A Brief History of Production Functions

SK Mishra

North-Eastern Hill University, Shillong (India)

9. October 2007

Online at http://mpra.ub.uni-muenchen.de/5254/
MPRA Paper No. 5254, posted 10. October 2007
A Brief History of Production Functions

SK Mishra
Dept. of Economics
North-Eastern Hill University
Shillong (India)

I. Introduction: Production function has been used as an important tool of economic analysis in the neoclassical tradition. It is generally believed that Philip Wicksteed (1894) was the first economist to algebraically formulate the relationship between output and inputs as \( P = f(x_1, x_2, \ldots, x_m) \) although there are some evidences suggesting that Johann von Thünen first formulated it in the 1840’s (Humphrey, 1997).

It is relevant to note that among others there are two leading concepts of efficiency relating to a production system: the one often called the ‘technical efficiency’ and the other called the ‘allocative efficiency’ (see Libenstein et al., 1988). The formulation of production function assumes that the engineering and managerial problems of technical efficiency have already been addressed and solved, so that analysis can focus on the problems of allocative efficiency. That is why a production function is (correctly) defined as a relationship between the maximal technically feasible output and the inputs needed to produce that output (Shephard, 1970). However, in many theoretical and most empirical studies it is loosely defined as a technical relationship between output and inputs, and the assumption that such output is maximal (and inputs minimal) is often tacit. Further, although the relationship of output with inputs is fundamentally physical, production function often uses their monetary values. The production process uses several types of inputs that cannot be aggregated in physical units. It also produces several types of output (joint production) measured in different physical units. There is an extreme view that (in a sense) all production processes produce multiple outputs (Faber, et al., 1998). One of the ways to deal with the multiple output case is to aggregate different products by assigning price weights to them. In so doing, one abstracts away from essential and inherent aspects of physical production processes, including error, entropy or waste. Moreover, production functions do not ordinarily model the business processes, whereby ignoring the role of management, of sunk cost investments and the relation of fixed overhead to variable costs (wikipedia-a).

It has been noted that although the notion of production function generally assumes that technical efficiency has been achieved, this is not true in reality. Some economists and operations research workers (Farrel, 1957; Charnes et al., 1978; Banker et al., 1984; Lovell and Schmidt, 1988; Seiford and Thrall, 1990; Emrouznejad, 2001, etc) addressed this problem by what is known as the ‘Data Envelopment Analysis’ or DEA. The advantages of DEA are: first that here one need not specify a mathematical form for the production function explicitly; it is capable of handling multiple inputs and outputs and being used with any input/output measurement; and efficiency at technical/managerial level is not presumed. It has been found useful for investigating into the hidden relationships and causes of inefficiency. Technically, it uses linear programming as a method of analysis. We do not intend to pursue this approach here.
Starting in the early 1950’s until the late 1970’s production function attracted many economists. During the said period a number of specifications or algebraic forms relating inputs to output were proposed, thoroughly analyzed and used for deriving various conclusions. Especially after the end of the ‘capital controversy’, search for new specification of production functions slowed down considerably. Our objective in this paper is to briefly describe that line of development. In the schema of Ragnar Frisch (1965), we will first concentrate on “single-ware” or single-output production function. Then we would move to “multi-ware” or multi-output production function. Finally, we would address the pros and cons of the aggregate production function.

II. Single Output Production Function: Humphrey (1997) gives an outline of historical development of the concept and mathematical formulation of production functions before the enunciation of Cobb-Douglas function in 1928. Paul Douglas, on a sabbatical at Amherst, asked mathematics professor Charles W. Cobb to suggest an equation describing the relationship among the time series on manufacturing output, labor input, and capital input that Douglas had assembled for the period 1889–1922, and this led to their joint paper.

An implicit formulation of production functions dates back to Turgot. In his 1767 Observations on a Paper by Saint-P’eravy, Turgot discusses how variations in factor proportions affect marginal productivities (Schumpeter, 1954). Malthus introduced the logarithmic production function (Stigler, 1952) and Barkai (1959) demonstrated how Ricardo’s quadratic production function was implicit in his tables. Ricardo used it to predict the trend of rent’s distributive share as the economy approaches the stationary state (Blaug, 1985).

Johann von Thünen was perhaps the first economist who implicitly formulated the exponential production function as 

$$P = f(F) = A \prod_{i} (1 - e^{-a_{i}F_{i}})$$

where \(F_{1}, F_{2}\) and \(F_{3}\) are the three inputs, labour, capital and fertilizer, \(a_{i}\) are the parameters and \(P\) is the agricultural production. Lloyd (1969) provides a complete account of von Thünen’s exponential production functions and their derivation. He was also the first economist to apply the differential calculus to productivity theory and perhaps the first to use Calculus to solve economic optimization problems and interpret marginal productivities essentially as partial derivatives of the production function (Blaug, 1985). Mitscherlich (1909) and Spillman (1924) rediscovered von Thünen’s exponential production function.

In The Isolated State, vol-II, von Thünen wrote down the first algebraic production function in as 

$$p = hq^{n},$$

where \(p\) is output per worker (Q/L), \(q\) is capital per worker (C/L) and \(h\) is the parameter that represents fertility of soil and efficiency of labour. The exponent \(n\) is another parameter that lies between zero and unity. Multiplying both sides of von Thünen’s function by \(L\) (labour), we have 

$$Lp = h q^n L = hC^n L^{1-n} = P = \text{Output}.$$ 

Thus, we have the Cobb-Douglas production function hidden in von Thünen’s production function (Lloyd, 1969). The credit for presenting the first Cobb-Douglas function, albeit in disguised or indirect form, must go to von Thünen in the late 1840s rather than to Douglas and Cobb in 1928 (Humphrey, 1997). Further, von Thünen was uneasy to realize the implication of his production function in which labour alone cannot
produce anything. Hence he modified his function to \( P = h(L + C)^\alpha L^{\beta} \). This equation, which von Thünen estimated empirically for his own agricultural estate and which he declares he discovered only after more than 20 years of fruitless search, states that labor produces something even when unequipped with capital (Humphrey, 1997). It is surprising, however, that modern economists never formulate a production function in which labour alone can produce something.

Velupillai (1973) points out how Wicksell formulated his production function in 1900-01 that is identical to Cobb-Douglas function. In his 1923 review of Gustaf Akerman’s doctoral dissertation *Realkapital und Kapitalzins*, Wicksell wrote his function as \( P = e^{L^\alpha C^\beta} \) with the exponents adding up to unity.

