output increases, even when the amount of K remains the same and the amount of L being held constant as before. This is shown in Fig.7 by the points q_1^{I} and q_1^{II} which correspond to K_1 of the variable factor. It should be noted that this increase in the output has not involved any change in the ratio of factors used as the amounts of the two remain constant. Consider another amount K_2 of K which is combined with the same fixed amount of L. In the absence of technical progress, the output of X increases to q_2^{I}, but with technical progress, the output rises to q_2^{II}. Now the total increase in X can be divided in two parts

1) This illustration has become very common. For reference see A.E. Ott, Production Functions, Technical Progress and Economic Growth, International Economic Papers, No.11.

$$\overrightarrow{q_1^{I}\ q_2^{II}} = \overrightarrow{q_1^{I}\ q_2^{II}} + \overrightarrow{q_2^{I}\ q_2^{II}}$$

$$\text{or} = \overrightarrow{q_1^{I}\ q_1^{II}} + \overrightarrow{q_1^{II}\ q_2^{II}}$$

In the first case, we have first expressed the effect of technical change first and that of technical progress next, and in the second case the order has been reversed.

Movement from q_1^{I} to q_1^{II} does not imply any change in the proportion of factors used. Such a technical movement is called neutral technical progress. But the movement from q_1^{I} to q_2^{II} does involve a change in the proportion of factors also, we call such a technical movement 'non-neutral' technical progress. We shall devote the next lecture to discussing the neutral vs. non neutral technical progress in some details.

A tough problem is faced in separating out the effect of technological changes and technological progress from a technological movement. The problem is really how far the technological movement can be accounted by reduction in the amounts of factors used per unit and how far it is just a substitution of one factor for another. Obviously the latter is not a technological progress but merely a technological change. We shall devote a succeeding lecture to discuss the attempts that have been made in this connection, their shortcomings and further scope of research if any.

Yet another problem of mathematico-economic nature arises in relation to proper representation of the technological progress as such. In recent years some attempts have been made in finding a mathematical, what amounts to the same a logical, expression for the technological progress. We shall discuss these attempts in a separate lecture.

During the last three decades a tremendous amount of theoretical and empirical research has been done on discrete production functions. Mostly these researches have utilized and are based on linear models. Discrete production functions can be divided into two classes. Firstly, there are single-technique models; by this we mean that there is only one technique to produce a commodity or a group of commodities considered together. Leontief's input-output models fall typically in this class. The other class can be described to consist of alternative techniques model. In these models several or more than one techniques are included for producing each commodity or group of commodities. The approach which is employed in finding the optimal techniques for producing the given assortment of commodities is popularly known as linear programming. We shall briefly discuss the single and alternative-technique production functions in a later lecture.

References:

A.E. Ott , Production Functions, Technical Progress and Economic Growth, International Economic Papers, No.11.

R. Dorfman ,
P.A. Samuelson,
& R. Solow , Linear Programming and Economic Analysis, New York 1958, Chap. 11.

oo 0 oo

Lecture 7.

NEUTRAL VS. NON-NEUTRAL TECHNICAL PROGRESS.

I.

Technical progress, as we have defined earlier, must result either in an increase in output with the same amount of factors or in a decrease in all or at least one of the factors, the others remaining the same for producing the same amount of output. Now if we define technical progress in the first way, we can easily see that it must maintain the proportion of factors used in the production as before. Neutral technical progress, as the terms signify, implies a technical progress which maintains the relative importance of factors by using them in the same proportion as before. Technical progress defined in the first way is, therefore, necessarily neutral. Technical progress defined in the second way may or may not be neutral; if it reduces all the inputs used in the same proportion, it is neutral, otherwise not.

Technical progress as defined above in the alternative ways is quite unambiguous. Technical progress, however, can take place in a less clear manner. Taking the case of two factors, labour and capital, it is just possible that a technical progress involves an increase in the amount of labour (capital) but a decrease in the amount of capital (labour) for the production of the same amount of output. In such a case we cannot say without ambiguity whether this is a case of technical progress or a technical change. However, if we further assume that prices of factors remain constant, we can give a clear cut definition of technical progress. Let $C_o = p_L L_o + p_K K_o$ be the initial cost of production where p_L and p_K are the constant prices of factors L and K, respectively. A technical progress is said to have taken place if the cost of producing the same amount of output at constant prices of factors, is changed to C_1 such that

$$(1) \qquad C_1 = p_l L_1 + p_K K_1 < C_o$$

(1) can happen in several ways, either L is reduced K remaining constant and vice-versa, or both L and K are reduced in the same or different proportions, or L is reduced but K is increased but relatively in a smaller degree, so that $C_1 < C_o$ and vice-versa. The technical progress will be neutral in one case only when $\frac{L_1}{K_1} = \frac{L_o}{K_o}$. In all other cases it will be non-neutral. If $\frac{L_1}{K_1} > \frac{L_o}{K_o}$ the technical progress is labour intensive, if $\frac{L_1}{K_1} < \frac{L_o}{K_o}$, it is labour saving.

II.

Technical progress can be studies in relation to changes in capital-output and labour-output ratios, by dividing the cost by the amount of output, P, so that

$$(2) \qquad \frac{C}{P} = \frac{P_L L + P_K K}{P} = P_L \frac{L}{P} + P_K \frac{K}{P} .$$

Given the prices of factors, technical progress will reduce $\frac{C}{P}$. This reduction may be due to a reduction in labour-output ratio, $\frac{L}{P}$, accompanied by a reduction in the capital-output ratio, $\frac{K}{P}$, or $\frac{K}{P}$ may remain constant or may even increase. In the last two cases, it is clear that the technical progress will be labour saving. In the first case, i.e., when the reduction in $\frac{L}{P}$ is accompanied by a reduction in $\frac{K}{P}$, we can say definitely that the technical progress is labour saving only when $\frac{L_1}{K_1} < \frac{L_o}{K_o}$, where L_1 and K_1 are the amounts of labour and capital after the technical progress has taken place and L_o and K_o the initial amounts to produce the same amount of output. Or, to say the same when

$$(3) \quad \begin{cases} \dfrac{L_1}{L_o} < \dfrac{K_1}{K_o} \quad \text{or} \quad \dfrac{L_1 - L_o}{L_o} < \dfrac{K_1 - K_o}{K_o} \\[2ex] \text{or} \quad \dfrac{-\Delta L}{L_o} < \dfrac{-\Delta K}{K_o} \quad \text{or} \quad \dfrac{\Delta L}{\Delta K} > \dfrac{L_o}{K_o} \end{cases}$$

(3) means that when a technical progress is such that it leads to a reduction in both the labour output and capital output ratios, it will be labour saving only when the ratio of reduction in labour to that in capital is greater than the initial ratio between labour and capital.

It should be noted that technical progress necessarily implies that in (2) $P_L \Delta L + P_K \Delta K < 0$ for all values of ΔL and ΔK (positive or negative) given the values of P_L and P_K.

In the reverse case, a technical progress can take place when a reduction in capital-output ratio is accompanied by a reduction in labour-output ratio or by a constant or even increased labour-output ratio. In the case when both $\frac{K}{P}$ and $\frac{L}{P}$ are reduced, it will be labour saving if $\frac{\Delta L}{\Delta K} > \frac{L_o}{K_o}$ as stated above and capital saving if $\frac{\Delta L}{\Delta K} < \frac{L_o}{K_o}$.

A technical progress is neither labour saving nor capital saving but neutral if

$$\frac{\Delta L}{\Delta K} = \frac{L_o}{K_o}$$

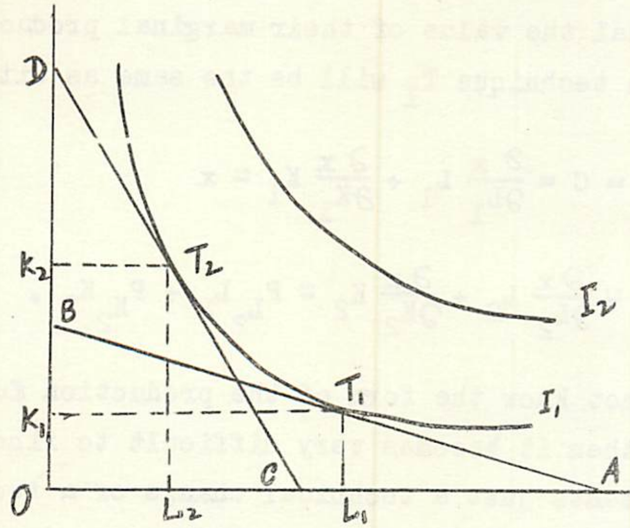
that is when the ratio of the changes in labour and capital after the technical progress is equal to the ratio between labour and capital used initially to produce the same amount of output.

III.

