ISSN 0111-1760

How Sound are the Foundations of the Aggregate Production Function?

J Felipe & JSL McCombie

No. 0116

Jesus Felipe is at the Georgia Institute of Technology and JSL McCombie is Fellow and Director of Studies of Economics, Downing College, Cambridge. We are grateful, as usual, for the comments and encouragement of Geoff Harcourt. We are also grateful for the comments received at seminars given at the University of Otago and the Victoria University of Wellington, New Zealand.
Email: jesus.felipe@inta.gatech.edu *and* jslm2@cam.ac.uk

September 2001

Department of Economics
School of Business
University of Otago
PO Box 56
Dunedin
NEW ZEALAND
Ph: 00 64 3 479 8725
Fax: 00 64 3 479 8174
Email: economics@otago.ac.nz

How Sound are the Foundations of the Aggregate Production Function?

J Felipe[1]

&

JSL McCombie

Abstract: The aggregate production function has been subject to a number of criticisms ever since its first empirical estimation by Cobb and Douglas in the 1920s, notably the problems raised by aggregation and the Cambridge Capital Theory Controversies. There is a further criticism due initially to Phelps Brown (and elaborated, in particular, by Simon and Shaikh) which is not so widely known. This critique is that because at the aggregate level only value data can be used to estimate production function, this means that the estimated parameters of the production function are merely capturing an underlying accounting identity. Hence, no reliance can be placed on estimates of, for example, the elasticity of substitution as reflecting technological parameters. The argument also explains why good statistical fits of the aggregate production functions are obtained, notwithstanding the difficulties posed by the aggregation problem and the Cambridge Capital Controversies noted above. This paper outlines and assesses the Phelps Brown critique and its extensions. In particular, it considers some possible objections to his argument and demonstrates that they are not significant. It is concluded that the theoretical basis of the aggregate production function is problematic.

JEL Classification: O3 O4

Keywords: Cobb-Douglas, production function, income identity

Introduction

It is somewhat paradoxical that one of the concepts most widely used in macroeconomics, namely the aggregate production function, is the one whose theoretical rationale is perhaps most suspect. The serious problems raised by the Cambridge Capital Theory Controversies dominated “high theory” in the late 1960s and early 1970s, and eventually led to an agreement that reswitching and capital reversing were theoretically possible (Harcourt, 1972). This posed serious problems for the justification of the use of the neoclassical one-sector aggregate production function as a “parable”. The revival of interest in growth theory with the development of endogenous growth theory is still squarely in the tradition of the neoclassical

[1] Jesus Felipe is at the Georgia Institute of Technology and JSL McCombie is Fellow and Director of Studies of Economics, Downing College, Cambridge. We are grateful, as usual, for the comments and

growth model. Pasinetti (1994) was compelled to remind the participants at a recent conference on economic growth that

> this result [that there is no unambiguous relationship between the rate of profit and the capital-labour ratio], however uncomfortable it may be for orthodox theory, still stands. Surprisingly, it is not mentioned. In almost all 'new growth theory' models, a neoclassical production function, which by itself implies a monotonic inverse relationship between the rate of profits and quantity of capital per man, is simply *assumed*. (emphasis in the original)

If this were not enough, there is a whole series of further problems concerned with the question of whether or not micro-production functions can be aggregated to give a macro-relationship which can be shown to reflect the underlying technology of the economy in some meaningful way (Walters, 1963, Fisher, 1987, 1992). Indeed, Blaug (1974), who can be scarcely viewed as sympathetic to the Cambridge UK view of the interpretation, or importance, of the Capital Theory Controversies, nevertheless considers that the aggregation problem effectively destroys the rationale of the aggregate production function. "Even if capital were physically homogeneous, aggregation of labour and indeed aggregation of output would still require stringent and patently unrealistic conditions at the economy–level". Moreover, "the concept of the economic meaningful aggregate production function requires much stronger and much less plausible conditions than the concept of an aggregate consumption function. And yet, undisturbed by Walter's conclusions or Fisher's findings, economists have gone on happily in increasing numbers estimating aggregate production functions of even more complexity, barely halting to justify their procedures or to explain the economic significance of their results." (Blaug, 1974).

