Haberdashers' School

Occasional Papers Series in the Humanities

Occasional Paper Number Eighty-Four

Notes on the Elasticity of Factor Substitution

Ian St John

Economics Department

istjohn@habselstree.org.uk

June 2025

Haberdashers' School Occasional Paper Number Eighty-Four

June 2025

The Elasticity of Factor Substitution

Ian St John

Abstract

This paper explains the concept of the Elasticity of Factor Substitution, being a measure of how easily capital and labour can be substituted for each other in the production process. This concept is important since it helps to predict how readily production processes can adjust to differences in the relative supplies of factors of production, becoming more or less intensive in their utilisation of capital or labour as relative factor prices change. The value of the Elasticity of Factor Substitution (σ) depends on the percentage change in the capital-labour ratio divided by the percentage change in the Marginal Rate of Technical Substitution (MRTS), which measures how much of one factor must be given up for a given increase in another whilst holding total output constant. As the MRTS varies between zero and infinity so does the Elasticity of Factor Substitution, according to whether the factor inputs cannot be substituted for each other at all (MRTS = 0) or are perfect substitutes for each other (MRTS = ∞). Typically, we assume that the two factors are *imperfect* substitutes for each and that the MRTS varies as the ratio of factor inputs varies. In the case of a Cobb-Douglas production function the changes in the MRTS and the K/L ratio exactly counter-balance each other, with the result that the Elasticity of Substitution is constant and always equal to one. Evidence suggests that actual production functions have constant elasticities of substitution but not generally equal to one. Hence Arrow, Chenery, Minhas, and Solow developed the Constant Elasticity of Substitution (CES) production function, the derivation and properties of we summarise here.

The Elasticity of Factor Substitution is a measure of the degree to which one factor input used in a production process (such as capital) can be substituted for another factor input (such as labour) when the ratio of factor prices changes. The concept was developed in the 1930s, with notable contributions by John Hicks, in his *The Theory* *of Wages* (1932), A.P. Lerner, whose 'Notes on the Elasticity of Substitution' appeared in the *Review of Economic Statistics* in 1933, and Joan Robinson, in her *The Economics of Imperfect Competition* (1933). Assuming a production function with two factor inputs (such as capital and labour):

Q = f(K, L)

then the Elasticity of Substitution is defined as:

$$\sigma = \frac{\text{Proportionate Change in the ratio K/L}}{\text{Proportionate change in Marginal Rate of Technical Substitution of Labour for Capital}}$$

$$\sigma = \frac{\frac{d\left(\frac{K}{L}\right)}{K/L}}{\frac{d(MRTS)}{MRTS}} = \frac{d\left(\frac{K}{L}\right)}{d(MRTS)} \cdot \frac{MRTS}{\frac{K}{L}}$$

As we explain below, the Elasticity of Factor Substitution is also equal to:

$$\sigma = \frac{\text{Percentage change in } \left(\frac{K}{L}\right)}{\text{Percentage change in } \left(\frac{w}{r}\right)}$$

where w is the wage rate and r is the rental price of capital. Expressed in this latter way, the significance of σ becomes clearer: it tells us how far a change in relative factor prices will lead to a relative change in the ratio of factor inputs.

The Elasticity of Substitution takes values ranging from zero to infinity. What determines the Elasticity of Substitution is how rapidly the Marginal Rate of Technical Substitution of labour for capital ($MRTS_{LK}$) changes as labour is substituted for capital (or *vice versa*). If, as labour is substituted for capital (K/L declines), the $MRTS_{LK}$ declines rapidly (meaning that more and more labour must be substituted for capital to keep output constant – which is to say, labour is *less* able to substitute for capital) then the substitution of labour for capital is difficult and σ is less than one; but if, by contrast, as L increases relative to K the $MRTS_{LK}$ decreases slowly or not at all, then σ will be greater than one and substituting labour for capital will be easy. The value of σ thus obtained is important in determining what will be the effects of a change in the ratio of the price of labour and capital on the relative use of labour and capital in the production process. When σ is low (say 0.1) then a 10 per cent change in the MRTS (equivalent to a 10 per cent change in relative factor prices) will cause a small (1 per cent) change in factor proportions (0.1 x 10). But if the elasticity of substitution is high (say 5) then a 10 per cent change in the MRTS (or relative factor prices) will cause a (5 x 10) equals 50 per cent change in factor proportions.[1]

The Marginal Rate of Technical Substitution

The MRTS measures how much of one factor must be added to a production process to compensate for the deduction of a unit of another factor in order for total output (Q) to remain constant. The concept can be illustrated with reference to an isoquant.

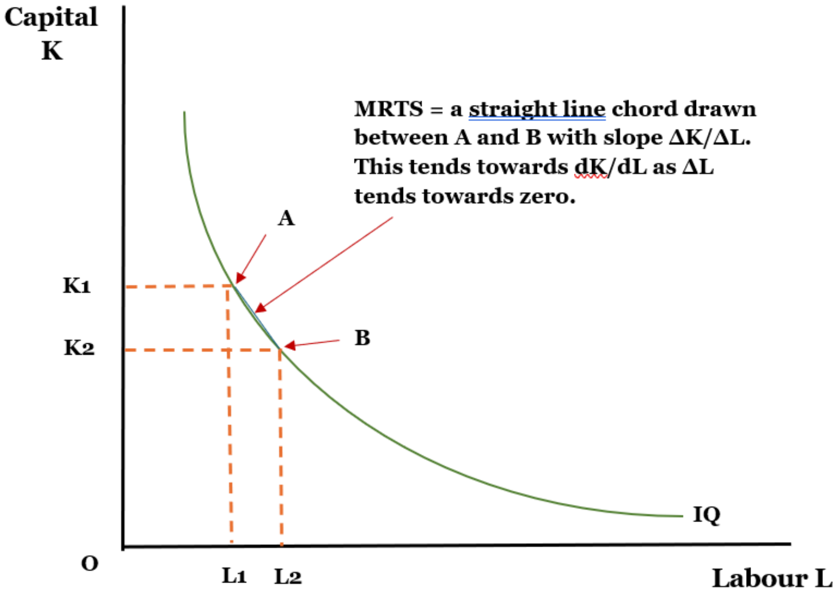

Figure 1. The Marginal Rate of Technical Substitution as the Gradient of an Isoquant