Works of Turgot, von Thünen, and Wicksell might not have been known to Paul Douglas, but it is surprising to know that before his collaboration with Cobb, Sidney Wilcox, a research assistant of Douglas, had formulated in 1926 a production function of which the Cobb-Douglas function is only a special case (Samuelson, 1979). Wilcox’s production function was, perhaps, ignored by Douglas and till date it has remained in obscurity.

Likewise, what is now well known as the Leontief production function was formulated by Jevons, Menger and Leon Walras. In the Walrasian model of general equilibrium the proportions of output to inputs are fixed and no substitution among inputs is entertained.

Bertrand Russell tells us that certain academic achievements of individual scholars are a culmination of the research efforts of an entire epoch. Such achievements are superb in exposition though containing only a little of originality, and yet they are known to the posterity by the name of those scholars in whose work they found their final expression. These achievements set forth a paradigm and thus put a halt on a further progress of science for quite a long period. This is true of Aristotle, Newton and many others (Russell, 1984: p. 521). So, it is also true of Cobb-Douglas. A notable change in his formulation of production function came only in 1961 – after a gap of 33 years – with the work of Arrow, Chenery, Minhas and Solow (1961), which, however, is only an extension, not an alternative paradigm.

Of course, two generalizations of the Cobb-Douglas production function appeared in the literature before 1961. One of them is described as \( P = Ae^{\alpha K \beta L} \) and the transcendental production function of Halter et al. (1957) specified as \( P = Ae^{\alpha K \beta L} \); \( 0 < \alpha, \beta < 1 \). The first of these functions is a neoclassical production function only in a restricted region in the non-negative orthant of the (K, L) plane, the region in which the marginal products are nonnegative and diminishing marginal rate of substitution holds. When \( \alpha > 0, 0 < \beta < 1 \) the marginal product of labour is nonnegative only if \( \beta / \alpha \geq K / L \). Moreover, it does not contain a linear function as its special case. The linear production function is important in view of the Harrod-Domar fixed coefficient model of an expanding economy and therefore every neoclassical production function, the Cobb-Douglas or its generalizations, must contain the linear production function.
\( P = aK + bL \) to be consistent with it (Revankar, 1971). The transcendental function is a neoclassical production function if \( a, b < 0 \) and \( L \leq -\beta/b \) and \( K \leq -\alpha/a \) (Revankar, 1971). The transcendental function also does not contain a linear function. However, these functions allow for variable elasticity of substitution.

In the Cobb-Douglas production function the elasticity of substitution of capital for labour is fixed to unity – e.g. one percent for one percent. The production function formulated by Arrow et al. permitted it to lie between zero and infinity, but to stay fixed at that number along and across the isoquants - irrespective of the size of output or inputs (capital and labour) used in the production process. This function is well known as the Constant Elasticity of Substitution (CES) production function. It encompasses the Cobb-Douglas, the Leontief and the Linear production functions as its special cases.

Two mutually interrelated difficulties with the CES production function came to light very soon. The first is in the constancy of the elasticity of substitution (between inputs) along and across the isoquants and the second is in defining the said elasticity when more than two inputs are used in production. For three inputs there would be three elasticities (say, \( \sigma_{ij}, \sigma_{ik}, \sigma_{jk} \)) and for more inputs there would be many more. Uzawa (1962) and McFadden (1962, 1963) proved that it is impossible to obtain a functional form for a production function that has an arbitrary set of constant elasticities of substitution if the number of inputs (factors of production) is greater than two. Mathematical enunciations of these assertions are now known as the impossibility theorems of Uzawa and McFadden.

It is rather predictable what economists would do afterwards. That is: to find a functional form of (a neoclassical) production function that would permit variable elasticities of substitution (among different inputs) along and across the isoquants as well as larger number of inputs to be included in the recipe.

Mukerji (1963) generalized CES for constant ratios of elasticities of substitution. Bruno (1962) suggested a generalization of CES production function to permit the elasticity of substitution to vary. Liu and Hildebrand (1965) formulated a variable elasticity of production function as

\[
P = A[(1 - \delta)K^\eta + \delta K^{-\omega}L^{1-\omega}]^{1/\eta}.
\]

It reduces to CES for \( m=0 \). For this function, the elasticity of substitution is given as

\[
\sigma = (1 - \eta + m\eta/S_k)^{-1}
\]

where \( S_k \) is the share of capital. In 1968 Bruno formulated his Constant Marginal Share (CMS) production function

\[
P = AK^\alpha L^{1-\alpha} - mL \quad \text{or} \quad P/L = A(K/L)^\alpha - m,
\]

which implies that productivity of labour increases with capital-labour ratio at a decreasing rate. The CMS production function contains the linear production function and defines the elasticity of substitution,

\[
\sigma = 1 - [m\alpha/(1 - \alpha)](L/P).
\]

As the output-labour ratio increases (e.g. with economic growth), the elasticity of substitution in this function tends to unity and thus the CMS tends to the Cobb-Douglas production function. The CMS function has nonnegative marginal productivity of labour over the entire \( (K, L) \) orthant only if \( m \leq 0 \) (and \( 0 < \alpha < 1 \)). But then the elasticity of substitution will never be less than unity. This feature of the CMS function puts some limitation on it.
Lu and Fletcher (1968) generalized the CES production function to permit variable elasticity of substitution. Their “Variable Elasticity of Substitution” function is specified as

\[ P = A \left[ \delta K^\beta + (1 - \delta) L^{(1-\beta)} K^\beta \right]^{1/\beta}. \]

Assuming that competition has led to minimal cost conditions in production, the elasticity of substitution is obtained as

\[ \sigma = (1 + \beta)^{-1} [1 - c(1 + wL/(rK))] \]

where \( wL \) and \( rK \) are the share of labour and capital in output (net value added). It may be noted that the assumption of minimal cost conditions and use of factor shares in defining the elasticity of substitution limit the importance of this function in empirical investigation. Ryuzo Sato and Hoffman (1968) also introduced their variable elasticity of production function.