The discussion of the problem becomes more interesting and at the same time more complicated, as it should be, when we relax the assumption of constant factor prices. Then there will be two aspects to consider: 1) Whether the change in the relative amounts of factors represents a technical progress or just a technical change, and 2) in case it represents a technical progress whether it is neutral or non-neutral. As we have stated earlier, a technical movement can be called a technical progress only when the cost of production for a given amount of output or per unit is reduced. Now if the cost has been reduced just because the price of one or both the factors has been reduced for some reason or the other, this will be neither a technical progress nor a technical change. The situation relevant to our discussion is that when factor prices influence and in turn are influenced by the relative amounts of factors used in a systematic way, that is, when factor utilisation and factor prices are interdependent. We assume that a competitive equilibrium prevails that the ratio of the marginal productivity of factors tends to equal the ratio of their prices.

What is meant by a technical movement which is just a technical change that is bereft of technical progress is easily demonstrated by the following diagram;

Fig.1.



In Fig.1, I represents the isoquant for producing a certain given amount of output x. Initially the ratio of the factor prices is given by the slope of AB, so that

$$(4) \qquad \tan \quad a = \frac{P_{L_1}}{P_{K_1}} = \frac{\partial x}{\partial L_1} \Big/ \frac{\partial x}{\partial K_1} .$$

After a technical movement, the ratio of factor prices and the amount of factors used is changed. The ratio of factor prices is now given by the slope of CD, so that

$$(5) \qquad \tan \quad c = \frac{P_{L_2}}{P_{K_2}} = \frac{\partial x}{\partial L_2} \Big/ \frac{\partial x}{\partial K_2} .$$

In the initial position L_1 and K_1 amounts of factors were used for producing the amount of output and now L_2 and K_2 amounts are used to produce the same amount; $L_2 < L_1$ but $K_2 > K_1$ and so $\frac{L_2}{K_2} < \frac{L_1}{K_1}$. This is clearly a case of labour saving <u>technical change</u>. In the same way we can say that a transition from T_2 to T_1 will be a capital saving technical change.

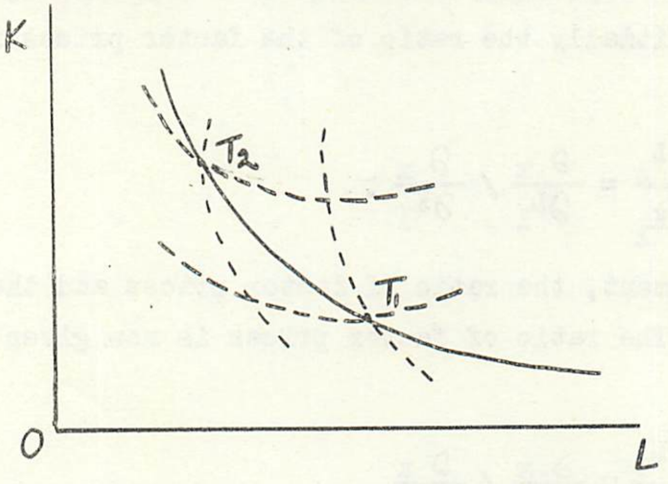
It has been easy for us to state that the transition from T_1 to T_2 and vice-versa is a case of technical change and not a technical progress simply because we had already assumed the shape of the curve and the form of the corresponding production function, and had further assumed that it does not change after the technical change. If we know the form of the function, we can easily derive the isoquants and find out whether the altered amounts of labour and capital lie on the same isoquant or not.

If we assume a linear homogeneous production function so that factor prices tend to equal the value of their marginal products, the cost of producing the output with technique T_1 will be the same as with technique T_2

$$(6) \quad \begin{cases} P_{L_1} L_1 + P_{K_1} K_1 = C = \dfrac{\partial x}{\partial L_1} L_1 + \dfrac{\partial x}{\partial K_1} K_1 = x \\ \qquad = \dfrac{\partial x}{\partial L_2} L_2 + \dfrac{\partial x}{\partial K_2} K_2 = P_{L_2} L_2 + P_{K_2} K_2 . \end{cases}$$

But if we do not know the form of the production function after the technical movement, then it becomes very difficult to find out whether the technical movement represents just a technical change or a technical progress. The situation can be explained by the following figure.

Fig.2.



In Fig.2, T_1 and T_2 represent two combinations of labour and capital. A technical movement is involved in a transition from T_1 to T_2 and vice-versa. But the technical movement is just a technical change, if the form of the production function is such that T_1 and T_2 lie on the same line. But if the production function is such that its isoquants are not represented by the continuous lines, but by the broken or the dotted lines, then T_1 and T_2 are combination of factors which give different levels of outputs. The picture is now quite complicated; T_1 and T_2 represent two different combination of factors and two different levels of output. In the absence of exact knowledge about the form of the production function it is not possible to state whether a transition from T_1 to T_2 represents a technical change or a progress (positive or negative).

IV.

We can to begin with suppose that the amounts of factors L_1 and K_1 remain unaltered and that a technical movement takes place. Let the original production function be $x = F(L, K)$ and the function after the technical movement be $x^* = F^*(L, K)$. Now the movement represents a technical progress when

$$(7) \qquad x^* = F^*(L_1, K_1) > x = F(L_1, K_1).$$

The above is very clear and straight simply because we have supposed the factor amounts to remain the same as did Hicks in 1932[1]. Hicks further stated that "labour-saving" inventions increase the marginal product of capital more than they increase the marginal product of labour; "capital saving" inventions increase the marginal product of labour more than that of capital; "neutral" inventions increase them in the same proportion[2]. If, as under assumptions of perfect competition, price of a factor equals its marginal productivity and the value output is imputed back to the factors, we have from (7)

$$(8) \qquad \begin{cases} \dfrac{\partial x^*}{L_1} L_1 + \dfrac{\partial x^*}{K_1} K_1 > \dfrac{\partial x}{\partial L_1} L_1 + \dfrac{\partial x}{\partial K_1} K_1 \\ \text{or} \left\{ \dfrac{\partial x^*}{\partial L_1} - \dfrac{\partial x}{\partial L_1} \right\} L_1 + \left\{ \dfrac{\partial x^*}{\partial K_1} - \dfrac{\partial x}{\partial K_1} \right\} K_1 > 0 \end{cases}$$

In (8), if both the brackets are positive, the inequality necessarily holds, but it is not necessary that they must be so.

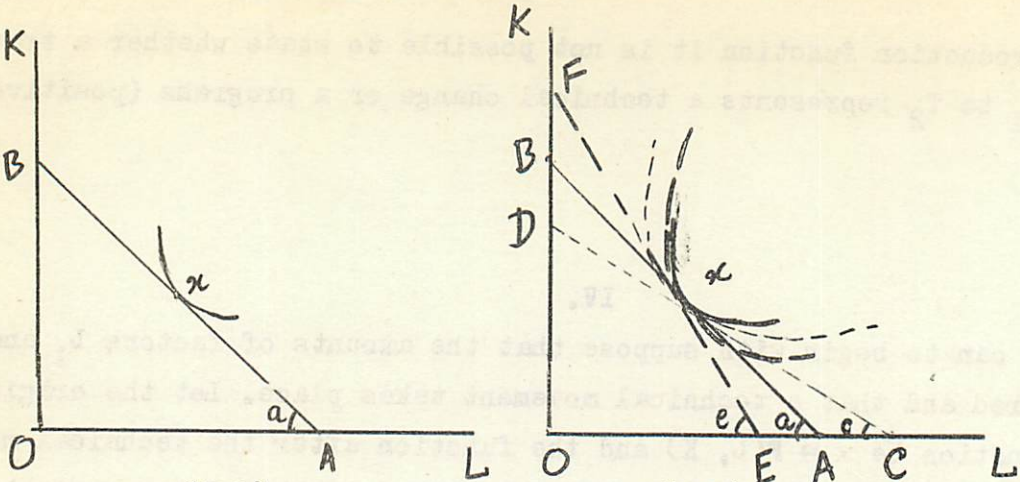
All that is needed for the technical progress to take place is that out of the two expressions, the positive should be greater than the negative. Now the Hicksian classification of technical progress can be easily demonstrated by the following figures.

---

1) J.R. Hicks, The Theory of Wages, (Macmillan) 1932, Second Edition 1963.

2) Ibid. p. 122.

Fig.3.



(3.1) (3.2)

In Fig.(3.1) L_1 and K_1 produce an output given by the level of the isoquant. The ratio of marginal productivity of labour and capital can be represented by the slope of AB, i.e. tan a. In Fig.(3.2), the same amounts L_1 and K_1 are shown to produce a higher amount of output. But the ratio of the marginal productivity of labour and capital may increase as in the case of the broken isoquant and the corresponding broken price line or may remain unchanged as in the case of the continued isoquant and the continued price line AB which is parallel to AB in Fig.(3.1), or it may decrease as in case of the dotted isoquant and the corresponding dotted price line CD. In the first case, according to Hicks, the progress is capital saving, in the second, neutral and in the third it is labour saving.

Two aspects of Hicksian definition are worth considering. The first is that, as the factor amounts and hence the factor ratios remain unchanged, it is only the expected changes in the factor ratios that are implied. Hence the technical progress is capital saving, neutral or labour saving only in the expected sense. Evidently, if the marginal productivity of one factor rises in comparison with the other, it will be expected to be utilised in greater proportion. Secondly, it has a bearing on the distributive shares of the factors. A capital saving technical progress will lead to a higher percentage of output produced to be distributed to labour and a labour saving to capital, while a neutral technical progress will maintain the distributive shares. In Hicks' own words "In every case, however, a labour-saving invention will diminish the relative share of labour. Exactly the same holds, mutatis mutandis, of a capital-saving invention."[1)]

1) Ibid. p. 122.