An isoquant shows combinations of capital and labour that yield a constant level of output. Isoquant IQ represents a given and constant level of output, this output associated with varying combinations of capital and labour inputs. Starting at point A with the combination of capital and labour K1/L1, in moving to point B capital inputs fall from K1 to K2 while labour inputs rise from L1 to L2. The ratio of the change in capital input to the change in the labour input is the Marginal Rate of Technical Substitution of labour for capital:

$$MRTS_{LK} = -\frac{\Delta K}{\Delta L}$$

The MRTS is always negative since if K decreases L must increase and *vice versa*. As the change in capital diminishes, the degree to which labour inputs need to change

[1] D. Heathfield and S. Wibe, *An Introduction to Cost and Production Functions* (Macmillan Education, Basingstoke, 1987), p. 59.

to keep output constant also diminishes – which is to say that B tends towards A. Hence $\Delta K/\Delta L$ tends towards dK/dL, which is the gradient of the isoquant at point A. Thus:

$$MRST_{LK} = -\frac{dK}{dL}$$

Thus, the MRTS is equal to the gradient of an isoquant. It depends upon the ratio of the Marginal Productivities of the two factors. This can be shown by taking the total differential of the production function:

$$Q = f(K, L)$$

$$dQ = \frac{\partial Q}{\partial K}dk + \frac{\partial Q}{\partial L}dL$$

Since dQ = 0 for a given isoquant (output is constant), then:

$$\frac{\partial Q}{\partial K}dk + \frac{\partial Q}{\partial L}dL = 0$$

$$\frac{\partial Q}{\partial K}dk = -\frac{\partial Q}{\partial L}dL$$

$$\frac{dK}{dL} = -\frac{\partial Q/\partial L}{\partial Q/\partial K} = \frac{MPL}{MPK}$$

Where MPL and MPK are the marginal products of labour and capital respectively.

Typically, we assume that isoquants are convex to the origin, such that their gradient declines continually as we move down the line. This is because capital and labour are considered to be imperfect substitutes for each other such that as capital inputs are reduced by successive units, more and more labour is required to compensate for each unit reduction in the capital input. Hence the $MRTS_{LK}$ is said to increase as we move down any given isoquant and the slope of the isoquant becomes flatter. For such an isoquant, the Elasticity of Substitution is greater than zero and less than infinity. We return to this case below. However, it is also theoretically possible that the MRTS does *not* vary along an isoquant yielding cases where the Elasticity of Substitution is zero or tending to infinite.

1. Elasticity of Substitution is Zero (σ = 0)

In this case technology requires the use of fixed factor proportions and it is not possible to substitute one factor for another. The $MRTS_{LK}$ is zero. One might think of a pilot and an aeroplane: each plane requires one pilot and pilots can't be substituted for planes or planes for pilots. Fixed-factor ratios generate L-shaped isoquants.

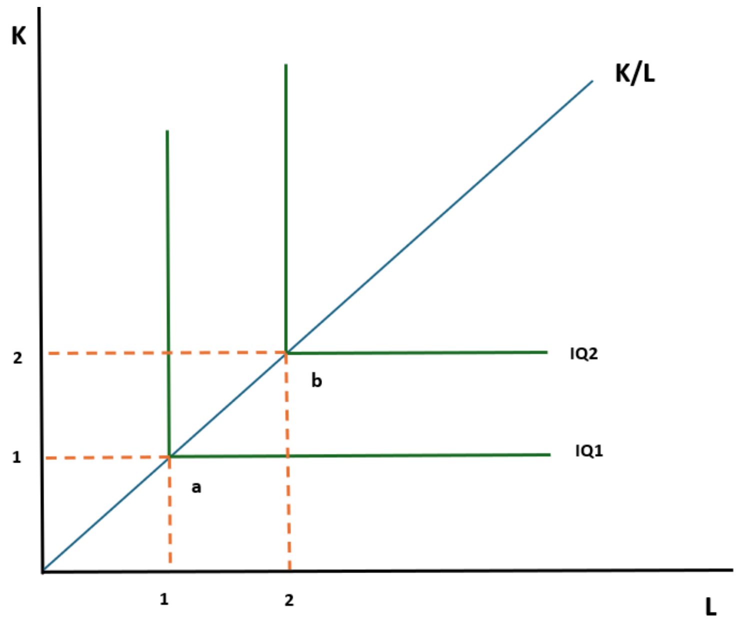

Figure 2. Isoquants where the Factor Inputs are Perfect Complements

In this example, capital and labour must be used in fixed 1:1 ratios, as in our plane-pilot illustration. At point **a** there is one pilot and one plane, and at point **b** there are two pilots and two planes. At a point such as **a** there is no reason to increase the number of pilots beyond one if the number of planes remains at one: the extra pilot contributes nothing and hence as L increases along the isoquant IQ1 output remains the same. Equally, if the number of pilots remains at one and the number of planes is increased, moving up the isoquant away from **a**, output does not increase as the new planes sit idle for there is no extra pilot to fly them. This is why the isoquant has a L shape. The MRTS is zero as the two factors L and K cannot be substituted for one another. In this case, a firm would only occupy the corner point of a given isoquant. Having arrived at the technically necessary ratio of capital to labour it would have no incentive to employ more of one or other factor since this would only add to costs – adding nothing to output.