In his doctoral dissertation Revankar (1967) expounded his generalized production functions that permit variability to returns-to-scale as well as elasticity of substitution. In contrast with the production functions that (rather unrealistically) assume the same returns to scale at all levels of output, Zellner and Revankar (1969) found a procedure to generalize any given (neoclassical) production function with specified constant or variable elasticities of substitution such that the resulting production function retains its specification as to the elasticities of substitution all along but permits returns-to-scale to vary with the scale of output. Their Generalized Production Function (GPF) is given as

\[ P e^{\theta P} = c f^h \]

where \( f \) is the basic function (e.g. Cobb-Douglas, CES, etc) as the object of generalization, \( c \) is the constant of integration and \( \theta, h \) relate to parameters associated with the returns-to-scale function. In particular, if the Cobb-Douglas production function is generalized, we have

\[ P e^{\theta P} = AK^{\rho} L^{\rho(1-\alpha)} \]

This function is interesting from the viewpoint of estimation also. It has to be estimated so as to maximize the likelihood function since the Least Squares and Max Likelihood estimators of parameters do not coincide. The return to scale function is given by

\[ \rho(P) = \rho/(1+\theta P) \]

Depending on the sign of \( \theta \), the returns-to-scale function monotonically increases or decreases with increase in \( P \). However, as we know, the returns to scale first increases with output, remains more or less constant in a domain and then begins falling. This fact is not captured by the Zellner-Revankar function since it gives us a linear returns-to-scale function. Revankar (1971) presented his Variable Elasticity of Substitution (VES) production function as

\[ P = AK^{\rho(1-\alpha)} [L + (\mu - 1) K]^{\rho/\alpha} \]

with restriction on parameters:

\[ A, \rho > 0; \ 0 < \delta < 1; \ 0 \leq \delta \mu \leq 1 \text{ and } L/K > (1-\mu)/(1-\delta \mu). \]

The elasticity of substitution function is given as

\[ \sigma(K, L) = 1 + [(\mu - 1)/(1-\delta \mu)] K/L = 1 + \beta K/L. \]

Revankar’s VES does not contain the Leontief production function (while the CES does contain it), but it contains Harrod-Domar fixed coefficient model, the linear production function and the Cobb-Douglas function.

Brown and Cani (1963) had generalized the CES production to allow for non-constant returns to scale. Nerlove (1963) had generalized the Cobb-Douglas production function to allow for variable returns to scale. Ringstad (1967) advanced on the same line. Their production function is given as

\[ P^{(\ln + \ln P)} = A e^{K^\delta}; \ c \geq 0, \] that contains the Cobb-Douglas production function for \( c=0 \).

On the side of generalizing the CES production function to more than two factors of production, Uzawa (1962) made fundamental contributions. Further, Kazuo Sato
(1967) generalized the CES production function so as to incorporate more than two inputs in the recipe. To illustrate Sao’s schema, let \( P = f(x_1, x_2, x_3, x_4) \) where \( P \) is output and \( x_i \) is an input. Now combine \( x_1 \) and \( x_2 \) in a manner of CES to obtain \( Z_1 \) and similarly, combine \( x_3 \) and \( x_4 \) to obtain \( Z_2 \). At the second level, combine \( Z_1 \) and \( Z_2 \) to obtain \( P \). This schema would provide (constant) elasticities of substitution between \( x_1 \) and \( x_2 \) (say, \( \sigma_{12} \)) and between \( x_3 \) and \( x_4 \) (say, \( \sigma_{34} \)) at the first level and between \( Z_1 \) and \( Z_2 \) (say, \( s_{12} \)) at the second level. The nesting schedule of inputs would depend on the nature of inputs and production technology.

There are ample empirical evidences that suggest capital-skill complementarity (Griliches, 1969), or the wage differential between skilled and unskilled workers. It requires two types of labour (skilled and unskilled) to be separately dealt with in specifying the production function. Obviously, the elasticity of substitution of capital for skilled labour is very little – they are complementary to each other, while the case of unskilled labour is entirely different. To specify such models, the two-level CES production technology with capital, skilled labor and unskilled labor as inputs may be more suitable.

Diewert (1971) made two very important generalizations of production functions. First, he obtained a functional form that can incorporate many inputs in the recipe and the second that such a functional form permitted variable elasticities of substitution. He generalized the Leontief production function to attain an arbitrary set of Allen-Uzawa elasticities or shadow elasticities of substitution. He also obtained the generalized linear production function that can attain an arbitrary set of direct elasticities of substitution at a given set of inputs and input prices. Diewert’s generalized linear production function is:
\[
P = h \left( \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i^{1/2} x_j^{1/2} \right); \quad a_{ij} = a_{ji} \geq 0
\]
where \( h \) is a continuous, monotonically increasing function that tends to \(+\infty\) and has \( h(0) = 0 \). Restricting the sum of coefficients to unity, this function provides convexity to the isoquants and implies a well-behaved cost function.

Griliches and Ringstad (1971), Berndt and Christensen (1973) and Christensen, Jorgenson and Lau (1973) introduced their translog production functions that permit more than two inputs as well as variable elasticities of substitution (which is a necessity in view of Uzawa-McFaddan theorems). For \( n \) inputs \((x_i)\), the translog production function is specified as
\[
\ln(P) = a_q + \sum_{i=1}^{n} a_i \ln(x_i) + 0.5 \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} \ln(x_i) \ln(x_j); \quad b_{ij} = b_{ji}
\]
Unfortunately, this function is not invariant to the units of measurement of inputs and output (Intriligator, 1978). Further, on account of inclusion of \( \ln(x_i), \ln(x_j) \), their product and their squared values, estimation of parameters of this function often suffers from multicollinearity problem.

Kadiyala (1972) proposed a production function that includes CD, CES, Lu-Fletcher, Revankar and Sato-Hoffman production functions as its special cases. The function is defined as
In the function above, \( E(t) \) is the efficiency parameter, which also absorbs the neutral technical progress. Further, \( \beta_1 \) and \( \beta_2 \) bear the same sign as \( (\beta_1 + \beta_2) \). Kadiyala showed that for \( \omega_2 = 0 \) his function is CES, for \( \omega_2 = 0 \) it is Lu-Fletcher and for \( \omega_1 = 0 \) it is Sato-Hoffman production function. He also showed that the function may be generalized for more than two inputs, but to take care of the Allen or Uzawa-MacFadden partial elasticities of substitution one has to involve all the input ratios and the functional form may thus be quite lengthy and complicated. In spite of its generality, Kadiyala’s production function has remained in obscurity.