V.

We have seen that when we keep the factor prices constant and the factor amounts variable and also when we keep the factor amounts constant and factor prices variable, it is possible to state unequivocally whether technical progress has taken place or not. For in the first case, technical progress can be assessed by the reduction in the cost after the technical movement for producing the same volume of output, and in the latter case, the technical progress can be gauged by the increase in the volume of output. Things become shoddy, when both the factor prices and the factors amounts are changed. Since the factor prices are their own prices', i.e., equal to their marginal productivities, in full competitive equilibrium, the value of output will be equal to the cost of production. So with changes in the factor amounts; the volume and value of output will change, the cost being equal to the value of output, the reduction of cost criteria cannot be applied. Similarly when the factor amounts vary, the variations of the volume of output does not provide a ready guidance for the interpretation of the technical progress.

Let us consider two combinations of factors L_1, K_1 and L_2, K_2, denoted by points T_1 and T_2, so that $L_1 \gtrless L_2$ when $K_1 \lessgtr K_2$, respectively. Let F^1 be the initial production function and F^2 the production function after the technical movement, then it is not sufficient for F^2 to represent a technical progress that $F^2(L_2, K_2) > F^1(L_1, K_1)$. It is necessary that

$$F^2(L_2, K_2) > F^1(L_2, K_2)$$

and $$F^2(L_1, K_1) > F^1(L_1, K_1)$$

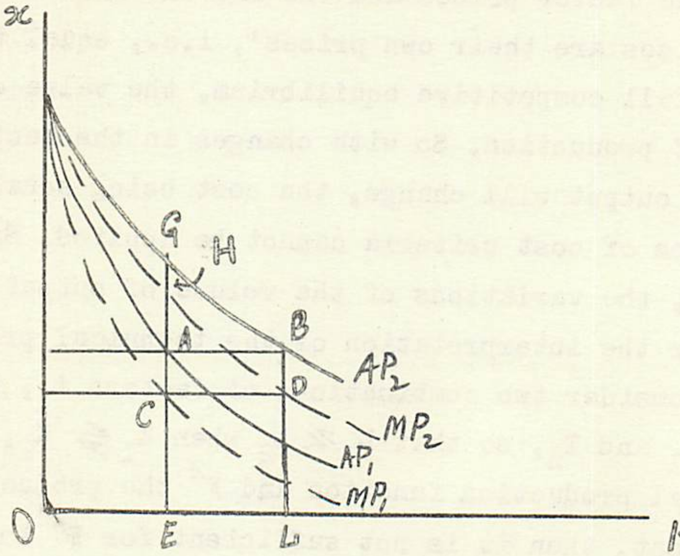
or more strictly, F^2 represents a technical progress when $F^2 > F^1$ for any combination of L and K. How should technical progress be represented in concrete functional form will be taken up in the next lecture.

The definition given above does not amount actually to more than what is generally known as the shifting of the production curve in case of technical progress as contrasted with the movement along the same curve in case of a technical change.

VI.

Harrod[1] has given an alternative definition of neutral technical progress and there has been a minor controversy whether Harrod's and Hicks' definitions are in essence the same or different. According to Harrod, technical progress is labour-saving if, at a constant rate of interest it raises the share of capital and lowers the share of wages in the total national product, neutral if it leaves the shares of both factors unchanged and capital saving, if it raises the share of labour and lowers the share of capital. Harrod's idea can be explained by a diagram introduced by Joan Robinson[2] for the same purpose.

Fig.4.



In Fig.4, AP_1 and AP_2 are average product curves with respect to capital corresponding to initial situation and that after the technical progress respectively with constant amount of labour and $\frac{\partial x}{\partial K_1}$ and $\frac{\partial x}{\partial K_2}$ are the corresponding marginal productivity curves of capital. In full equilibrium before the invention, the marginal product of capital, CE, is equal to the rate of interest, and amount of capital employed with the given amount of labour is OE. Total product is equal to OE · AE, which is divided between capital and labour in the ratio given by EC to CA.

Now let us suppose that when full-equilibrium is restored after the technical progress has taken place, the amount of capital employed with the given amount of labour is OL whereas the marginal productivity of capital is equal to the rate of interest, such that the average product of capital BL

1) R.F. Harrod, Review of Robinson's Essays in the Theory of Employment, Economic Journal, 1937.

2) J. Robinson, The Classification of Inventions, Review of Economic Studies, 1937-38; Reprinted in Readings in the Theory of Income Distribution, Ed. by W. Fellner and B.F. Haley, London 1950.

equals AE. In the first position, the ratio of the shares of labour and capital in the total product is AC to CE and in the second position after the technical progress it is BD to DL. Thus the relative share of capital is unchanged when the BD is equal to AC as it is in the figure. But the share of capital is increased or decreased by the technical progress according as BD is less or greater than AC. According to Harrod, if the share of capital remains the same at a constant rate of interest, the technical progress is neutral, and labour saving if it rises and capital saving if it falls.

Mrs. Robinson has shown[1)] that the elasticity of curve AP_1 at A is $\frac{AE}{AC}$ and the elasticity of curve AP_2 is $\frac{BL}{BD}$. In the figure AE = BL and CE = DL; therefore $\frac{AE}{AC} = \frac{BL}{BD}$. It means that when the technical progress is neutral in the sense of Harrod, the average product curve rises iso-elastically. A capital saving technical progress lowers the elasticity of average productivity curve, while a labour saving, or more properly, capital using technical progress raises the elasticity of the average product curve.

VII.

As stated above, according to Harrod a technical progress will be neutral if the average productivity curve of capital rises iso-elastically, so that, with a constant rate of interest, the relative share of capital in the total product is unchanged. But Mrs. Robinson shows that in Hicks' sense the progress must be neutral if the elasticity of substitution is unity, while if the elasticity of substitution is less or greater than unity; the invention must be capital saving or labour saving, to a corresponding extent in Hicks' sense.

This can be demonstrated in the following way. In Fig.4, GE is the average product of the original amount of capital, OE, with new technique and HE its marginal product. Then with the new improved techniques and the old amounts of factors, total product is GE x OE which is divided between capital and labour in the proportion EH to GH.

1) Economics of Inperfect Competition, p.36. This can be easily derived. For a given amount of labour, average productivity is $\frac{x}{K}$. The elasticity of $\frac{x}{K}$ with respect to K is equal to $\frac{K}{x} \cdot \frac{dx}{dK} - 1$. $\frac{K}{x} \cdot \frac{dx}{dK}$ is the ratio of marginal productivity, i.e. it is equal to $\frac{CE}{AE}$ or $\frac{DL}{BL}$. Hence the elasticity of curve AP_1 at A with respect to K is $-\frac{AC}{AE}$ and that of AP_2 at B is $-\frac{BD}{BL}$. (This is negative reciprocal of what Mrs. Robinson states and so it has to be checked).

Mrs. Robinson argues that if the elasticity of substitution is equal to unity over the relevant range[1), it follows that the ratio of the income of labour to the total product is independent of the amount of capital. Therefore, $\frac{GE}{GH} = \frac{BL}{BD}$. But $\frac{BL}{BD} = \frac{AE}{AC}$, therefore $\frac{GE}{GH} = \frac{AE}{AC}$. It follows that the marginal product of labour is raised by the invention (with a constant amount of capital) in the same proportion as total output and the invention is neutral in Hicks' sense. If the elasticity of substitution is less than unity, then $\frac{GE}{GH}$ is correspondingly greater than $\frac{AE}{AC}$ (as in Fig. 4), and the invention as labour saving in Hicks' sense, while if the elasticity of substitution is less than unity the invention is capital saving, to a corresponding entent in Hicks' sense.

Harrod's definition of neutral technical improvement or progress has led to a good deal of discussions, most of it quite subtle. While it is not possible to discuss the subtleties here, a few points that appear interesting may be stated.

The main point of discussion about Harrod's definition of neutral technical improvement has been due to the discrepancy that seems to arise with labour amount given and under the assumption of constant rate of interest, a technical progress[2) may lead to additions of capital stock, if it has to remain neutral. But if the ratio of capital to labour changes; then the generally accepted definition of neutrality ceases to hold. How can then Harrod give such a definition of neutrality? It has been shown, particularly by Kennedy[3), that according to Harrod's definition of neutrality, capital labour ratio does not necessarily change. Making use of a figure, originally given by Mrs. Robinson[4) which is similar to her earlier figure (Fig. 4 here), Kennedy shows that if the technical progress takes place in the consumption goods sector,

---

1) Elasticity of substitution is defined as $\frac{dr}{r} / \frac{d\frac{L}{K}}{L/K}$ where $r = \frac{dK}{dL}$. If σ is unity, $\frac{dr}{r} = \frac{d\,L/K}{L/K}$, i.e. the relative change in the ratio of the marginal productivities is equal to relative change in the ratio of factors. Hence, if factor price equals its marginal productivity, then the ratio of income of the two factors will not change, if $\sigma = 1$.