When factors must be used in fixed proportions we say they are *perfect complements*. They *must* be used together (like the proverbial horse and carriage). In this case the production function is:

Q = min(aK, bL)

Here a and b are fixed technical parameters reflecting how many units of K and L are required to make a unit of Q (in our case Q is a flight and a and b are both 1), while 'min' means that Q is determined by the smaller of the two values aK and bL.[1] If aK <

[1] W. Nicholson, *Microeconomic Theory: Basic Principles and Extensions* (The Dryden Press, Illinois, Second Edition, 1978), pp. 197-199.

bL, then if K and L are applied in their technically fixed ratios Q is determined by K only:

Q = aK

Given K, then adding L beyond the fixed ratio of L to K makes no addition to output. So, in our example, if there were 100 planes and 200 pilots, then it is the lower number of 100 planes that determines output – 100 pilots will be superfluous. By contrast, if bL < aK, then it will be the number of pilots that fixes the maximum number of flights. When aK = bL then both factors are optimally utilised – as in the case of one pilot/one plane. Hence:

aK = bL

$$\frac{K}{L} = \frac{b}{a}$$

Here, the ratio of capital to labour in production (K/L) is constant and determined by the technically fixed ratio b/a. Since the MRTS and the change in the ratio K/L are both zero, the Elasticity of Factor Substitution (σ) is zero.[1]

2. Elasticity of Substitution is Infinite (σ = ∞): Perfect Substitutes

In this case one factor can be substituted for another continuously at a fixed rate. For example, every time we reduce the K input by 1 we can maintain output at a given amount by increasing the number of workers by 3; in which case the $MRTS_{LK}$ is -1/3 and this ratio never changes along the isoquant which is consequently a straight line. Such an isoquant is drawn below.

[1] If we assumed that two pilots were required to fly a plane (e.g. a pilot and a co-pilot) then the production function would become:

$Q = [(1)K, \left(\frac{1}{2}\right)L]$

Thus, if K = 100 and L = 100, then Q = (100, 50) and it would be the lower number 50 that would determine Q. In other words, with 2 pilots per flight, 100 pilots can fly 50 planes, so 50 planes would be used and 50 would remain idle.

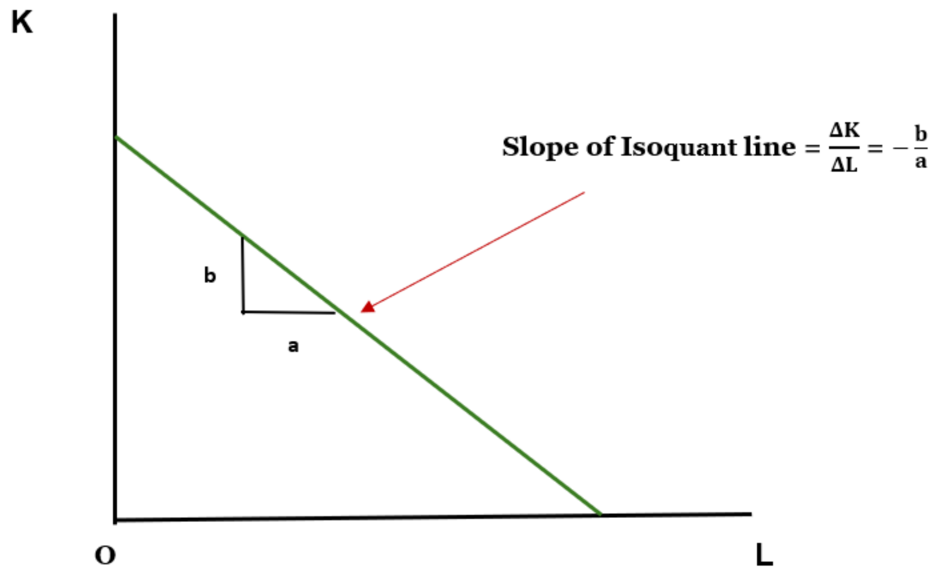

Figure 3. Isoquant when Capital and Labour are Perfect Substitutes

An example of a production function yielding such an isoquant is:

Q = f(K, L) = aK + bL

In this case:

$dQ = \frac{\partial Q}{\partial K}dk + \frac{\partial Q}{\partial L}dL$

dQ = adK + bdL = 0

dKa = – dLb

$\frac{dK}{dL} = -\frac{b}{a}$

The $MRTS_{LK}$ is constant and equal to -b/a. Given that:

$\sigma = \frac{\%\Delta \text{ in } K/L}{\%\Delta MRTS}$

then the fact that the percentage change in the MRTS is zero means that σ tends towards ∞.

3. Imperfect Substitutes (o < σ < ∞)

Our third case is more typical: namely, when two factors *can* be substituted for each other in the production process, but they are *not* perfect substitutes. In this case, the MRTS is negative but less than infinity and greater than zero and it *varies* along an isoquant. This is what is assumed in a conventional convex isoquant, where, as we move down an isoquant, the gradient of the curve becomes less negative and we say that the MRTS is diminishing – which means that the amount of capital given up per additional unit of labour input falls as we move down an isoquant. In other words, labour is an imperfect substitute for capital. At first, when we have lots of capital and little labour, then a significant amount of capital would need to be given up if an extra worker were employed in order to keep output constant. If, however, we have lots of workers in a factory but only a few machines, then an extra worker would make hardly any difference to output and so very little capital would need to be given up to keep output constant. This is the point about the factors not being perfect substitutes. Yes, if you substitute machines for workers you can maintain or increase output, but as the number of machines rises and the number of workers declines, it gets harder and harder to increase output with machines since the workers left fulfil vital functions like turning the machines on, setting them to run properly, clearing away waste products and so forth (in the context of a super market, who would reset the check-out machine when there is an unknown item in the bagging area?). Similarly, a farm could cut its capital inputs and increase the number of workers and maintain output – but bit by bit the workers would have fewer and fewer tools and machines to work with and maintaining output when there is not even a spade to dig with would be almost impossible. When an isoquant is curved then both the K/L ratio and the MRTS change as we move along its length.