By the middle of the 1970’s, generalization of Cobb-Douglas and CES production functions (in their classical form) was almost complete. In the classical form, these functions assume that the marginal rate of substitution between any two factors of production is associated only with relative factor prices and it is independent of technical progress or level of output or, in technical terms, the technological progress is Hicks-neutral. To explain this concept a little more, we note that a technological change describes a change in the set of feasible production possibilities. A technological change is Hicks-neutral (Hicks, 1932) if the ratio of capital’s marginal product to labour’s marginal product is unchanged for a given capital-labour ratio. A technological change is said to be Harrod-neutral if the technology is labour-augmenting, and it is Solow-neutral if the technology is capital-augmenting. It is very easy to incorporate Hicks-neutral technical change into (a classical) production function. The production function \( P = f(x) \) is modifies as \( P = e^n f(x) \) where \( e^n \) captures the technological change that does not modify the elasticity of substitution between factors of production.

Ryuzo Sato (1975) observed that so far the marginal rate of substitution function had been specified as \( \ln(w/r) = \ln(a) + (1/\sigma) \ln(K/L) \) where \( w \) and \( r \) are the prices of labour (L) and capital (K), respectively, and \( \sigma \) is the elasticity of substitution. In view of the homotheticity assumption (assertion that a function \( g \) is a continuous positive monotonically increasing function of a homogenous function \( f \), such that if \( y = f(x) = f(x_1, x_2); \ z = g(y) = g(f(x)) = g(f(x_1, x_2)) \) then \( dg/dx > 0 \) ) implicit in the (classical versions of) production functions (Cobb-Douglas or CES, etc. discussed so far) the marginal rate of substitution was considered to be independent of the level of production or neutral technical change. Empirical data in many cases, however, suggested that the factor price ratio varies even at a constant input ratio. An introduction of factor-augmenting technical progress (of Harrod or Solow) fails to perform due to impossibility of identification of the bias (of technical progress) and substitution effect (Sato, 1970). Therefore, Sato (1975) relaxed the homotheticity assumption so that the level of output and the degree of neutral technical progress explicitly affect the factor combinations or \( \ln(w/r) = \ln(a) + (1/\sigma) \ln(K/L) + b \ln(P) + c \ln(T(t)), \) where \( P \) is the production and \( T(t) \) is the time dependent index of biased technical progress. From this specification he obtained the ‘most general’ class of CES function, of which the classical CES (as well as the Cobb-Douglas) and the ‘non-homothetic Cobb-Douglas” production functions are only special cases. Sato’s generalization permitted decomposition of income and substitution effects (in the factor market) and distinction between normal and ‘inferior’ inputs.
Sato’s CES function \( F(X_1, X_2, f) = C_1(f)X_1 + C_2(f)X_2 + H(f) = 0 \), where \( X_i = \delta x_i^\beta + \theta_i \) (for \( \sigma \neq 1 \)) or \( X_i = \delta_i \ln(x_i) + \theta_i \) (for \( \sigma = 1 \)), \( \delta \) and \( \theta \) are appropriate constants, \( \beta = (1 - \sigma)/\sigma \) for a non-unitary (and non-zero) elasticity of substitution \( (\sigma) \), \( x_i \) is a factor of production, and \( C_x(f), H(f) \) and \( f \) are defined appropriately according to homotheticity (or otherwise) and separability. Since all homothetic functions are separable, Sato obtained a three-fold classification of CES. First, when \( dC_i / df \equiv 0 \) and \( dH / df \neq 0 \), we obtain ordinary CES and Cobb-Douglas functions depending on whether \( \sigma \) is non-unitary or unitary. Secondly, when \( F(X_1, X_2, f) = -X_1 + C(f)X_2 = 0 \), the constants \( \theta_i = (1/\rho - \delta_i)\beta - \delta_i \) and \( f = V(X_1 / X_2) \), \( \beta \delta_i < 0, \beta \delta_2 > 0 \), where \( \rho \) is the non-homogeneity parameter, we have separable non-homothetic CES function or Cobb-Douglas function depending on whether \( \sigma \) is non-unitary or unitary (Sato, 1974). In this case, \( dC_i / df \neq 0, dH / df \neq 0 \) and \( C_i \neq mC_2 \) where \( m \) is a constant. Separable CES functions are linear solutions of \( F(.) \). Finally, we have non-separable CES functions as the non-linear solutions of \( F(X_1, X_2, f) = -X_1 + C(f)X_2 + H(f) = 0 \) in terms of \( f \) or \( C(f) \). In this case too, \( dC_i / df \neq 0, dH / df \neq 0 \) and \( C_i \neq mC_2 \) where \( m \) is a constant. Examples of non-separable CES are: if \( C(f) = af \) and \( H(f) = bf^2 \) then \( f \) is given by \([-aX_2 \pm \{(aX_2)^2 + 4bX_1\}^{1/2}]/(2b) > 0 \); if \( C(f) = af^2 \) and \( H(f) = bf \) then we have \( f = [-b \pm \{b^2 + 4aX_1X_2\}]/(2aX_2) > 0 \). Note that, in general, the function \( C(f) \) is:
\[
C(f) = \frac{\beta^{-1}x_1^\beta - \beta^{-1}x_1(h(f))^{-\beta}}{-\beta^{-1}x_2^\beta + \beta^{-1}x_2(h(f))^{-\beta}}, \quad \beta = (1 - \sigma)/\sigma, \quad \sigma \neq 1, \quad x_i \text{ are initial values of } x_i.
\]

Sato showed that his generalized CES might easily be extended to \( n \)-inputs case as well as variable elasticity of substitution. If the elasticity of substitution depends on the level of output then the ratio of factor prices \( (w/r) = (x_i / x_j)^{\sigma(f)}C(f), \quad \sigma(f) > 0, \quad C(f) > 0 \), in which case \( \beta = (1 - \sigma(f))/\sigma(f) \). The elasticity of substitution is constant along an isoquant but it varies across the isoquants, as output varies. It allows for a case when an isoquant in the \((x_1, x_2)\) plane may be a Cobb-Douglas or (an ordinary) CES but another isoquant can be a non-homothetic CES. For \( n \)-inputs case, define for any two inputs \( i \) and \( j \) the ratio of factor prices as \( \omega_{ij} \) and the elasticity of substitution between them as \( \sigma_{ij} = \sigma = \partial \ln(x_i / x_j)/\partial \ln(\omega_{ij}), \quad i \neq j, \quad i, j = 1, 2, \ldots, n \) then \( \omega_{ij} = (x_i / x_j)^{\sigma_{ij}C(f)}, \quad C(f) > 0 \). We have then \( F(X, f) = \sum_{i=1}^{n} C(f)X_i + H(f) = 0 \) where \( X_i = \delta x_i^\beta + \theta \) or \( X_i = \delta \ln(x_i) + \theta_i \) for a non-unitary or unitary \( \sigma \) respectively. In the \( n \)-inputs case we may permit variability to the elasticity of substitution across the isoquants by defining \( \sigma = \sigma(f) \) =constant at any \( f \). This grand generalization of the CES (and Cobb-Douglas) functions possibly concluded an era of investigations on this topic.