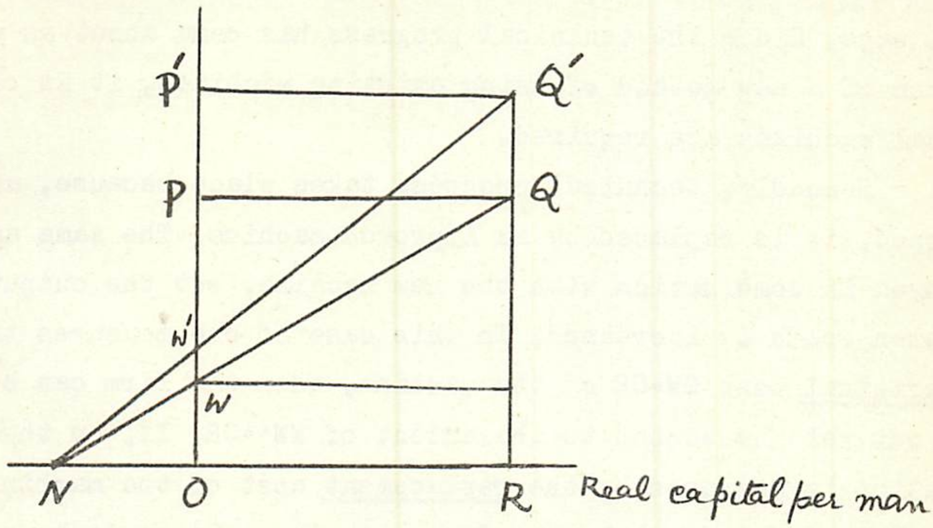
2) As defined by Harrod which leaves the share of output to labour and capital unchanged.

3) C.F. Kennedy, Technical Progress and Investment, The Economic Journal, 1961.

4) J. Robinson, Accumulation of Capital, (Macmillan) 1956.

the share of output to capital and labour may remain unchanged along with unchanged ratio of capital to labour (which here means without additions to capital) at a constant rate of interest. Fig.5 reproduces the figure given by Mrs. Robinson and used by Kennedy.

Fig.5.



In Fig.5, output per man is measured on the vertical axis and real capital (the value of capital in terms of wage units) per man on the horizontal axis. Suppose initially that the real wage is OW, and that the values of output per man and of real capital per man are given by the point Q. Then we have the value of capital per man in terms of goods equal to OW.° OR and the rate of profit on capital equal to $\frac{WP}{OW \cdot OR}$. Let us consider an increase in output per man. As under constant rate of interest, the condition of neutrality is that the capital output ratio is constant, it implies for any value of OP, $\frac{WP}{OW \cdot OR}$ and $\frac{OW \cdot OR}{OP}$ are constant since OP = OW + ØP, it follows that both OW/OP and OR are constant. Thus a neutral technical progress means a change from a point Q to a point Q' (vertically above Q), while the real wage rate rises in the same proportion as output per man.

It can be seen now that if capital is measured in wage units, then no increase in the value of capital is required by neutral technical progress, but if a goods standard is chosen, then the value of capital stock must rise in the same proportion as output per man.

Kennedy then studies three cases. In the first he assumes that all machines are everlasting and do not become obsolete. Technical progress takes place as a result of the invention of new methods of using the existing machines and not by the invention of new machines. As a result of the introduction of some new methods of production, the output per man in the consumption sector rises, but the condition of neutrality requires that the number of men employed with each machine remains constant. In Fig.5, the value of capital rises from OW·OR to OW'· OR, but this change is due entirely to the change in the real wage. Since the technical progress has come about as a result of the adoptation of a new method of using existing machines, it is clear that no additional machines are required.

Secondly, technical progress takes place because, as each machine is scrapped, it is replaced by an improved machine. The same amount of labour is employed in combination with the new machine, but the output per man of the consumption goods is increased. In this case if one measures the depreciation by the original cost OW·OR of the machine, then the firm can be said to have carried out net investment to the entent of WW'·OR. If, on the other hand, the depreciation is measured by the replacement cost of the machine, then no net investment is carried out by the firm when it replaces it by an improved machine. For as a result of the rise in the real wage, the replacement cost of the old machine has risen to OW'·OR, even though the machine is in fact replaced by an improved machine also costing OW'·OR.

Thirdly, technical progress will speed up the rate of replacement of machines as a result of more rapid obsolescence. In this case the situation is similar to the second case.

Harrod[1] himself has tried to defend or, as he would like to put it, to explain his definition of neutrality.'He suggests that, for this purpose and in this context, in order to obtain the required symmetry and conceptual tidiness, we should measure the quantity of capital in different ways. The quantity of capital should be measured by the average time of waiting multiplied by the number of man-hours (or other non-capital factors of production as currently valued in terms of man-hours) in respect of which there is waiting. Then, if the average length of the productive process is unchanged, however much the non-capital factors, in respect of which there is waiting, rise in goods value. Also with the rate of interest unchanged, the share of capital will remain unchanged.'

---

1) The Neutrality of Improvements, Economic Journal, 1961.

It seems that this ingeneous definition implies that the amount of labour remains constant, though it has not been stated explicitly by Harrod in the context of his re-explanation of his definition of neutral technical progress. Once this is explicitly taken into account, Harrod's explanation becomes round-about, it assumed what it purports to explain. Naturally, if the man-hours are given, the sum of the man-hours already used for the production of goods and the rest unused will be equal to the total man-hours given, and if the average length of the productive process is given, then by the definition of Harrod, the amount of capital will be constant, because this is equal to the man-hours multiplied by the average waiting time which is in turn equal to the average length of productive process.

References:

A.E. Ott , "Production Functions, Technical Progress and Economic Growth", International Economic Papers, No.11.

J. Robinson , "The Classification of Inventions", Review of Economic Studies, 1937-38.

C.F. Kennedy, "Technical Progress and Investment", The Economic Journal, 1961.

oo 0 oo

Lecture 8.

ECONOMIC REPRESENTATION OF TECHNICAL PROGRESS.

I.

We have already defined or explained technical progress in the last lecture. As can be noted it is not easy always to disentangle technical progress from mere technical change. It is all the more difficult to find a quantitative expression for technical progress which can be meaningfully used in economic analysis and growth theory. Yet a growth model or a growth theory will be incomplete, if it does not take into account technical progress.

All that we have been able to say in the last lecture is to make the facile statement that a technical progress can be said to have certainly taken place if the amount of output from some amounts of inputs is higher than that was obtained from the same amounts of inputs whatever the proportion of factors may be. This means that if the original production function is F and the function after the technical progress is denoted by F^{*}, then

(1) $$F^{*}(L,K) > F(L,K)$$

for all values of L and K. Apparently (1) is a very trivial statement. But some recent attempts to find quantitive expressions for technical progress are essentially based on this.

One way how to assure that $F^{*} > F$ for all values of the factors of production is to define

(2) $$F^{*} = \gamma F$$

where γ is more than unity for a positive technical progress.[1] It can be less than unity if we have negative technical progress, which case, of course, is of no practical value. Technical progress, as it can be easily conjectured, takes time to come by. So γ may be considered as a function of time. What will be the shape of this function is a tricky problem. Obviously technical progress has no causal relationship with passage of time as such. Hence to express technical progress coefficient γ, as a function of time is not satisfactory. That technical progress is correlated, if not causally related, with time has been implied both by Tinbergen and Solow.

Finding several faults with the above approach, to be discussed below, Kaldor[2] argues that technical progress necessarily involves capital accumulation. So he represents technical progress as a function of capital per head.

Another method of expressing technical progress has been recently suggested by Arrow.[3] This is that technical progress is really a process of learning. There is more to be said for this method, but as we shall see the way Arrow expresses learning is not very convincing and the process of learning itself is not the whole or the major element in technical progress.

II.

In this section, we shall discuss Solow's[4] approach of representing technical progress as a function of time. Solow starts with a production function

$$(3) \qquad P = F(K, L, t)$$

---

1) This way of expressing technical progress was first adopted by J. Tinbergen: "Zur Theorie der langfristigen wirtschaftsentwicklung", Weltwirtschaftliches Archiv, 1942.

2) N. Kaldor, "Capital Accumulation and Economic Growth", in The Theory of Capital, ed. by F.A. Lutz and D.C. Hague, (Macmillan) 1961.

3) K.J. Arrow, "The Economic Implications of Learning by Doing", The Review of Economic Studies, June 1962.