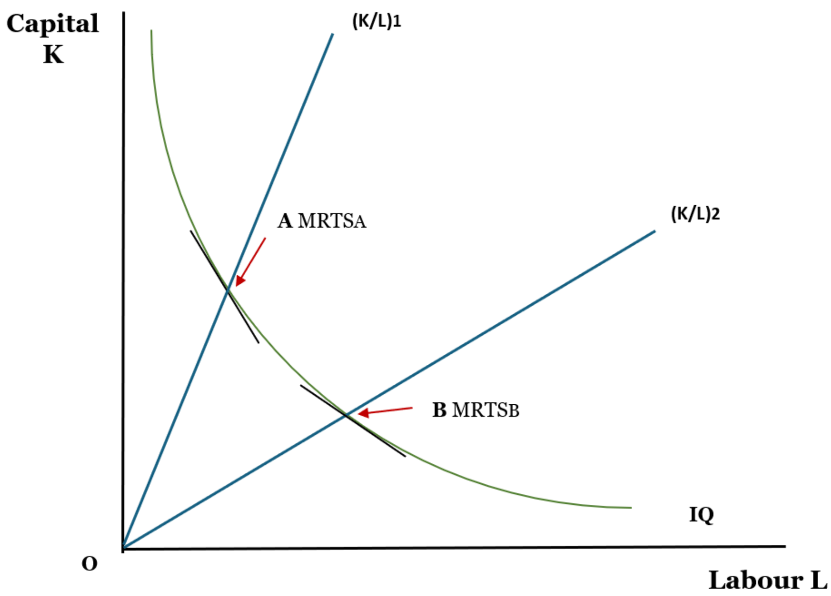

Figure 4. An Isoquant when Capital and Labour are Imperfect Substitutes

Consider the two points, A and B, on the isoquant depicted in **Figure 4**. Each point represents a particular combination of factor inputs (K/L) and to each there corresponds a MRTS, given by the gradient of the isoquant at that point (dK/dL). As we move down the isoquant from point A to point B two things change:

i. The ratio of K to L
ii. The MRTS of L for K

It is the relationship between the simultaneous changes in both that determines how the Elasticity of Substitution behaves along the isoquant. The sharper the curve of the isoquant the lower will be the Elasticity of Substitution – which, at the limit of a L-shaped isoquant, is zero – while the flatter the isoquant the greater the Elasticity of Substitution, tending to infinite if the isoquant is a straight line.

To understand how changes in the MRTS cause changes in the ratio of K to L we need to recall that a cost-minimising firm will always locate at that point on a given isoquant where the gradient of the isoquant is equal to the gradient of an iso-cost line. An isocost line shows the different combinations of capital and labour a firm can use for a given cost outlay. For each given level of Total Costs there is a different isocost line. Given a firm has two factor inputs, capital and labour, and the cost per unit of labour time used (e.g. one hour) is the wage rate w, and the rental cost per time unit of capital employed is r, then the firms Total Costs are:

$TC = wL + rK$

Where L is the amount of labour in hours and K is the amount of capital services in hours. Expressing this in terms of K:

$rK = TC – wL$

$K = \frac{TC}{r} - \frac{w}{r}L$

This is a linear function. The intercept of the vertical axis, when the firm only buys capital and employs no labour at all, is TC/r, i.e. the total spending on inputs divided by the price per unit of capital, r. Clearly if we divide a total amount of cost spending by the price per unit of capital, we arrive at the total amount of capital the firm can acquire. The *slope* of the line is w/r, which is determined by the ratio of input prices. The higher the wage rate, or the lower the rental price of capital, the steeper the isocost line – which again makes sense: if a firm employs less capital, the amount of extra labour it can employ is larger if the wage is low or the cost of capital is high.

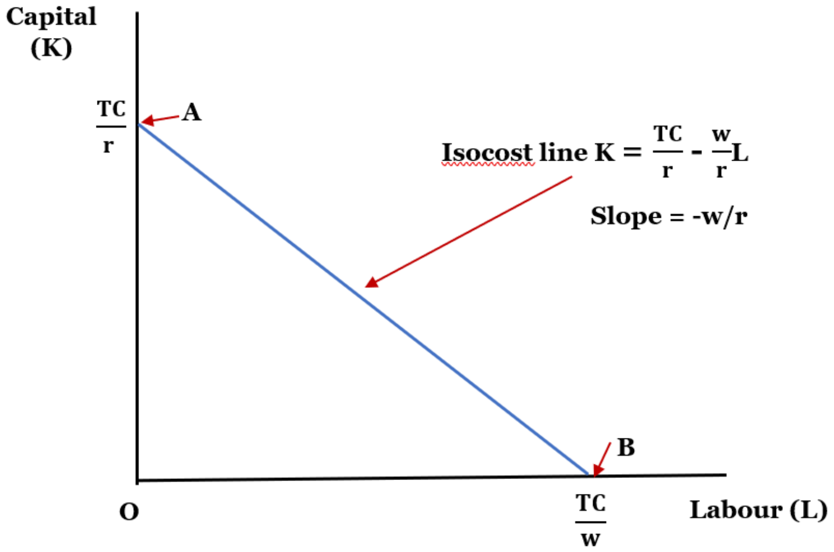

Figure 5. General Form of an Isocost Line

At point A the firm spends all its outlay on capital, so the quantity of capital hired is TC/r, while B is where it spends all its total cost outlay on labour, so the maximum quantity of labour it can hire is TC/w. The slope of the line is -w/r.

Here we have shown just one isocost line. In fact, there will be innumerable potential isocost lines, each corresponding to a particular level of costs. A higher total cost yields an isocost line that intercepts the y and x axes at higher values, while a lower cost outlay will yield isocost lines below the example shown.

To return to our initial question. Suppose the firm wishes to produce a *given* output Q at the lowest possible cost, what combinations of capital and labour will it employ? We can easily answer this question by combining isoquants with isocost curves. In the case of *cost minimisation*, we take a given output as fixed and to this there corresponds exactly one isoquant. We then shift an isocost line inwards until it is *just tangential to the isoquant*. This is the lowest possible cost at which a given output can be produced given the available technology and factor prices. The point at which this tangency position occurs on the isoquant will yield the cost-minimisation combination of capital and labour. This is illustrated in **Figure 6**.