The defence of this procedure, however, has been eloquently put forward by Solow (1966), who can hardly be accused of not being fully aware of the aggregation problem. "I have never thought of the macroeconomic production function as a rigorously justifiable concept. … It is either an illuminating parable, or else a mere device for handling data, to be used so long as it gives good empirical results, and to be abandoned as soon as it doesn't, or as soon as something else better comes along" (Solow, 1966). Wan (1971) reflects this view when he argues that Solow's (1957) path-breaking approach may be defended on the grounds that "one may argue that the functional relation between Q and K, L is an 'empirical law' in its own right. In the methodological parlance of Samuelson, this is an operationally meaningful law, since it can be empirically refuted". (Q, K and L are output, capital, and labour.) From the first studies by Douglas and his various collaborators (see, for example, Cobb and Douglas, 1928, Douglas, 1944 and 1976) it has often been found that the aggregate production function

---

encouragement of Geoff Harcourt. We are also grateful for the comments received at seminars given at

gives a close statistical fit, especially using cross-sectional data, with the estimated output elasticities close to the factor shares.

However, the problem is that these studies generally could *not* have failed to find a close correspondence between the output elasticities and the factor shares. This arises from the fact that ideally the production function is a microeconomic concept, specifying the relationship between physical outputs and inputs (such as numbers of widgets, persons employed, and identical machines). On the other hand, at the aggregate level, constant price value data are used for capital and output.[2] Yet this is not an innocuous procedure, as it has been argued that this undermines the possibility of empirically testing the aggregate production function. As Simon (1979a) pointed out in his Nobel Memorial lecture, the good fits to the Cobb-Douglas production function "cannot be taken as strong evidence for the classical theory, for the identical results can readily be produced by mistakenly fitting a Cobb-Douglas function to data that were in fact generated by a linear accounting identity (value of output equals labor cost plus capital cost)".

Specifications of aggregate production functions, using value data, may be nothing more than approximations to an accounting identity, and hence can convey no information, *per se*, about the underlying technology of the "representative firm". This is not a new critique, but first came to prominence in Phelps Brown's (1957) criticism (later formalised by Simon and Levy (1963)) of Douglas's cross-industry results. Shaikh (1974, 1980, 1987) generalised it to time-series estimation of production functions and Simon (1979b) also considered the criticism in the context of both cross-section and time-series data. The criticism was re-examined and extended by Felipe and McCombie (2000, 2001 a&b, 2002 a&b), Felipe (2001a and 2001b), McCombie (1987, 1998, 2000, 2000-2001, 2001), McCombie and Dixon (1991) and McCombie and Thirlwall (1994).

Once it is recognised that all that is being estimated is an underlying identity, it can be shown how it is always possible, with a little ingenuity, to obtain a perfect statistical fit to a putative production function, which exhibits constant returns to scale and where the estimated "output elasticities" equal the factor shares. It can also shown how the results of estimation of production functions which find increasing returns to scale and externalities are simply due to misspecification of the underlying identity and the estimated biased coefficients may actually be predicted in advance (McCombie, 2000-2001, Felipe, 2001a).

---

the University of Otago and the Victoria University of Wellington, New Zealand.

[2] It should be noted that apart from a very few "engineering" production function studies, all other estimations of production functions have used value data, even at low levels of aggregation, for example, the 3 or 4 digit SIC.

The purpose of this paper is to provide a survey of the main elements of this critique and also to consider some counter-criticisms which have been made.[3] We conclude that the latter leave the central tenet unaffected.

The Cobb-Douglas Production Function and the Accounting Identity

The problem that the accounting identity poses for the interpretation of the aggregate production function was first brought to the fore by Phelps Brown (1957) in his seminal paper "The Meaning of the Fitted Cobb-Douglas Function".[4] (It had also been partly anticipated by Bronfenbrenner (1944).) It is one of the ironies of the history of economic thought that this article, which challenged the whole rationale for estimating aggregate production functions, was published in the same year as Solow's (1957) "Technical Change and the Aggregate Production Function". The latter, of course, was largely responsible for the beginning of the neoclassical approach to the empirical analysis of growth.

Phelps Brown's critique was addressed to the fitting of production functions using cross-sectional data and was specifically directed at Douglas's various studies (see Douglas, 1944), and we consider this first, before considering the case of time-series data.

Cross-Section Data

The fact that the crucial tenet of Phelps Brown's (1957) argument was presented rather obscurely and was buried in his paper did not help its reception, even though it was published in one of the US's leading economics journal. The following is the key passage:

> The same assumption would account for the observed agreement for the values obtained for α [the output elasticity of labour], and the share of earnings given by the income statistics. For on this assumption the net products to which the Cobb-Douglas is fitted would be made up of just the same rates of return to productive factors, and quantities of those factors, as also make up the income statistics; and when we calculate α by fitting the Cobb-Douglas function we are bound to arrive at the same value when we reckon up total earnings and compare them with the total net product. In α we have a measure of the percentage change in net product that goes with a 1 per cent change in the intake of labour, when the intake of capital is held constant; but when we try to trace such changes by comparing one industry with another, and the net products of the two industries approximately satisfy, $V_i = wL_i + rJ_i$, the difference between them will always approximate to the compensation at the wage rate w of the difference in labour intake.