4) R.M. Solow, "Technical Change and the Aggregate Production Function", Review of Economics and Statistics, 1957.

which includes the time variable t to allow for technical progress.' Neutral technical progress is defined as one that "leaves marginal rates of substitution untouched and simply increases the output obtainable from given inputs". Solow specifies the production function (3) as did Tinbergen by expressing it as

(4) $P = A(t)\, F(K,L)$

By assumung that F is a homogeneous function of degree one and that factor prices equal their respective marginal products, he obtains the simple result 1)

(5) $\bar{P} = \bar{A} + \alpha \bar{L} + \beta \bar{K}$

where $\bar{P}$, $\bar{A}$, $\bar{L}$, $\bar{K}$ are relative (percentage) rates of change of the respective variables per unit of time, and α and β are the ratios of the values of labour and capital inputs respectively to the value of output in the base period 2), $\alpha + \beta = 1$. Since $\bar{P}$, $\bar{L}$, $\bar{K}$, α and β can be derived empirically, $\bar{A}$, the rate of growth of the residual to be imputed to technical progress can be estimated. If $\bar{L} = 1.5\%$, $\bar{K} = 3\%$, $\alpha = .75$, $\beta = .25$, as suggested more or less by the aggregate American date, then 'in the absence of the residual $\bar{P}$ would be a weighted mean of $\bar{L}$ and $\bar{K}$, and since labour's weight is much larger than capital's, $\bar{Y}$ would be much closer to $\bar{L}$ than to $\bar{K}$ equalling 1.9%. Actually $\bar{P}$ approximated 3.5% and the difference between 3.5% and 1.9% yields an $\bar{A}$ of 1.6%.

As pointed out by Domar, the value of A, by its very nature is a residual, and its total value or even a substantial part of it can hardly be imputed to technical progress. In words of Domar, "It absorbs, like a sponge, all increases in output not accounted for by the growth of explicitely recognized inputs". Technical progress, if it has any meaning at all, should mean that the factors of production become more efficient in producing output, i.e. either they produce more of output of the initial quality with their amounts equal to the

---

1) cf. E.D. Domar: "On the Measurement of Technological Change", Economic Journal, 1961.

2) (5) can immediately be derived. Differentiate (4) with respect to t, to get

$$\frac{dP}{dt} = A\,\frac{F}{L}\cdot\frac{dL}{dt} + A\,\frac{F}{K}\cdot\frac{dK}{dt} + \frac{dA}{dt}\cdot F.$$

Dividing both sides by P and substituting $\alpha = A\,\frac{F}{L}\cdot\frac{L}{P}$ and $\beta = \frac{F}{K}\cdot\frac{K}{P}$ we get (5), (assuming that α and β are invariant with respect to changes in K and L).

initial amount, or they produce the same quantity of output but of better quality.[1] A technical progress if it has to be represented in a meaningful way economically, then it must express a causal connection between the inputs and the outputs. That is, it must purport to express the increase in efficiency of the factors of production, as far as it is possible. The heart of the matter is this that either labour or capital or both become more efficient in production. The increase in efficiency may be of two types, firstly one that is obtained from practice, workers working with the same machines or tools find quicker and easier ways of handling them and thus can produce more. The point is that no cost on training them has been incurred, and labour hours in terms of training cost remain the same and they produce more. Secondly, the same workers producing tools and machines find quicker and easier methods of producing more tools and machines. Labour hours in terms of training cost remain the same. The tools and machines also remain the same. Just through practice and experience production has increased, because of an all-round increase in efficiency of labour. This type of technical progress is the simplest and can be represented as a function of time, as practice does involve time. However, there does not seem to be an exact relationship between time and the amount of practice, the latter depends on the intensity with which the work has been performed during the time which has elapsed and also on the inherent and acquired capabilities of the workers. Granting that we can find a relationship between time and the improvement in the efficiency of labour all round through practice in the way described above, we can say that the production of all godds, capital and consumption, will increase without changing the capital-labour ratio. It does not really matter whether the efficiency of labour has increased more in the capital goods producing sector or in the consumption goods, so long as the number of workers per machine either producing machine or consumption goods is not altered, the capital labour ratio remains constant, though the output has increased per worker and per machine. In this simple example it seems that only the efficiency of labour has increased, but we can just as well say that the efficiency of machines has increased. It seems more proper in this connection to state that the efficiency of workers and machines has mutually and more or less equally increased. All this leads us to conclude that if the technical progress is of this simple type,

---

1) Quality is in reality a subjective evaluation, and depends on the preferences of the users of the commodity.

Tinbergen-Solow method of representing technical progress may be considered appropriate depending on the degree of success with which we can relate the extent of practice with the passage of time.

III.

Recently Arrow has tried to represent technical progress in terms of experience or practice. He takes cumulative gross investment as an index of experience. Each new machine produced and put into use is capable of changing the environment in which production takes place, so that learning is taking place with continually new stimuli. This at least makes plausible the possibility of continued learning in the sense, here, of a steady rate of growth in productivity. He further assumes that at any moment of time the new capital goods incorporate all the knowledge then available, but once built, their productive efficiency cannot be altered by subsequent learning.

To simplify the discussion, Arrow further assumes that the production process associated with any given new capital good is characterised by fixed coefficients, so that a fixed amount of labour is used and a fixed amount of output is obtained. Thus the type of technical progress discussed above is ruled out. It is reasonably assumed that new capital goods are better than old ones in the strong sense that, if we compare a unit of capital good produced at time t, with one produced at time $t_2 > t_1$, the first requires the cooperation of at least as much labour as the second, and produces no more product. A specific case of Arrow's model is given by him as follows.

Let G represent cumulative gross investment in period t and G' the cumulative gross investment in the period immediately before the time period in which the capital items created are to disappear in period t, assuming that capital goods have a fixed life-time and so they disappear in serial order. Arrow assumes capital output ratio constant, so that total output in period t is

(6) $$x = a(G-G')$$

where a is the constant output capital ratio. Labour employment is a decreasing function of cumulative gross investment, so that[1]

1) Total number of workers employed L is equal to the number of workers employed with capital of each vintage from the one that is to disappear in the current period up to the one t which is created in the current period. When a capital

(7) $$L = \int bG^{-n}\, dG - \int bG'^{-n}\, dG' \qquad \text{where } n > 0$$

$$= \frac{b}{1-n}\left(G^{1-n} - G'^{1-n}\right)$$

Substituting from (7) in (6) we get

(8) $$x = a\left[G - \left(\frac{L(1-n)}{b} - G^{1-n}\right)^{\frac{1}{1-n}}\right]$$

$$= aG\left[1 - \left(1 - \frac{1-n}{b} \cdot \frac{L}{G^{1-n}}\right)^{\frac{1}{1-n}}\right]$$

When n=1, (7) becomes

(7') $$L = b \log G - b \log G'$$

so that (8) can be written as follows

(8') $$x = aG(1 - e^{-\frac{L}{b}}) \quad \text{when } n = 1.$$

(8) and (8') are production functions showing increasing returns to scale in the variables G and L. It can be noted that x increases more than proportionately to scale changes in G and L which is also obvious intutively, since the additional amounts of L and G are used more efficiently than the earlier ones.

Arrow's point that it is 'gross rather than net investment which is the basic agent of technical change' will be readily agreed to.[1] This is because most of the funds set aside for depreciation and replacements etc. can be used for any type of machines and equipments as the net investments can be done.

As pointed out by Kaldor[2], experience or learning involves time. A meaningful technical progress function or learning function must relate the process of accumulation with passage of time, that is, it must not depend only on the magnitude of accumulated gross investment, but also on the speed of this accumulation. A priori, it is difficult to see whether the rate of learning or technical progress will be higher when the gross investment accumulates at a higher speed or a lower speed as distinct from its magnitude. It seems that when

---

1) cf. N. Kaldor, 'Comment', The Review of Economic Studies, June 1962.

2) N. Kaldor, op.cit.

a new technical process is discovered or invented, there may be some hesitation or delay in making heavy gross outlay in the lines of production, in which the progress has taken place by junking out the older capital stock. And it is only when the new processes have been proved more profitable than the existing ones, with an adequate degree of certainty that substantial amounts of gross outlays might be made. Confined to the trivial case of an increase in 'learning' with the existing types of machines, this seems to be more dependent on the passage of time rather than the magnitude of gross investment cumulated or not.

For Arrow's procedure to be satisfactory, it must also be necessary to show that the validity of the reverse tendency, i.e., the gross investment is constant when the technical progress or 'learning by experience' is zero. Since obviously this is not the case in reality, to relate practice or learning by experience to accumulated gross investment does not seem very reasonable.

IV.