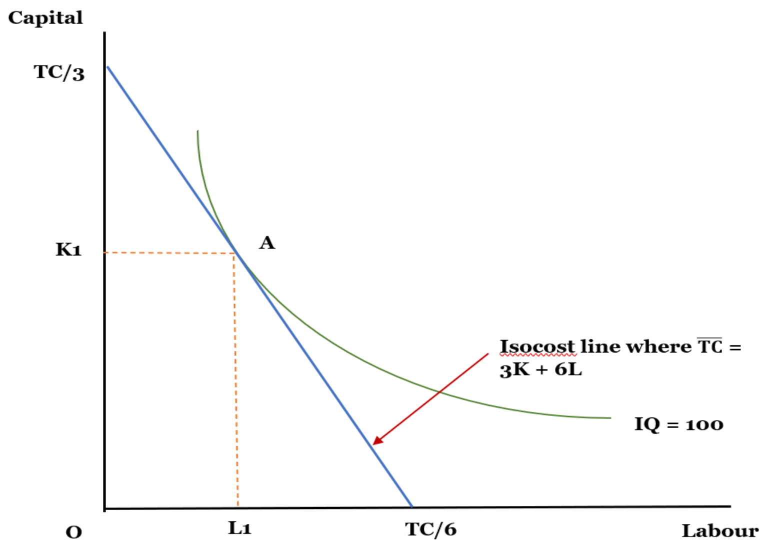

Figure 6. Obtaining optimum combinations of Capital and Labour from the Tangency Point between the Isocost and Isoquant Curves

In **Figure 6**, given the firm wishes to produce output 100, represented by the isoquant, the lowest possible isocost curve compatible with IQ = 100 is the one drawn, which is tangential to IQ = 100 at point A. Point A, with the corresponding Labour inputs of L1 and Capital inputs K1, is the cost-minimising position of a firm seeking to produce an output of Q = 100. Any isocost curve *below* that shown would not be compatible with the quantities of Capital and Labour necessary to produce Q = 100 and is thus not feasible. It would be possible to produce Q = 100 with isocost curves *above* that shown, but these would represent higher levels of cost and would thus not be cost-minimising. Thus, the cost-minimising firm would employ OK1 units of Capital and OL1 units of Labour. Note that the firm will employ relatively more Capital than Labour which reflects the fact that the price of Labour is *double* the price of Capital.

It is this result which enables us to better understand the meaning of the Elasticity of Substitution. At the tangency point A the gradient of the isoquant is equal to the gradient of the isocost line. That is:

$$\frac{dK}{dL} = \frac{w}{r}$$

Now we already know that the gradient of the isoquant is the MRTS which is equal to:

$$\frac{dK}{dL} = -\frac{\partial Q/\partial L}{\partial Q/\partial K} = \frac{MPL}{MPK}$$

It therefore follows that a cost-minimising firm will employ that combination of capital and labour where:

$$\frac{MPL}{MPK} = \frac{w}{r}$$

In other words, a cost-minimising firm will employ the combination of capital and labour (K/L) where the Marginal Rate of Technical Substitution, which is equal to the ratio of the Marginal Product of Labour over the Marginal Product of Capital, is equal to the ratio of the wage of labour (w) to the rental cost of capital (r). It is for this reason that we can re-write the Elasticity of Substitution as:

$$\sigma = \frac{\frac{d\left(\frac{K}{L}\right)}{K/L}}{\frac{d(MRTS)}{MRTS}} = \frac{\frac{d\left(\frac{K}{L}\right)}{K/L}}{\frac{d(w/r)}{w/r}}$$

Hence we can say that the Elasticity of Substitution measures the degree to which a change in relative factor prices will bring about a change in the relative use of factors in the production process. When σ = 0, and the two factors must be used in fixed proportions, a relative rise in (say) wages will *not* cause any tendency for firms to substitute capital for labour; when σ tends to ∞ then a small rise in wages will cause an extremely large change in the relative utilisation of capital and labour; while if σ = 1 then a relative increase in wages relative to the rental price of capital will cause an equal percentage change in the capital-to-labour ratio.

Constant Elasticity of Substitution (CES)

It is quite possible that the Elasticity of Substitution will vary along an isoquant. But economists have been chiefly interested in isoquants that have a constant Elasticity of Substitution along their length: what are known as *Constant Elasticity of Substitution (CES) production functions*. Two cases are especially noteworthy.

Elasticity of Substitution (σ) Equals One

An example of a production function yielding σ = 1 along the length of the isoquant is a Cobb-Douglas production function. A Cobb-Douglas production function has the form:

$Q = f(K, L) = AK^{\alpha}L^{\beta}$

Where A, α, and β are all constants. A is a measure of technical efficiency (total factor productivity). α and β represent the shares of total output received by capital and labour respectively and are usually assumed to sum to one: i.e. $\alpha + \beta = 1$. This yields the case of constant returns to scale, meaning that if K and L are both increased by some common factor (m), then output Q will also increase by m. This can be demonstrated as follows.

$Q = AK^{\alpha}L^{\beta}$

Now increase both K and L by some scalar m:

$Q = A(mK)^{\alpha}(mL)^{\beta}$

$Q = Am^{\alpha}K^{\alpha}m^{\beta}L^{\beta}$

$Q = Am^{\alpha+\beta}K^{\alpha}L^{\beta}$

$Q = AmK^{\alpha}L^{\beta}$ (since $\alpha + \beta = 1$)

$Q = mQ$

Thus, increasing both factors by m increases total output by m. For example, if both capital and labour are doubled (m = 2) then Q will double (2Q). This is constant returns to scale. If $\alpha + \beta > 1$ there are *increasing returns to scale* and if $\alpha + \beta < 1$ there are *decreasing returns to scale*.

We can show that a Cobb-Douglas production function has an Elasticity of Substitution of 1 whatever its returns to scale. We know that:

$$\sigma = \frac{d\left(\frac{K}{L}\right)}{d(MRTS)} \cdot \frac{MRTS}{K/L}$$

We know also that:

$$MRTS_{LK} = -\frac{\partial Q/\partial L}{\partial Q/\partial K} = -\frac{MPL}{MPK}$$

For a Cobb-Douglas function:

$$\frac{\partial}{\partial K}(AK^{\alpha}L^{\beta}) = \alpha AK^{\alpha-1}L^{\beta} = MPK$$

$$\frac{\partial}{\partial L}(AK^{\alpha}L^{\beta}) = \beta AK^{\alpha}L^{\beta-1} = MPL$$

Hence the MRTS equals:

$$MRTS_{LK} = \frac{\beta AK^{\alpha}L^{\beta-1}}{\alpha AK^{\alpha-1}L^{\beta}} = -\frac{\beta}{\alpha}\frac{K}{L}$$

Thus, the MRTS depends upon the ratio of K to L and the fixed ratio (β/α).