[3] Some of these criticisms are not referenced as they were mostly made by way of private communication with the authors. Lest it be thought we are setting up straw horses, the sources are available on request from the authors.

[4] We define an aggregate production function as one where output is a value, rather than a physical, measure, regardless of the precise unit of observation, i.e., the exact level of the Standard Industrial Classification.

> The Cobb-Douglas α and the share of earnings in income will be only two sides of the same penny. (Phelps Brown, 1957, p.557)

V, *w*, *L*, *r*, and *J* are output (value added at constant prices), the average wage rate, the labour input, the average observed rate of return and the constant price value of the capital stock.[5] (We use *V* and *J* to refer to the value measures; *Q* and *K* are used below to denote the physical measures of output and capital.) The subscript *i* denotes the *i*th firm or industry.

The argument therefore seems to be this. The output elasticity of the Cobb-Douglas production function is defined as $\alpha = (\partial V/\partial L)(L/V)$ and given the assumptions of the neoclassical theory of factor pricing, the marginal prodiuct of labour equals the wage rate, $\partial V/\partial L = w$. Given these assumptions, it also follows that $\alpha = a = wL/V$, the share of labour in value added.

However, as Phelps Brown pointed out in the quotation cited above, there is also an accounting identity that defines the measure of value added for all units of observation, whether they be the firm, the 1, 2, 3 or 4 digit SIC, or the whole economy. This is given by:

$$V_i = wL_i + rJ_i \qquad (1)$$

It should be emphasised that there are no behavioural assumptions underlying this equation, in that it is compatible with any degree of competition, increasing or decereasing returns to scale and the existence or not of a well-baehaved underlying production function. J is the constant price measure of the value of the capital stock (normally calculated by the perpetual inventory method) and *r* is the observed rate of profit normally calculated as the product of the share of profits in value added $(1-\alpha)$ and the output-capital ratio, i.e., $r = (1-\alpha)V/K$

Consequently, partially differentiating the accounting identity, $V_i = wL_i + rJ_i$, with respect to *L* gives $\partial V/\partial L = w$ and it follows that $(\partial V/\partial L)/(L/V) = a = wL/V$. The argument stemming from the identity has not made any economic assumptions at all (e.g., it does not rely on the marginal productivity theory of factor pricing or the existence of perfectly competitive markets and optimising behaviour of firms). Consequently, the finding that the putative output elasticities equal the observed factor shares cannot be taken as a test that factors of production are paid their marginal products. This is a position, however, that was not accepted by Douglas (1976) himself. "A considerable body of independent work tends to corroborate the original Cobb-Douglas formula, but more important, the approximate coincidence of the estimated coefficients with the actual shares received also strengthens the

[5] The notation has been changed to make it consistent with the rest of this paper.

competitive theory of distribution and disproves the Marxian." However, it is noticeable that, in his survey, Douglas fails to mention the Phelps Brown (1957) paper.

If the output elasticity of labour and the share of labour's total compensation are merely "two sides of the same penny", could it be that the Cobb-Douglas is simply an alternative way of expressing the income identity and, as such, has no implications for the underlying technology of the economy? This was the proposition that Simon and Levy (1963) proved some eight years later.

Following Simon and Levy (1963) and Intriligator (1978), the isomorphism between the Cobb-Douglas production function and the underlying accounting may be simply shown. The Cobb-Douglas, when estimated using cross-section (firm, industry or regional) data, is specified as:

$$V_i = A L_i^{\alpha} J_i^{\beta} \qquad (2)$$

where A is a constant.

Equation (2) may be expressed as:

$$ln\frac{V_i}{\overline{V}} = \alpha\, ln\frac{L_i}{\overline{L}} + \beta\, ln\frac{J_i}{\overline{J}} \qquad (3)$$

where $\overline{V}, \overline{L},$ and $\overline{J}$ are the values of some reference observations, such as those of the average firm or the base year.