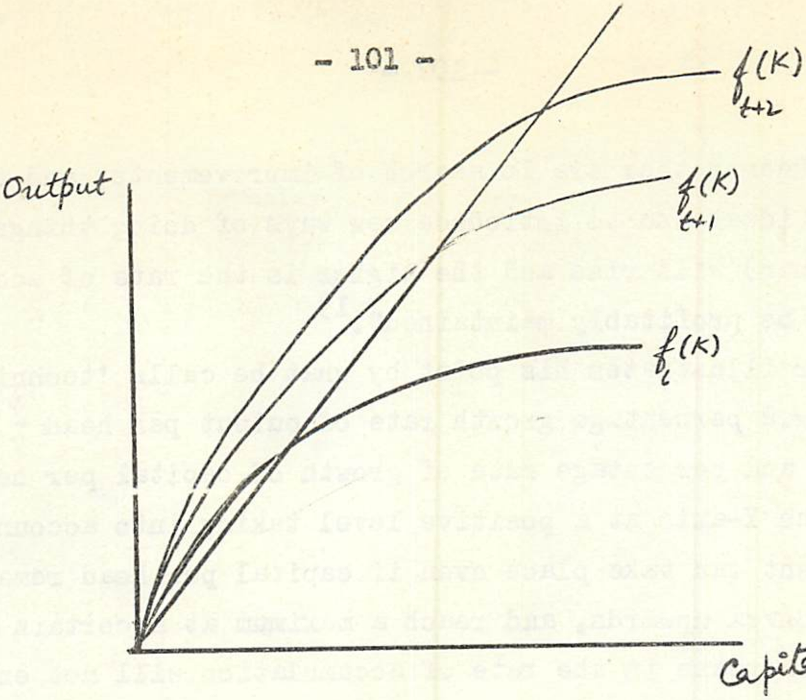
We have briefly discussed two ways of representing technical progress. In the first case it was related to time only and in the secodnd case to magnitudes of accumulated gross investment. Both methods are of course, far from satisfactory. One of the well-known sceptics about the former method and also a critic of the latter as we have noted above is Kaldor. We shall give here his own crude method of represeting the technical progress and his comments about the first method.[1)]

Kaldor takes exception to the neutrality of technical progress and opines that the latter can hardly be expected to follow a uniform course as implied by neutrality. For the assumption of neutral technical progress means that the production curve shifts in such a manner that the slopes of the tangents, of the functions f_t, f_{t+1}, f_{t+2}, etc., remain unchanged along any radius from origin (Fig.1.). The tangent of the angle between a ray from the origin and the output axis measures the capital output ratio and the slope of the production curves gives the marginal productivity of capital which under assumptions of competitive equilibrium is equal to the rate of profit or the share of capital. Thus in order that a constant rate of profit over time may be consistent with a constant capital output ratio and a constant rate of growth, the slope of each successive production curves at the points where the ray from the origin interests themmust be the same.

---

1) This is based on Kaldor's paper "A Capital Accumulation and Economic Growth",

Fig.1.

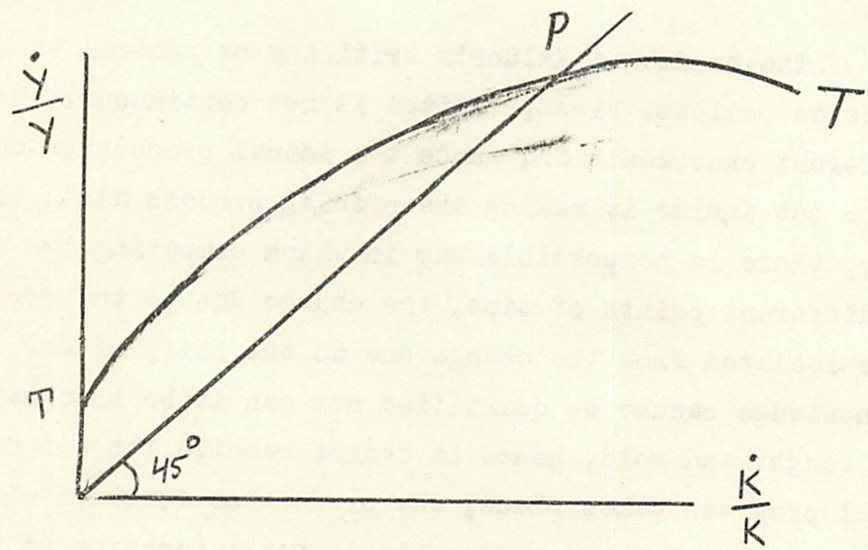


The burden of Kaldor's criticism of neutral technical progress a la Harrod is as follows. First, capital is not continuously adaptable in respect of its different endowments and hence the actual production does not move along the curve but inside it making the pricing process along the curve irrelevant. Secondly, there is no possible way in which comparing two different positions, at two different points of time, the change due to the movement along the curve could be isolated from the change due to the shift of the curve. Thirdly, technical knowledge cannot be quantified nor can it be brought under exclusive ownership or bought and sold, hence it cannot receive its own marginal product. When technical progress takes place, the production function will not be linear homogeneous in terms of labour and capital, but a function of higher order in these factors. In that case the share of profit must be necessarily less than the marginal product of capital, and there is no sure ground for relating rate of profit to a given capital output ratio on a technical ground. Fourthly, the rate of shift of the production function due to changing state of knowledge cannot be treated as an independent function of (chronological) time, but depends on the rate of accumulation of capital itself, i.e., speed of movement along the curve, making any attempt to isolate the shift of the curve from the movement along the curve even more non-sensical.

According to Kaldor, technical progress depends on two things, the rate of accumulation of capital and the society's capacity to absorbe technical change in a given period. "The more 'dynamic' are the people in control of production, the keener they are in search of improvements, and the readier they are to adopt new ideas and to introduce new ways of doing things, the faster production (per man) will rise and the higher is the rate of accumulation of capital that can be profitably maintained".[1)]

Kaldor illustrates his point by what he calls 'technical progress function'. In Fig.2 percentage growth rate of output per head $\frac{\dot{Y}}{y}$, has been plotted along Oy and percentage rate of growth of capital per hear along Ox. The curve cuts the Y-axis at a positive level taking into account that a certain rate of improvement can take place even if capital per head remained unchanged, but it will be convex upwards, and reach a maximum at a certain point beyond which a further increase in the rate of accumulation will not enhance further the rate of growth of output.

Fig.2.



It means that as the rate of accumulation increases, the rate of increase of output increases but at a diminishing rate, and reaches a maximum point depending on society's dynamism (inventiveness and readiness to change). Given the fact that the technical progress curve starts at a positive level and is convex, it must be intersected by a radius of 45°. In Fig.2 this is done at point P. At this point all the conditions of neutral technical progress are satisfied: the capital output ratio will remain constant at a constant rate of growth, constant distributive shares, and a constant rate of profit on capital.

Kaldor asserts that there is a tendency of the system to move towards a position where output and capital both grow at the same rate, and where therefore the rate of profit on capital will remain constant at a constant margin of profit on turnover. This he justifies by the suppositions (i) that the prospective rate of profit on existing capital will be higher than the currently realised rate of profit on existing capital whenever production is rising faster than the capital stock; (ii) that a rise in the prospective rate of profit causes an increase in the rate of invesment, relative to the requirements of a state of steady growth, and vice versa.[1]

In the model actually Kaldor uses a linear function 'for the sake of convenience' as he puts it. His technical progress function showing the relationship between the rate of growth of output per worker, G_o, and the rate of growth of capital per head, $G_k - \lambda$, (λ being the rate of growth of labour and G_k that of capital), is as follows,

$$G_o = \alpha' + \beta' (G_k - \lambda)$$

where $\alpha' > 0,\ 1 > B' > 0.$

Thus, in Kaldor's simple treatment of technical progress, the per capita output does not grow at the same rate as per capita capital ($G_k - \lambda$), but at a rate which is sum of a constant α' indicating the rate of growth of per capita output even when the per capita growth of capital is zero and a multiple of the rate of growth of per capita capital.

V.

It is obvious by now that disentangelment of technical progress from technical change is a complex and difficult problem, and a satisfactory method of representing technical progress has yet to be discovered. The difficulties are analogous to the concept and measurement of marginal productivity, only a bit more complicated. Just as in the case of the measurement of marginal productivity all factors units must be uniform, so it seems that we can make a headway in the study of technical change and progress when the units of factors used remain uniform and unchanged. Unfortunately, in the case of technical progress, though not in that of technical change, we can hardly visualize a situation where factor units can be regarded unchanged, when some technical progress have had taken place. For if the quality of a factor unit is judged by its productivity, no technical progress can take place without involving a change in the quality of one or more of the factors. And if the factors are not the same in

1) Kaldor, op.cit. p. 214.

the initial and the current situation, then a comparison whether of productivity or anything else is hardly meaningful.

However, an alternative approach, though not wholly flawless, can be adopted. It is that we do not consider the quality of the factors by their productivity but by the cost incurred on them. This is in itself not sufficient for factor costs vary as their productivity changes, given the pattern of (derived) demand for them. Factor costs are, therefore, dependent on their productivity and hence the former cannot be used in measuring the latter. However, the change in the productivity of a factor does not directly influence its own price excepting in so far it is used in producing itself, which can be neglected. Thus, a rough measure of the increase in factor productivity can be obtained by estimating the reduction in cost.

The problem becomes more tractable when we visualise an economy where the supply of the basic factor, i.e., unskilled labour is imfinitely elastic or its supply price remains constant. Furthermore, if we assume competitive equilibrium, so that prices of different factors such as different types of skilled labour and different assortment of capital goods equal their cost of production, then the reduction in the cost of production of certain commodity will certainly become an exact measure of technical progress. It seems that only in this ideally simple situation, an accurate measure of technical progress is possible. ~~situation, an accurate measure of technical progress is possible~~. The main difficulty regarding the study of the present problem or for the matter of any other economic problem is, that it is not easy to determine the supply price of unskilled labour and that whether it remains constant or not. In reality, perfectly competitive equilibrium also hardly exists. The problem of the fixation of the extent of technical progress, is, therefore, likely to remain uncertain and inexact.