The importance of energy in the economic system was well stressed by Nicholas Georgescu-Roegen (1971), well known for versatility and competence in economics, mathematical sciences, physical sciences, life sciences and philosophy alike. However,
the energy crisis due to Jom-Kippur Iraq-Iran wars that threatened the U.S. and many other economies (in the mid 1970’s and thereafter) led many economists to formulate energy-dependent as well as other productions functions that included energy and materials (besides conventional labour and capital) as inputs (Tintner et al., 1974; Hudson and Jorgenson, 1974; Berndt and Wood, 1979). Kümmel (1982) and Kümmel et al. (1985; 1998/2000) introduced the linear exponential (LINEX) production function that was based on physics and technology as much as on economic considerations. It may be noted that the economists with technology or physics background have almost always pleaded for incorporation of technological considerations into production function, while the economists of the other category have often limited themselves to mere economic considerations. Tjalling Koopmans (1979) appears to be in agreement with an engineer’s view of economics, saying: “Economics is not dismal but incomplete. The things missed are very important” and a physical scientist saying: “Economists are technologically radical. They assume everything can be done.” By the way, it is worth noting that Koopmans was a physical scientist (as well as an economist) of high order (see Wikipedia-b). He is well known for his theorem in molecular orbital theory in quantum chemistry.

Hollis Chenery (1949, 1950) was, perhaps, the first economist to demonstrate how engineering information could be used to improve the empirical studies of production and to provide a bridge over the gap between the theoretical and empirical analysis of production (Wibe, 1984). In his paper Soren Wibe presents a detailed survey of engineering production functions, so we would not repeat them here.

We return to the linear exponential function (Lindenberger and Kümmel, 2002), which depends linearly on energy (E) and exponentially on quotients of capital (K), labor (L) and energy (E). The constant $P_0$ is a technology parameter indicating changes in the monetary valuation of the original basket of goods and services making up the output unit, P. All inputs and output are measured relative to some fixed quantity, $E_0$, $K_0$, $L_0$ and $P_0$ (and thus all of them are indices with a fixed base). Creativity-induced innovations and structural change make $a_0$, $c_0$, and $P_0$ time-dependent. In these variables and parameters the function is specified as $P = P_0 E \exp[a_0\{2-(L+E)/K\}+a_0c_0(L/E-1)]$.

Lindberger (2003) extended the above LINEX function to ‘service production functions’ that may be defined as $P = P_0 L(E/L)^{\alpha\beta\gamma} \exp[a_0\{2-(L+E)/K\}]$ or alternatively, $P = P_0 L\exp[a_0\{3-2(L/K)-(LE/K^2)\}+a_0c_m^2\{1-(L/E)\}]$ that emphasizes labour-dependence of service production, and yields (non-negative) production elasticities $\alpha = 2a_0(L/K)(E/K+1)$, $\gamma = a_0(c_m^2(L/E)-(LE/K^2))$; $\beta = 1-\alpha-\gamma$. One may estimate the function by any suitable algorithm (see Mishra, 2006).

**III. Multiple Output Production Function:** In the preceding narrative we have dealt with the case when a producing agent turns out a single product. Now we will turn to a case of multiple or joint production. It is worth mention that Kurz (1986) presents a very lively account of multiple output production function as visualized by the classical and the (early) neo-classical economists. Salvadori and Steedman (1988) review the concept in the Sraffian framework. A number of studies have been carried out that deal with this
topic. In particular, studies in agricultural (or farm) economics have addressed this problem more frequently (see Chizmar and Zak, 1983; Just et al., 1983; Mundlak, 1963; Mundlak and Razin, 1971; Weaver, 1983; Jawetz, 1961). Studies in the economics of household production and allocation of time between work and leisure (e.g. Becker, 1965; Pollak and Wachter, 1975; Gronau, 1977; Graham and Green, 1984) also have dealt with the joint production function. Of late, it has found favours in the upcoming field of environmental economics (Baumgärtner, 2004).

The first important question regarding a joint production function is its definition and existence. While in case of a single output technology, the production function was defined as the maximal output obtainable from a given input vector (Shephard, 1970), no simple maximal output existed for a multi-output technology, so that a multi-output production function could be defined and its existence proved readily. Building upon the work of Bol and Moechlin (1975), Al-Ayat and Färe (1977) examined necessary and sufficient conditions for the existence of a joint production function within a general framework of production correspondences. Without enforcing the strong disposability of inputs or outputs it was shown that a joint production function exists if and only if both input and output correspondences are strictly increasing along rays. Subsequently, Färe (1986) put forth three alternative ways in which multi-output production function might be defined. The first, as defined by Shephard (1970), it is an isoquant joint production function in the manner that for a given pair of input and output vectors, the input vector and the output vector belonged to the isoquant of input correspondence and the isoquant of output correspondence respectively. The second follows Hanoch (1970) in which an efficient joint production function characterizes input and output vectors that are simultaneously input and output efficient. The third concept of joint production might relate weakly efficient input vectors to weakly efficient output vectors.

The main finding of Färe is: let L(u) and L(v) denote all input vectors yielding output u and v (respectively), and P(x) and P(y) denote all output vectors obtainable by inputs x and y (respectively); then, for non-negative inputs and outputs (such that P(x) is not zero and L(u) is not null) a joint production exists if and only if the sets S{L(u)} and S{L(v)} are disjoint and the sets S{P(x)} and S{P(y)} are disjoint. More formally, a joint production function exists if the following conditions are satisfied:

\[
S\{L(u)\} \cap S\{L(v)\} = \phi, \quad u \mathcal{R} v
\]

\[
S\{P(x)\} \cap S\{P(y)\} = \phi, \quad x \mathcal{R} y
\]

\[
x, u \geq 0 \text{ such that } P(x) \neq \{0\}, \quad L(u) \neq \phi
\]

The relationship \( \mathcal{R} \) is \( \lambda x (\lambda \neq 1), \), >, and \( \geq \) as applicable. Weak disposability of inputs and outputs is assumed. Färe also proved that under certain conditions all the three definitions of joint production function are equivalent.