In which case:

$$\sigma = \frac{d\left(\frac{K}{L}\right)}{d\left(-\frac{\beta K}{\alpha L}\right)} \cdot \frac{-\frac{\beta K}{\alpha L}}{K/L} = \frac{d\left(\frac{K}{L}\right)}{d\left(-\frac{\beta K}{\alpha L}\right)} \cdot -\frac{\beta}{\alpha}$$

Since β/α is fixed, $d\left(-\frac{\beta}{\alpha}\frac{K}{L}\right) = -\frac{\beta}{\alpha}d\left(\frac{K}{L}\right)$. Which means we have:

$$\sigma = \frac{d\left(\frac{K}{L}\right)}{d\left(-\frac{\beta K}{\alpha L}\right)} \cdot -\frac{\beta}{\alpha} = \frac{d\left(\frac{K}{L}\right)}{-\frac{\beta}{\alpha} d\left(\frac{K}{L}\right)} \cdot -\frac{\beta}{\alpha} = \frac{-\frac{\beta}{\alpha}}{-\frac{\beta}{\alpha}} = 1$$

Hence, we have shown that for a Cobb-Douglas production function with constant returns to scale has an Elasticity of Substitution of 1 for any ratio of K to L, i.e. at any point on a given isoquant. What this means is that for any movement along an isoquant, the percentage change in capital intensity (K/L) is the same as the percentage change in the MRTS.[1]

A Numerical Example

Suppose our Cobb-Douglas production function is of the form:

$Q = 10L^{0.6}K^{0.4}$

Assuming that L = 100 and K = 200, we have:

$Q = 10(100)^{0.6}(200)^{0.4}$

Since the power terms sum to one (0.6 + 0.4) this production function exhibits constant returns to scale. The Marginal Rate of Technical Substitution is:

$$MRTS_{LK} = -\frac{\partial Q/\partial L}{\partial Q/\partial K} = -\frac{MPL}{MPK} = \frac{0.6(10)L^{-0.4}K^{0.4}}{0.4(10)L^{0.6}K^{-0.6}} = \frac{6L^{-0.4}K^{0.4}}{4L^{0.6}K^{-0.6}} = -\frac{3}{2}\frac{K}{L} =$$

$$-1.5\left(\frac{200}{100}\right) = -1.5 \text{ x } 2 = -3.$$

$$\sigma = \frac{d\left(\frac{K}{L}\right)}{d\left(-1.5\frac{K}{L}\right)} \cdot \frac{-1.5\frac{K}{L}}{\frac{K}{L}} = \frac{d\left(\frac{K}{L}\right)}{d\left(-1.5\frac{K}{L}\right)} \cdot -1.5 = \frac{d\left(\frac{K}{L}\right)}{-1.5d\left(\frac{K}{L}\right)} \cdot -1.5$$

Given that L = 100 and K = 200, then:

$\frac{K}{L} = \frac{200}{100} = 2$

Now suppose capital increases from 200 to 300 and labour decreases from 100 to 50 (in other words, that we are moving up an isoquant). In this case:

$\frac{K}{L} = \frac{300}{50} = 6$

Hence:

[1] S. Estrin, D. Laidler, and M. Dietrich, *Microeconomics* (Pearson, Harlow, Fifth Edition, 2008), p. 155.

$d\left(\frac{K}{L}\right) = 6 - 2 = 4$

Substituting this value for d(K/L) into

$$\sigma = \frac{d\left(\frac{K}{L}\right)}{-1.5d\left(\frac{K}{L}\right)} \cdot -1.5 = \frac{4}{-1.5(4)} \times -1.5 = \frac{4}{-6} \times -1.5 = 1$$

Which shows, as expected, that the Elasticity of Substitution along a Cobb-Douglas isoquant is 1.

Constant Elasticity of Substitution (CES) Production Function

A Cobb-Douglas production function always has an Elasticity of Substitution of one. However, in 1961, Arrow, Chenery, Minhas, and Solow formulated a production function with a constant Elasticity of Substitution which can take *any* positive value.[1] Their CES production function has the following form:

$$Q = A[\alpha K^{-\rho} + (1 - \alpha)L^{-\rho}]^{-1/\rho}$$

A is a total factor productivity parameter and is > 0

α is a factor distribution parameter – in this case the distribution of the total product Q between capital (K) and labour (L). It is $0 \leq \alpha \leq 1$

ρ (pronounced roe) is a measure of the degree of factor substitution. Its values are **ρ** ≥ -1.

This function is homogeneous in degree one, meaning it exhibits constant returns to scale. We can show this as follows. Imagine the values for capital and labour are both scaled up or down by the quantity j. Then:

$$Q = A[\alpha K^{-\rho} + (1 - \alpha)L^{-\rho}]^{-1/\rho}$$

becomes:

$$Q = A[\alpha(jK)^{-\rho} + (1 - \alpha)(jL)^{-\rho}]^{-1/\rho}$$

$$Q = A\{j^{-\rho}[\alpha K^{-\rho} + (1 - \alpha)L^{-\rho}]\}^{-1/\rho}$$

$$Q = (j^{-\rho})^{-1/\rho}A[\alpha K^{-\rho} + (1 - \alpha)L^{-\rho}]^{-1/\rho}$$

$$Q = j(Q)$$

[1] K. Arrow, H. Chenery, B. Minhas, and R. Solow, 'Capital-Labor Substitution and Economic Efficiency', *The Review of Economics and Statistics*, Vol XLIII, No. 3 (August 1961), pp. 225-250.

Hence, multiplying the factor inputs by a scalar j increases total output Q by the same scalar j, which is what is meant by constant returns to scale.