References:

R.M. Solow, "Technical Change and the Aggregate Production Function", Review of Economics and Statistics, 1957.

N. Kaldor, "Capital Accumulation and Ecobomic Growth", in The Theory of Capital edited by F.A. Lutz and D.C. Hague, (Macmillan) 1961.

K.J. Arrow, "The Economic Implications of Learning by Doing", The Review of Economic Studies, June 1962.

Lecture 9.*

DISCRETE PRODUCTION FUNCTION.

I.

So far we have generally assumed that factors can be freely substituted for each other or the possibilities of substitution are infinite. As stated earlier, it is quite obvious that this assumption is not realistic. In the last three decades a great bulk of theoretical and empirical work has been done on the basis of discrete production functions. We shall confine ourselves to an elementary discussion of the original contributions in this field made by Walras, Leontief, and von Neumann[1] and briefly note their distinctive features.

It is convenient to give the Walras equation-system first to enable ourselves to study how the production function is treated.

Let us consider an economy with n commodities $x_j (j=1,...n)$ and m factors of production $q_i (i=1,...m)$. Let the technical production possibilities be characterised by mn fixed numbers a_{ij}, representing the physical amount of the i_{th} resource used up in the manufacture of a unit of the j_{th} commodity.[2] The market demand for commodities is expressed as

---

*) This lecture is based on Prof. Bent Hansen's "Lectures in Economic Theory", Part I.

1) L. Walras, Elements d'Economic Politique Pure, (1873), English translation 1954.
W.W. Leontief, The Structure of American Economy, 1919-39, Oxford 1941, Studies in the Structure of American Economy, Oxford, 1953.
J. von Neumann, A Model of General Economic Equilibrium, <u>The Review of Economic Studies</u>, 1945-46.

2) In some versions of Walras' model (cf. B. Hansen: Lectures in Economic Theory, 2nd revised edition, Part I and II) a_{ij}'s are assumed to be freely variable, dependent upon the price of factors and commodities. This case is not interesting to us in the present content, for it depicts infinite possibilities of substitution to our earlier discussions.

(1) $$x_j = F^j(p_1,\ldots,p_n;\ r_1,\ldots,r_m) \qquad j=1,\ldots,n,$$

where $p_j (j=1,\ldots,n)$ is the price of commodity j, and r_i $(i=1,\ldots,m)$ the price of factor i.

The cost of producing each commodity equals its price hence,

(2) $$p_j = \sum_{i=1}^{m} a_{ij}\, r_i \qquad j=1,\ldots,n.$$

Equilibrium requires that all factors of production are used up in production, no more, no less, so that

(3) $$q_i = \sum_{j=1}^{n} a_{ij}\, x_j \qquad i=1,\ldots,m.$$

The supply of factors is expressed as

(4) $$q_i = G^i(p_1,\ldots,p_n;\ r_1,\ldots,r_m) \qquad i=1,\ldots,m.$$

In (1) to (4) we have 2m + 2n equations in 2m + 2n unknown variables x_j's, P_j's, q_i's and r_i's. This is Walras' model. The main thing to see is how the production function enters the model. It is contained in (2) and (3), and is in fact described by the coefficients of technical production a_{ij}'s. We can express these in a tabular form:

(5) $$\begin{Bmatrix} a_{11} & \ldots & a_{12} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{i1} & \ldots & a_{i2} & \ldots & a_{in} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{m1} & \ldots & a_{m2} & \ldots & a_{mn} \end{Bmatrix}$$

(5) can be called a matrix of technology or a matrix of production possibility or just a discrete production function. It should be noticed that (5) gives only one way of producing each commodity[1] and that factors of production are treated quite separately from the commodities produced.

1) In a fully fledged model of Walras when a_{ij}'s are treated freely variable, there are introduced infinite ways of producing each commodity leading to the other extreme.

II.

Of the two characteristics of Walras' model, i.e. singleton techniques of production (or in the alternative case of an infinite number of techniques) for producing each commodity and the complete separation of factors or inputs from outputs. Leontief's model modifies and amplifies the situation as regards the latter. Leontief's matrix of technology is derived from his input-output table. It consists of a certain number of horizontal and the same number of vertical entrances, one for each industry or each group of industries flanked on the left and bottom side by entries for final demand and factor inputs. Entries on a row corresponding to a certain sector express the magnitudes or values of deliveries from that sector to other sectors, while entries on the column corresponding to that sector express the magnitude or values of receipts from other sectors to that sector. Thus the goods belonging to each sector are treated as both inputs and outputs. The same good is output of one industry and inputs to other industries or sectors. This approach is clearly closer to facts for in reality there is no water-tight deivsion between factors or inputs and outputs. The following is a sketch of Leontief's type input-output model.

<u>Table I.</u>

Receipts from other sectors	Deliveries to sectors						Final demand			Total output
	No 1	No 2	...	No j	...	No n	Consumption	Investment	Export	
No 1	x_{11}	x_{12}	...	x_{1j}	...	x_{1n}	D_{1C}	D_{1K}	D_{1X}	O_1
No 2	x_{21}	x_{22}	...	x_{2j}	...	x_{2n}	D_{2C}	D_{2K}	D_{2X}	O_2
:	...	...	...	...	...	...	...	...	...	...
No i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{in}	D_{iC}	D_{iK}	D_{iX}	O_i
:	...	...	...	...	...	...	...	...	...	...
No n	x_{n1}	x_{n2}	...	x_{nj}	...	x_{nn}	D_{nC}	D_{nK}	D_{nX}	O_n
Labour	q_{L1}	q_{L2}	...	q_{Lj}	...	q_{Ln}				
Capital	q_{K1}	q_{K2}	...	q_{Kj}	...	q_{Kn}				
Land	q_{A1}	q_{A2}	...	q_{Aj}	...	q_{An}				
Total inputs	I_1	I_2	...	I_j	...	I_n				

It is clear that total output of sector i

$$O_i = \sum_{j=1}^{n} x_{ij} + D_{iC} + D_{iK} + D_{iX} \qquad i=1,\ldots,n, \tag{6}$$

and total inputs in sector j

$$I_j = \sum_{i=1}^{n} x_{ij} + q_{Lj} + q_{Kj} + q_{Aj} \qquad j=1,\ldots,n. \tag{7}$$

It is obvious that inputs can be summed only when they are expressed in value terms. If we express the outputs also in value terms then (in equilibrium)

$$O_i = I_i \qquad i=1,\ldots,n. \tag{8}$$

Table I can be modified in various ways depending on the purpose for which it is constructed and on the availability of time and resources. If we divide each element in Table I with the total output of the corresponding sector, we obtain the elements of the matrix of technology

$$a_{ij} = \frac{x_{ij}}{O_i} \qquad i=1,\ldots,n \quad j=1,\ldots,n. \tag{9}$$

If we express all elements a_{ij}'s for all i's and j's in a tabular form, we get a square matrix of technology similar to (5).

The main contribution of Leontief's approach seems to be its pragmatism. It is no wonder that this approach has led to such a tremendous amount of empirical research. Walras' approach was based on abstract logic in the sense that the technical coefficients of production were either given or were to depend on supply, demand and cost functions and the equilibrium conditions. So long as the supply and demand functions were not derived empirically, there was no possibility of estimating the technical coefficients. The separation of factor inputs and outputs did not make it any ~~less difficult~~ for the Walras model to be empirically tested.

Leontief's model, while amenable to empirical studies, is simpler and cruder than Walras'. In Leontief's model, the role of prices remains suppressed; technical coefficients depict the state of production expost, making it less suitable for future use; Leontief's technical coefficients are really some sort of averages of the technical coefficients that hold in different firms using different techniques for producing the same commodity or group of commodities, and hence it is not possible to identify Leontief's technical coefficients with any actual coefficients that exist.

III.

As stated above, Leontief's technical coefficients are some sort of statistical catch-all and they may be heavily biassed in favour of obsolete techniques specially when a major portion of production is carried on with older capital stock rather than the newer ones. This and the fact that the model gives only one technique for producing a commodity or a group of commodities, it is futile to try to find out optimal techniques of production from Leontief's type of matrix of technology. In both Walras' and Leontief's treatment technology matrix once derived or arrived at is fixed and serves as datum. von Neumann's model removes this short-coming and incorporates alternative processes for producing commodities. The characteristic feature of von Neumann's model is that it proceeds on the basis of processes of production as contrasted with individual commodities (Walras) or sectors (Leontief). Each process, in theory, uses several or all goods and may produce several or even all goods. Thus von-Neumann's model allows the possibility of joint production. The following is the sketch of von Neumann's model.

There are n different goods and m technical processes, m > n. A process j working at unit level is described in terms of the inputs of goods required and the outputs of goods produced:

$$\text{inputs required at unit level of process } j \quad \begin{Bmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{Bmatrix} \longrightarrow \begin{Bmatrix} O_{1j} \\ \vdots \\ O_{nj} \end{Bmatrix} \quad \text{outputs forthcoming at unit activity level of activity } j$$

where x_i $(i=1,\ldots,n)$ and O_i $(i=1,\ldots,n)$ give the amounts of inputs and outputs respectively, some x's and some O's may be zero, but if some $x_{ij} = 0$ then $O_{ij} > 0$, and vice versa. This is needed to hold the system together.