Econometric analysis of joint production perhaps dates back to the work of Klein (1947). Since then a number of studies have been conducted to estimate the parameters of multi-output or joint production functions. Methodologically those studies may be classified under four heads: those formulating process analysis models; those formulating simultaneous equations systems; those formulating composite macro function; those
formulating composite implicit macro function. Some important works are briefly reviewed as follows.

Mundlak (1963) approached estimation of joint production function through aggregation. His method lies in specifying the individual micro production function for each (joint) product as well as the manner of aggregating them to an analogous macro production function. The macro production function is then estimated and its relationship with the micro production functions is investigated. However, the possibilities of establishing the relationship among the macro and micro production functions depend on availability of information on allocation of inputs used for different (joint) products. Mundlak also proposed formulation and estimation of a general implicit production function. This led to his further work (Mundlak, 1964) in which he formulated the problem of estimation of multiple/joint production functions as an exercise in estimation of an implicit function. If \( X \) are inputs and \( Y \) are output then the implicit function \( g(f(X) - \phi(Y)) = 0 \) is expressed in terms of the composite input function \( f(X) \) and the composite output function \( \phi(Y) \). Mundlak illustrated his approach by the transcendental specification (proposed by Halter, et al., 1957) of the composite functions

\[
f(X) = a_0x_1^a x_2^b \exp(b_1x_1 + b_2x_2), \quad f(Y) = y_1^a y_2^b \exp(d_1y_1 + d_2y_2) \]  

and the simple implicit function \( g(X,Y) = f(X) - \phi(Y) = 0 \). It may be noted, however, that generally output is considered to contain errors due to specification of \( f(X) \) such that any output vector \( y_k = f(X) + u_k \) but inputs are considered non-stochastic. This consideration would lead to the specification \( g(f(X) - \phi(Y)) = \epsilon \) where \( \epsilon \) is the disturbance term. The least squares estimation of such functions has remained problematic. Mundlak and Razin (1971) also was basically an attempt to aggregation of micro functions to macro function.

Vinod (1968) addressed the problem of estimation of joint production function by Hotelling’s canonical correlation analysis (Hotelling, 1936; Kendall and Stuart, 1968). Later he improved his method to take care of the estimation problem if the data on output (of different products) or inputs were collinear (Vinod, 1976). His method summarily lies, first, in transforming the input vectors \( X \) and the output vectors \( Y \) into two composite (weighted linear aggregate) vectors, \( U = Xw \) and \( V = Y\omega \) respectively where the weights, \( w \) and \( \omega \), are (mathematically derived) such as to maximize the squared (simple product moment) coefficient of correlation between \( U \) and \( V \), and then transforming \( U \) and \( V \) back into \( X \) and \( Y \) respectively. He showed that the back transformation of the composite vectors \( U \) and \( V \) into \( X \) and \( Y \) poses no problem when the number of inputs is equal to or larger than the number of output. However, when that is not the case, one has to resort to some sort of least squares estimation (resulting from his suggested use of the least squares generalized inverse in the transformation process).

There were strong reactions to Vinod’s method of estimation of joint production functions [Chetty (1969), Dhrymes and Mitchell (1969), Rao (1969)]. Rao pointed out that to be economically meaningful the production function must be convex and the transformation curve concave. However, the method proposed by Vinod did not yield composite output function (transformation function) that satisfied these requirements. Dhrymes and Mitchell (much like Chetty) pointed out that Vinod’s formation was partly
erroneous and partly a “very complicated way of performing ordinary least squares.” If the ordinary least squares method applied to estimate each production function separately and independently (ignoring the fact that they relate to joint products) were inconsistent then so would be the canonical correlation method. While acceding to the errors pointed out by the critics, in his reply Vinod (1969) disagreed on the inconsistency issue shown to exist in his method and argued that the critics (Dhrymes and Mitchell) had to establish the necessity and would not merely put up some particular cases thereof. It is interesting, however, to note that Vinod undermined the role of a single counterexample in demolishing the mathematical property of a method.

Apart from the problems pointed out above, Vinod’s method cannot be useful when production functions are intrinsically nonlinear such that it is not possible to transform them (by some simple procedure such as log-linearization, etc) into linear equations. Secondly, it may not be correct to form the composite output function in Vinod’s manner. Thirdly, it is not necessary that the specification of production functions is identical for all products. It is possible that while one of the products follows the CES, another follows the nested CES (Sato, 1967) and yet another follows the Diewert (1971) or any other specification. However, it is possible to specify production functions of different types for different products and estimate their parameters jointly, although without making any composite output function (Mishra, 2007-b and 2007-c)

Since the early work of Manne (1958) process analysis has amply exhibited its ability to deal with the economics of joint products. However, it requires a large database and solving large programming models. Further, it precludes the calculation of price and substitution elasticities that may have important policy implications. Griffin (1977) used a method similar to process analysis supplementing it with pseudo data to ascertain appropriate types of production frontier functions for different joint products of petroleum refinery. A pseudo data point shows the optimal input and output quantities corresponding to a vector of input and output prices. By repetitive solution of the process model for alternative price vectors, the shape of the production possibility frontier may be determined. However, as pointed out by Griffin himself, the efficiency of pseudo data approach to estimation of joint production functions ultimately rests on the quality of the engineering process model often difficult for an economist to build or evaluate. Even then, this approach does not rule out the possibilities of aggregation bias completely.

Just et al. (1983) formulated and estimated their multicrop production functions as a system of nonlinear simultaneous equation model. The methods of estimation were nonlinear two-stage and three-stage least squares. Chizmar and Zak (1983) discussed the appropriateness of simultaneous equation modeling of multiple products raised or manufactured simultaneously. However, they held that in case of joint products the implicit form single equation modeling would be appropriate.