The isoquant generated by such a production function is downward sloping when K is plotted on the vertical axis and L on the horizontal. To show this, we first need to calculate the marginal products of capital and labour.[1]

First the marginal product of labour Q_L. We calculate this by taking the partial derivative of Q with respect to L.

$$\frac{\partial}{\partial \mathrm{L}}\{\mathrm{A}[\alpha \mathrm{K}^{-\rho} + (1-\alpha)\mathrm{L}^{-\rho}]^{-1/\rho}\}$$

Setting $\alpha \mathrm{K}^{-\rho} + (1-\alpha)\mathrm{L}^{-\rho} = \mathrm{Z}$ we have:

$$\mathrm{Z} = \alpha \mathrm{K}^{-\rho} + (1-\alpha)\mathrm{L}^{-\rho}$$

Hence:

$$\mathrm{Q} = \mathrm{A}[\mathrm{Z}]^{-1/\mathrm{p}}$$

We now differentiate Q by L using the chain rule:

$$\frac{\partial \mathrm{Q}}{\partial \mathrm{L}} = \frac{\partial}{\partial \mathrm{L}}\mathrm{A}[\mathrm{Z}]^{-1/\mathrm{p}} \text{ x } \frac{\partial \mathrm{Z}}{\partial \mathrm{L}}$$

$$\frac{\partial}{\partial \mathrm{L}}\mathrm{A}[\mathrm{Z}]^{-1/\rho} = \mathrm{A}\left(-\frac{1}{\rho}\right)\mathrm{Z}^{-(\frac{1}{\rho}-1)} = \mathrm{A}\left(-\frac{1}{\rho}\right)\mathrm{Z}^{-(\frac{1}{\rho}+1)}$$

Differentiating Z by L we have:

$$\frac{\partial \mathrm{Z}}{\partial \mathrm{L}} = \frac{\partial}{\partial \mathrm{L}}[\alpha \mathrm{K}^{-\rho} + (1-\alpha)\mathrm{L}^{-\rho}]$$

$$\frac{\partial \mathrm{Z}}{\partial \mathrm{L}} = (1-\alpha)\text{-}\boldsymbol{\rho}\mathrm{L}^{-\rho-1}$$

Thus:

$$\frac{\partial \mathrm{Q}}{\partial \mathrm{L}} = \mathrm{A}\left(-\frac{1}{\rho}\right)\mathrm{Z}^{-(\frac{1}{\rho}+1)} \text{ x } (1-\alpha)\text{-}\boldsymbol{\rho}\mathrm{L}^{-\rho-1}$$

$$\frac{\partial \mathrm{Q}}{\partial \mathrm{L}} = \mathrm{A}(1\text{-}\alpha)\ \mathrm{L}^{-\rho-1}\mathrm{Z}^{-(\frac{1}{\rho}+1)} \qquad (\text{since -}1/\boldsymbol{\rho} \text{ x -}\boldsymbol{\rho} = 1)$$

Now, we already know that:

$$\mathrm{Q} = \mathrm{A}[\mathrm{Z}]^{-1/\mathrm{p}}$$

Hence:

[1] In the calculations which follow I have been greatly assisted by Gautham Arun of Haberdashers' School, Elstree.

$Z^{-1/\rho} = \frac{Q}{A}$

$Z = \left(\frac{Q}{A}\right)^{-\rho}$ (multiplying the exponents of both sides by -**ρ**)

So:

$Z^{-(\frac{1}{\rho}+1)}$ becomes $\left[\left(\frac{Q}{A}\right)^{-\rho}\right]^{-(\frac{1}{\rho}+1)} = \left(\frac{Q}{A}\right)^{1+\rho}$ [since $-\boldsymbol{\rho}(-\frac{1}{\rho} - 1) = 1 + \boldsymbol{\rho}$]

Thus:

$\frac{\partial Q}{\partial L} = A(1\text{-}\alpha)\, L^{-\rho-1} Z^{-(\frac{1}{\rho}+1)}$

can now be written:

$\frac{\partial Q}{\partial L} = A(1\text{-}\alpha)\, L^{-\rho-1} \left(\frac{Q}{A}\right)^{1+\rho}$

$\frac{\partial Q}{\partial L} = A\left(\frac{1}{A^{1+\rho}}\right)(1-\alpha)\, L^{-\rho-1} Q^{1+\rho}$

$\frac{\partial Q}{\partial L} = A^{-\rho}(1-\alpha)\, L^{-\rho-1} Q^{1+\rho}$

$\frac{\partial Q}{\partial L} = \frac{(1-\alpha)\, L^{-\rho-1}}{A^{\rho}} Q^{1+\rho}$

$\frac{\partial Q}{\partial L} = \frac{(1-\alpha)}{A^{\rho}} \frac{Q^{1+\rho}}{L^{\rho+1}}$

$\frac{\partial Q}{\partial L} = Q_L = MP_L = \frac{(1-\alpha)}{A^{\rho}} \left(\frac{Q}{L}\right)^{\rho+1}$

The marginal product of capital Q_K is calculated by taking the partial derivative of Q with respect to K:

$\frac{\partial}{\partial K}\{A[\alpha K^{-\rho} + (1-\alpha)L^{-\rho}]^{-1/\rho}\}$

We calculate this by the same use of the chain rule we used to arrive at the marginal product of labour, the expression for Q_K being:

$\frac{\partial Q}{\partial K} = Q_K = \frac{\alpha}{A^{\rho}} \left(\frac{Q}{K}\right)^{\rho+1}$

Note that in both cases the marginal products of labour and capital are positive since α is positive and less than one so both $(1\text{-}\alpha)/A^{\rho}$ and α/A^{ρ} are positive. We can also show that the gradient of the isoquant is negative. The gradient of an isoquant is dK/dL, which is equal to $-MP_L/MP_K$. Thus:

$$\frac{dK}{dL} = -\frac{QL}{QK} = -\frac{\frac{(1-\alpha)}{A^{\rho}}\left(\frac{Q}{L}\right)^{\rho+1}}{\frac{\alpha}{A^{\rho}}\left(\frac{Q}{K}\right)^{\rho+1}} = -\frac{(1-\alpha)}{A^{\rho}}\left(\frac{Q}{L}\right)^{\rho+1}\frac{A^{\rho}}{\alpha}\left(\frac{K}{L}\right)^{\rho+1} = -\left(\frac{1-\alpha}{\alpha}\right)\left(\frac{K}{L}\right)^{\rho+1}$$

which is negative since both the parentheses terms are positive.