As is the case with Walras' and Leontief's models, constant returns to scale is implied in von Neumann's model too. If the activity j $(j=1,\ldots,m)$ is worked at level, say, a_j, then the inputs will be $x_{1j}a_j \ldots x_{nj}a_j$ and the outputs will be $O_{1j}a_j \ldots O_{nj}a_j$.

The two major contributions that follow from representing the production function in the way von Neumann did is that a choice of processes is possible and thus an optimal set of processes may be attained, and that it leads to dynamic analysis. von Neumann also proved the existence of the system, which cannot be sketched here as it involves advanced mathematics.

Production will be possible only when

$$(11) \qquad \sum_{j=1}^{m} O_{ij}\, a_j^t \geq \sum_{j=1}^{m} x_{ij}\, a_j^{t+1} \qquad i=1,\ldots,n,$$

i.e. the amounts of individual outputs in period t must be equal to or greater than those required in the next period of production. Let us suppose that the equilibrium rate of growth is δ, so that in equilibrium we have

$$(12) \qquad a_j^{t+1} = (1+\delta)\, a_j^t \qquad j=1,\ldots,m.$$

Substituting (12) in (11) we have

$$(13) \qquad \sum_{j=1}^{m} O_{ij}\, a_j^t \geq \sum_{j=1}^{m} x_{ij}(1+\delta)\, a_j^t \qquad i=1,\ldots,n.$$

Another equilibrium condition is that in no process revenues should exceed costs, so that

$$(14) \qquad \sum_{i=1}^{n} p_i^t\, O_{ij} \leqq (1+r) \sum_{i=1}^{n} p_i^t\, x_{ij} \qquad j=1,\ldots,m,$$

where r is the rate of interest.

In (13) if for certain good the inequality holds, then its price will be zero, and in (14), if for certain activity or process the inequality holds then that activity will not be used. We now multiply (13) by p_i's and (14) by a_j's and know that in case of inequalities they become zero so that the inequalities disappear and we have

$$(15) \qquad \sum_{i=1}^{n} \sum_{j=1}^{m} O_{ij}\, a_j^t\, p_i^t = (1+\delta) \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}\, a_j^t\, p_i^t$$

$$(16) \qquad \sum_{j=1}^{m} \sum_{i=1}^{n} p_i^t\, O_{ij}\, a_j^t = (1+r) \sum_{j=1}^{m} \sum_{i=1}^{n} p_i\, x_{ij}^t\, a_j^t$$

From (15) and (16) it is obvious that $\delta = r$, i.e. the rate of equilibrium growth is equal to the rate of interest.

Without proving mathematically it can be intuitively observed that the equilibrium solution in von Neumann's model is optimal for if δ is less than the maximal, then it can seem that in (13) no equality will appear and then obviously δ can be raised, till at least there is one equality. Similarly it can be seen that the rate of interest is minimal, for, otherwise, there will not be any equality in (14).

von Neumann's model has led to a spate of theoretical research-work as did Leontief's model to empirical research work. We have seen the shortcomings of the latter. As regards the former, it paved the way for dynamic linear programming and the enormous literature that has followed. It has not been very successful in practice, simply because it has not been possible for any matrix of technology containing an exhaustive list or at least ~~an~~ already known list of alternative processes of production to be prepared. Hence, so far von Neumann's fundamental contribution has not been used in actual programming of economies, as such.

IV.

We have discussed above the three major attempts to construct discret production function under different sets of assumptions. One thing is common to all of them, they are all linear in inputs and imply constant returns to scale. These characteristics are quite justifiable for the purposes of short term production function, but the longer the period that is taken, the lesser the justification for these assumptions to hold. For long-term and medium-term planning, what is needed is not so much expost matrix of technology, but an ex-ante one and that too of the von Neumann type. Such a matrix of technology can be constructed by production engineers rather than the economists and statisticians, though the former may need the assistance of the latter. As stated above such a matrix of technology has not been constructed so far hence the discrete production functions outlined above are so far of ~~technological~~ theoretical value only. The matri of technology of the Leontief type is useful under static conditions or perhaps for countries whose economies have already settled on an equilibrium technology.

This is hardly true for underdeveloped countries and hence the limited utility of Leontief'type input-output tables for planning in these countries except in so far input-output tables are of some help in national accounting. Fortunately or unfortunately, the statistical data available in the underdeveloped countries are quite inadequate for constructing input-output tables and the ones that are available are of the nature of guess-work.

If production functions and techniques do help or matter in actual planning and programming, they can more gainfully be based on or derived from continuous functions, rather than the discrete ones. The economic theory in this field lags behind actual practice or reality perhaps more than it does in other fields of our economic life. This points up the need for greater efforts to be exerted by the economists and production engineers in this area.

(A.H.)

Appendix to Lecture 4.

Definition of elasticity of substitution is

(1) $$\frac{K}{L} = A \frac{(MP_L)^{\sigma}}{(MP}$$

Let the production function be

$$x = f(K,L), \quad f \text{ being linear homogeneous}$$

$$\therefore \quad MP_L = f(\tfrac{K}{L}) - \tfrac{K}{L}f'(\tfrac{K}{L})$$

$$MP_K = f'(\tfrac{K}{L}).$$

Case 1: $\sigma = 0$.

In (1) if we put $\sigma = 0$, we get the case of factor combination in fixed proportion of the input-output type:

(2) $$\frac{K}{L} = A \quad \text{or} \quad K = AL.$$

Case 2: $\sigma = 1$.

If we put $\sigma = 1$, in (1), we have

$$\frac{K}{L} = A \frac{f(\frac{K}{L}) - \frac{K}{L}f'(\frac{K}{L})}{f'(\frac{K}{L})}$$

or $$(1+A)\ \frac{K}{L} = A\frac{f(\frac{K}{L})}{f'(\frac{K}{L})}$$

$$\frac{d\frac{K}{L}}{(\frac{K}{L})} = B\frac{df(\frac{K}{L})}{f(\frac{K}{L})} \qquad \text{where } \frac{A}{1+A} = B$$

or $$\log \frac{K}{L} = B \ \log f(\tfrac{K}{L}) + \log C$$

$$\frac{K}{L} = \left(\frac{X}{L}\right)^{B} \cdot C$$

$$\therefore \quad X = K^{\frac{1}{B}} L^{1-\frac{1}{B}} \cdot D \qquad \text{where } D = C^{-\frac{1}{B}}$$

$$\text{or (3)} \quad X = DK^{1-\alpha} \cdot L^{\alpha} \qquad \text{where } \alpha = 1^{-\frac{1}{B}}$$

(3) is the Cobb-Douglas production function.

Case 3: σ = any constant

$$\frac{K}{L} = A\left[\frac{f(\frac{K}{L}) - \frac{K}{L} f'(\frac{K}{L})}{f'(\frac{K}{L})}\right]^{\sigma}$$

$$\therefore \frac{K}{L} + \left(\frac{K}{L}\right)^{\frac{1}{\sigma}} = \frac{f(\frac{K}{L})}{f'(\frac{K}{L})}, \qquad \text{putting } A=1 \text{ for simplicity}$$

$$\text{or} \quad \frac{df(\frac{K}{L})}{f(\frac{K}{L})} = \frac{d(\frac{K}{L})}{\frac{K}{L} + (\frac{K}{L})^{\frac{1}{\sigma}}}$$

$$= \left[\frac{1}{\frac{K}{L}} - \frac{(\frac{K}{L})^{\frac{1}{\sigma}-2}}{(\frac{K}{L})^{\frac{1}{\sigma}-1} + 1}\right] d\left(\frac{K}{L}\right)$$

$$\therefore \log f\left(\frac{K}{L}\right) = \log\left(\frac{K}{L}\right) - \frac{\sigma}{1-\sigma}\log\left(\left(\frac{K}{L}\right)^{\frac{\sigma}{1-\sigma}} + 1\right) + \log C$$

$$f\left(\frac{K}{L}\right) = C\frac{K}{L}\left(\left(\frac{K}{L}\right)^{\frac{\sigma}{1-\sigma}} + 1\right)^{-\frac{\sigma}{1-\sigma}}$$

$$\text{or} \quad \frac{X}{L} = C\left(1 + \left(\frac{K}{L}\right)^{\frac{\sigma-1}{\sigma}}\right)^{1-\frac{\sigma}{1-\sigma}}$$

$$\text{or} \quad X = C\left(L^{\frac{\sigma-1}{\sigma}} + K^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{1-\sigma}}$$

By putting some restrictions[1)] on the shares of labour and capital we get the SMAC

$$X = \gamma(\delta K^{-\rho} + (1-\delta) L^{-\rho})^{-\frac{1}{\rho}} \qquad \text{where } \rho = \frac{1-\sigma}{\sigma} \text{ and } \delta \text{ is the}$$

distribution parameter.

---

1) We shall complete the derivation as soon as I get a reference required.