III. Aggregate Production Functions: Now we turn to the most turbulent area of research in the economics of production. As long as a production function describes a recipe that relates output and inputs of a firm, the matters are more or less straight. However, when one makes an attempt to identify such a recipe at the industry, sector or economy level, the matters may be quite different. An industry is generally made up of
numerous firms producing similar products in which each firm uses inputs in its own accord with its own cost, returns-to-scale and market implications. A production function (or a cost curve for that matter) at an industry level is obtained by using the quantities of inputs and output aggregated over the constituent firms. The question arises: does this aggregate production function represent the core technological relation between inputs and output of a majority of firms? Or does it represent the production function of a “Platonic” or “archetypal” firm that might not be a real firm, but the empirically observed real firms are only the imperfect instances of that archetypal firm? A little further, a firm might not possess certain characteristics but an industry may possess them. A firm might be a price-taker in the factor market, but an industry might be a price-maker. Rigidities might not permit a real firm to adjust to the changes at the industry level and all input-output relations at the firm level would be sub-optimal. Deliberate keeping of room for adjustments and maneuvers and perpetual X-efficiency may invalidate the assumption of technical efficiency and vitiate a search for allocative efficiency. These issues and many others turn out to be more and more significant when one moves higher to macroeconomic levels.

Since Adam Smith (or even before him) economists of a particular hue have largely been preoccupied with intellectual work to prove that an autonomous management (and organization) of a society by capitalistic principles (characterizing self-interest guided agents operating in an institutional framework of private property, market economy, competition, accumulation of capital, etc) is the best, just, stable and most viable among all possible approaches to manage (and organize) a society. A society organized on that principle would grow (expand) indefinitely and deliver justice to all the agents. However, Karl Marx questioned the efficacy of the capitalistic system in not only delivering justice but also guaranteeing indefinite expansion and stability. After Marx, the (neoclassical) economists, therefore, had to invent strong arguments to defend the legitimacy, efficacy and supremacy of the capitalistic system. The Walrasian general equilibrium, Pareto-optimality of the competitive economy, aggregate production function, marginal productivity theory of distribution, product exhaustion theorem, and ultimately Harrod’s, Solow’s and von Neumann’s paths to expansion are only some major lemmas to prove the said grand proposition.

Whitaker (1975) brings to the light some unpublished works of Alfred Marshall and points out how Marshall formulated his aggregate production function $P = f(L,E,C,A,F)$, where $L$ is labour and $E$ its efficiency, $C$ is Capital, $A$ is level of technology and $F$ is fertility of soil. In this aggregate production function Marshall used the time derivative of the variables and therefore it may be regarded as the first neoclassical growth model, foreshadowing the Tinbergen-Solow growth model.

Knut Wicksell made significant contribution to demonstrate that nonhomogeneous production functions for firms are perfectly compatible with a linear homogeneous function for the entire industry. Suppose industry output expands and contracts through the entry and exit of identical firms, each operating at the same minimum unit cost. The result is to trace out a horizontal long-run industry supply curve that looks like it came from a constant-returns production function. Thus, he justified the
use of aggregate linear homogeneous functions such as the Cobb-Douglas function (Humphrey, 1997).

The Cobb-Douglas production function gave a readily convincing proof that in competitive equilibrium all inputs are paid their marginal product (and hence their respective real price), the entire product exhausts (as the sum of input elasticities of product sum up to unity), the constant returns to scale prevails, and the empirically observed constancy of relative shares of factors of production for long periods is fully explicable, justified and natural. This finding strengthened the foundations of using aggregate production function at the macroeconomic level. Humphrey (1997) has presented this development so lucidly that it is best to quote him: “Each stage saw production functions applied with increasing sophistication. First came the idea of marginal productivity schedules as derivatives of a production function. Next came numerical marginal schedules whose integrals constitute particular functional forms indispensable in determining factor prices and relative shares. Third appeared the path-breaking initial statement of the function in symbolic form. The fourth stage saw a mathematical production function employed in an aggregate neoclassical growth model. The fifth stage witnessed the flourishing of microeconomic production functions in derivations of the marginal conditions of optimal factor hire. Sixth came the demonstration that product exhaustion under marginal productivity requires production functions to exhibit constant returns to scale at the point of competitive equilibrium. Last came the proof that functions of the type later made famous by Cobb-Douglas satisfy this very requirement. In short, macro and micro production functions and their appurtenant concepts—marginal productivity, relative shares, first-order conditions of factor hire, product exhaustion, homogeneity and the like—already were well advanced when Cobb and Douglas arrived”.

During the post-Great-Depression period until the end of the 2nd World war economists investigated into the possibilities of growth without violent fluctuations. In this investigation the aggregate production function proved to be very useful. The Cobb-Douglas production function was found quite amenable to incorporation of technical change introduced into the production system from time to time without altering the basic conclusions on factor shares. The growth model of von Neumann (1937) deviated from using the Cobb-Douglas production function but retained the practice of aggregation. Consequent upon the development of linear programming as a method of optimization this line of investigation progressed very rapidly. Activity analysis of Koopmans, input-output analysis of Leontief, aggregate linear production function of Georgescu-Roegen (1951), separation theorems and generalization of von Neumann’s model by Gale (1956), careful proofs given by Nikaido (1968) all strengthened the foothold of aggregate production function in economic analysis.

However, the 1950’s did not welcome the aggregate production function wholeheartedly. Mrs. Joan Robinson (1953) viewed aggregate production function with a remark: “...the production function has been a powerful instrument of miseducation. The student of economic theory is taught to write \( Q = f(L, K) \) where \( L \) is a quantity of labor, \( K \) a quantity of capital and \( Q \) a rate of output of commodities. He is instructed to assume all workers alike, and to measure \( L \) in man-hours of labor; he is told something
about the index-number problem in choosing a unit of output; and then he is hurried on to
the next question, in the hope that he will forget to ask in what units $K$ is measured.
Before he ever does ask, he has become a professor, and so sloppy habits of thought are
handed on from one generation to the next.” A controversy began, especially regarding
the measurement of capital, in which Piero Sraffa, Joan Robinson, Luigi Pasinetti and
Pierangelo Garegnani (among others) argued against the use of aggregate production
function, and Paul Samuelson, Robert Solow, Frank Hahn and Christopher Bliss (among
others) argued in favour of using aggregate production function for explaining relative
factor shares. As argued by the first group, it is impossible to conceive of an abstract
quantity of capital which is independent of the rates of interest and wages. However, this
independence is a precondition of constructing an isoquant (or production function). The
iso-quant cannot be constructed and its slope measured unless the prices are known
beforehand, but the protagonists of aggregate production function use the slope of the iso-
quant to determine relative factor prices. This is begging the question.