Finally, following the exposition of Alpha Chiang, we can derive an expression for the Elasticity of Substitution for this production function.[1]

We know that a firm will minimise its costs of producing a given output when its isoquant is tangential to its isocost curve, where the gradient of the isocost line is P_L/P_K. In other words:

$$\frac{dK}{dL} = \left(\frac{1-\alpha}{\alpha}\right)\left(\frac{K}{L}\right)^{\rho+1} = \frac{w}{r}$$

$$\left(\frac{1-\alpha}{\alpha}\right)^{\frac{1}{1+\rho}}\left(\frac{K}{L}\right) = \left(\frac{w}{r}\right)^{\frac{1}{1+\rho}}$$

$$\left(\frac{K}{L}\right) = \frac{\left(\frac{w}{r}\right)^{\frac{1}{1+\rho}}}{\left(\frac{1-\alpha}{\alpha}\right)^{\frac{1}{1+\rho}}}$$

$$\left(\frac{K}{L}\right) = \left(\frac{\alpha}{1-\alpha}\right)^{\frac{1}{1+\rho}}\left(\frac{w}{r}\right)^{\frac{1}{1+\rho}}$$

To simplify, we write:

$$\left(\frac{\alpha}{1-\alpha}\right)^{\frac{1}{1+\rho}} = c$$

Hence:

$$\left(\frac{K}{L}\right) = c\left(\frac{w}{r}\right)^{\frac{1}{1+\rho}}$$

We now have K/L as a linear function of w/r. Thus, differentiating K/L by w/r we get:

$$\frac{d\left(\frac{K}{L}\right)}{d\left(\frac{w}{r}\right)} = \frac{c}{1+\rho}\left(\frac{w}{r}\right)^{\frac{1}{1+\rho}-1}$$

Having arrived at these results, we can now calculate the formula for the Elasticity of Substitution under a CES production function. Recall that the Elasticity of Factor Substitution is:

[1] A. Chiang, *Fundamental Methods of Mathematical Economics* (McGraw-Hill, New York, 2nd Edition, 1974), pp. 416-417.

$$\sigma = \frac{d\left(\frac{K}{L}\right)}{d(\text{MRTS})} \cdot \frac{\text{MRTS}}{K/L} \text{ or } \frac{\frac{d\left(\frac{K}{L}\right)}{d(\text{MRTS})}}{\frac{K/L}{\text{MRTS}}}$$

We have seen that, for a cost minimising firm, the Marginal Rate of Technical Substitution is equal to MP_L/MP_K = w/r. Hence any cost-minimising firm seeking an optimal combination of capital and labour inputs will respond to a change in relative prices of labour and capital by adjusting its use of labour and capital until:

$$\frac{\text{MPL}}{\text{MPK}} = \frac{w}{r} = \text{MRTS}$$

and:

$$d\left(\frac{\text{MPL}}{\text{MPK}}\right) = d\left(\frac{w}{r}\right) = d\text{MRTS}$$

So, with respect to our quotient formula for σ, the numerator term can be re-written as follows:

$$\frac{d\left(\frac{K}{L}\right)}{d(\text{MRTS})} = \frac{d\left(\frac{K}{L}\right)}{d\left(\frac{w}{r}\right)} = \frac{c}{1+\rho}\left(\frac{w}{r}\right)^{\frac{1}{1+\rho}-1}$$

With respect to the denominator, we have seen that:

$$\left(\frac{K}{L}\right) = c\left(\frac{w}{r}\right)^{\frac{1}{1+\rho}}$$

While:

$$\text{MRTS} = \left(\frac{w}{r}\right)$$

Thus:

$$\frac{K/L}{\text{MRTS}} = \frac{c\left(\frac{w}{r}\right)^{\frac{1}{1+\rho}}}{\left(\frac{w}{r}\right)} = c\left(\frac{w}{r}\right)^{\frac{1}{1+\rho}-1}$$

Combining these results we have:

$$\sigma = \frac{\frac{d\left(\frac{K}{L}\right)}{d(\text{MRTS})}}{\frac{K/L}{\text{MRTS}}} = \frac{\frac{c}{1+\rho}\left(\frac{w}{r}\right)^{\frac{1}{1+\rho}-1}}{c\left(\frac{w}{r}\right)^{\frac{1}{1+\rho}-1}} = \frac{1}{1+\rho}$$

Thus we arrive at the formula for calculating the Elasticity of Supply for a CES production function, namely:

$$\sigma = \frac{1}{1+\rho}$$

What this shows is that, for a given **ρ**, the Elasticity of Substitution is a constant, where **ρ** is a measure of the degree to which labour and capital can be substituted. But, unlike the Cobb-Doulgas function, σ can take different values according to the value of **ρ**. These can be summarised as follows:

-1 < **ρ** < 0 then σ > 1. As **ρ** tends towards -1 then σ tends towards ∞ which is the case of perfect factor substitution, corresponding to a linear isoquant.

ρ = 0 then σ = 1. This is the case of the Cobb-Douglas production function.

ρ > 0 < ∞ then σ < 1 tending towards 0.

Analysing production, capital-labour ratios, wage, and factor share data for US and Japanese industry over the period 1949-1955, Arrow, Chenery, Minhas, and Solow estimated that the Elasticity of Substitution (σ) was generally less than 1, being around 0.85. In other words, capital and labour are substitutable, but not perfectly so and slightly less than the Cobb-Douglas production function implies. This suggests that if the ratios at which factors of production such as labour and capital are present within an economy change, and the ratios of factor prices and hence the optimal MRTS change, the production process is sufficiently flexible in terms of Elasticity of Factor Substitution to permit the varying factor supplies to be employed. Subsequent research has broadly endorsed these findings. Most studies find that σ < 1, usually occupying something within the range of 0.4 to 0.8, meaning that a 1% rise in the ratio of the price of labour to the price of capital will cause around a 0.6% rise in the ratio of capital to labour.[1]

[1] D. Hamermesh, *Labor Demand* (Princeton University Press, Princeton, 1993), pp. 77-79.