This controversy (that began in 1953) lasted until the mid 1970’s. A very lively
account of it was presented by Harcourt (1969), who was himself a participant in the
controversy (from the Robinson’s side), and as Joseph Stiglitz (1974) notes, may not
possibly have been impartial in giving an unbiased account. However, it may be noted
that Stiglitz admits that he was a participant from the other side. A much more
comprehensive account of controversy may be found in Cohen and Harcourt (2003).

The controversy exposed all aggregate production functions, the Cobb-Douglas
production function in particular, and proved almost conclusively that it has no
economics in it as its properties stem from mere algebra. A series of three papers by
Anwar Shaikh on his “humbug production function” should have ousted out the Cobb-
Douglas production function from all serious endeavours in economic analysis. Felipe,
Fisher and their associates in a number of papers have exposed the validity of an
aggregate production function.

Once Paul Samuelson (1966) wrote: “Until the laws of thermodynamics are
repealed, I shall continue to relate outputs to inputs - i.e. to believe in production
functions. Unless factors cease to have their rewards to be determined by bidding in
quasi-competitive markets, I shall adhere to (generalized) neoclassical approximations in
explaining their market remunerations.” One fails to understand as to how the validity of
laws of thermodynamics and relating outputs to inputs entails validity of the proposition
that the aggregate of functions would be the function of aggregates, and if there is some
particular type of function that has this property then how that particular function is the
correct function describing the production technology of any economy and the relative
factor shares. Solow (1957) made an impressive empirical study to demonstrate how the
aggregate production function fits to the U.S. data for 1909-49, which confirm neutral
technical change, shift in production function, and therefore validate the artifact of
aggregate production function as a powerful tool of analysis. Samuelson (1962) invented
the so-called ‘surrogate production function’ which related the total quantity of output per
capita to the total quantity of capital per capita, which can be used to predict all
behaviour in the sense of the wage and profit rates that would prevail in different long run
equilibria, which, according to Donald Harris (1973), was based on too extremely restrictive assumptions and therefore was too fragile to hold true in reality.

Non-reswitching theorem was a major plank to support measurability of capital and its aggregation. Levhari and Samuelson published a paper which began, 'We wish to make it clear for the record that the nonreswitching theorem associated with us is definitely false. We are grateful to Dr. Pasinetti…' (Levhari and Samuelson 1966). The final outcome of the controversy may be best concluded in the words of Burmeister (2000): “… the damage had been done, and Cambridge, UK, 'declared victory': Levhari was wrong, Samuelson was wrong, Solow was wrong, MIT was wrong and therefore neoclassical economics was wrong. As a result there are some groups of economists who have abandoned neoclassical economics for their own refinements of classical economics. In the United States, on the other hand, mainstream economics goes on as if the controversy had never occurred. Macroeconomics textbooks discuss 'capital' as if it were a well-defined concept - which it is not, except in a very special one-capital-good world (or under other unrealistically restrictive conditions). The problems of heterogeneous capital goods have also been ignored in the 'rational expectation revolution’ and in virtually all econometric work."

The effects of ‘capital controversy’ beginning with a criticism of using aggregate production function for explaining (and justifying) relative factor shares had a disastrous effect on the neoclassical economics. Lavoie (2000) observed: “Capital reversing renders meaningless the neoclassical concepts of input substitution and capital scarcity or labor scarcity. It puts in jeopardy the neoclassical theory of capital and the notion of input demand curves, both at the economy and industry levels. It also puts in jeopardy the neoclassical theories of output and employment determination, as well as Wicksellian monetary theories, since they are all deprived of stability. The consequences for neoclassical analysis are thus quite devastating. It is usually asserted that only aggregate neoclassical theory of the textbook variety - and hence macroeconomic theory, based on aggregate production functions - is affected by capital reversing. It has been pointed out, however, that when neoclassical general equilibrium models are extended to long-run equilibria, stability proofs require the exclusion of capital reversing. In that sense, all neoclassical production models would be affected by capital reversing.

Gehrke and Lager (2000) observed: "These findings destroy, for example, the general validity of Heckscher-Ohlin-Samuelson international trade theory …, of the Hicksian neutrality of technical progress concept …, of neoclassical tax incidence theory …, and of the Pigouvian taxation theory applied in environmental economics …”.

What has remained with the believers in neoclassical aggregate production function is the complaint that neither Joan Robinson nor her associates developed an alternative set of theoretical (as opposed to descriptive) tools that avoid her concerns about the limitations of equilibrium analysis.

IV. Concluding Remarks: It is said that once Stanislaw Ulam very earnestly sought for an example of a theory in social sciences that is both true and nontrivial. Paul Samuelson, after several years supplied one: the Ricardian theory of comparative advantages. If this
is true then it speaks enough about the position of ‘production function’ in economics. Anwar Shaikh almost conclusively proved that the properties of the Cobb-Douglas production function are devoid of any economic content; they stem from the algebraic properties of the function. Measurement of capital was put in jeopardy by the capital controversy, not only for aggregate production function, but also for any production function even at the firm level. If one has interacted with engineers and the managers who are decision makers at certain level, one might know their view of the utility of production functions. Possibly, Mrs. Joan Robinson was right to comment that the production function has been a powerful instrument of miseducation. The student of economic theory is taught to write \[ Q = f(L, K) \] and that is its sole utility. The era of classical economics had ended (sometime in 1870’s) with the criticism of it by Marx and the birth of neoclassicism. The era of neoclassical economics possibly ended with the capital controversy sometime in 1970’s.

References

• Hicks, J.R. (1932), The Theory of Wages, 2nd Ed (1963), St Martin’s Press, NY.
• Klein, L.R. (1947) *The Use of Cross-Section Data in Econometrics with Application to a Study of Production of Railroad Services in the United States*, (Mimeographed) National Bureau of Economic Research, Washington, DC.


Revankar, N.S. (1967) Production Functions with Variable Elasticity of Substitution and Variable Returns to Scale, Doctoral Dissertation, Univ. of Wisconsin, USA.


• Thünen, J.H. von (1930) *Der isolierte Staat in Beziehung auf Landwirthschaft und Nationalökonomie*. 3 volumes: Fischer, Jena, Germany.

22