The following approximation holds for any variable X, when X and $\overline{X}$ do not greatly differ:

$$ln\frac{X}{\overline{X}} \approx \frac{X}{\overline{X}} - 1 \qquad (4)^{6}$$

Consequently, equation (2) may be written as:

---

[6] The reason for this is straightforward. $ln\frac{X}{\overline{X}}$ is the exponential growth or proportionate rate of change of X over the period, or distance, between the values and is approximately equal to $\Delta X/\overline{X}$, where $\Delta X \equiv X - \overline{X}$ and so $ln\frac{X}{\overline{X}} \approx \frac{X}{\overline{X}} - 1$. Alternatively, the approximation may be derived from a Taylor series expansion.

$$\frac{V_i}{\overline{V}} \approx \alpha \frac{L_i}{\overline{L}} + \beta \frac{J_i}{\overline{J}} + (1 - \alpha - \beta) \qquad (5)$$

or as:

$$V_i = \left(\alpha \frac{\overline{V}}{\overline{L}} \right) L_i + \beta \left(\frac{\overline{V}}{\overline{J}} \right) J_i + (1 - \alpha - \beta) \overline{V} \qquad (6)$$

A comparison with the income identity, namely, $V_i \equiv wL_i + rJ_i$, shows that $w = \alpha \frac{\overline{V}}{\overline{L}}$ or $w \frac{\overline{L}}{\overline{V}} = a = \alpha$. A similar relationship holds between the output elasticity of capital and capital's share. Moreover, $(1\text{-}\alpha\text{-}\beta)$ equals zero, so that the data will always suggest the existence of constant returns to scale, whatever the true technological relationship.

What is the implication of all this? Start from the accounting identity and undertake the approximation in the reverse order from that outlined above. The two procedures are formally equivalent. The consequence of this argument is that, for reasonably small variations of L and J and with w and r constant (the last two are not essential, as we shall see below), a Cobb-Douglas multiplicative power function will give a very good approximation to a linear function. Since the linear income identity exists for *any* underlying technology, we cannot be sure that all that the estimates are picking up is not simply the identity. The fact that a good fit to the Cobb-Douglas relationship is found implies nothing, *per se*, about such technological parameters as the elasticity of substitution.[7]

To see this consider Figure 1, which shows the accounting identity expressed as $V_i/L_i = w + rJ_i/L_i$ and the Cobb-Douglas relationship as $V_i/L_i = A(J_i/L_i)^a$. The observations must lie exactly on the accounting identity. We have assumed, for the moment, that w and r are constant. (If they show some variation, then the observations would be a scatter of points around the line where the slope and the intercept represent some average value of r and w.) The Cobb-Douglas approximation is given by the solid curved line, *cd*, which is tangent to the income identity, *ab*. If, however, we mistakenly statistically fit a Cobb-Douglas function to these data, we will find the best fit depicted by the dotted curved line, *ef*. The residuals will be autocorrelated.

[Figure 1 about here]

[7] Bronfenbrenner (1944) seemed to advance the same argument except that his cost identity expressed the costs facing the firm in the absence of abnormal profits. Marshak and Andrews (1944) argued that the zero profits equation would have a random term because the wiping out of profits and losses is hardly likely to be instantaneous. Consequently, they argue that the equation "is certainly subject to strong fluctuations from firm to firm. There is, therefore, certainly little likelihood that a function fitted, e.g., by the least squares, would give an approximation of [the cost function]." Of course, once it is appreciated that the definition of value added includes any monopoly profits and is exact for all time periods, then Marshak and Andrews reservations are no longer justified.

Of course, the empirical question arises as to how much variation in the data is required before the Cobb-Douglas ceases to give a plausible fit to the data. Fortunately, Simon (1979b) has provided the answer. He calculated the ratio between the predicted value given by the Cobb-Douglas function (V_{CD}) and that by the accounting identity, V_A, namely, $V_{CD}/V_A = A(L/J)^{\alpha}/(r + w(L/J))$. He found that when the L/J ranged from 16 to 1, the greatest error, or ratio, was only 10 per cent. He concluded "since in the data actually observed, most of the sample points lie relatively close to the mean value of L/J, we can expect average estimating errors of less than 5 per cent." (Simon 1979b).

The good approximation of the Cobb-Douglas to the accounting identity is also likely to carry through even when we allow w and r to change, provided the factor shares do not show very much variation. To see this, assume a continuum of firms and differentiated the accounting identity to give:

$$dV_i = (dw_i)L_i + w_i dL_i + (dr_i)J_i + r_i\, dJ_i \tag{7}$$

or

$$dV_i/V_i = a_i dw_i/w_i + a_i dL_i/L_i + (1-a_i)dr_i/r_i + (1-a_i)\, dJ_i/J_i \tag{8}$$

Let us assume factor shares are constant (and there are many reasons why this should occur other than because there is a Cobb-Douglas production function, e.g. firms pursue a constant mark-up pricing policy). Equation (8) may be integrated to give:

$$V_i = Bw_i^{a} r_i^{(1-a)} L_i^{a} J_i^{(1-a)} \tag{9}$$

where B is the constant of integration.

Provided that $w_i^{a} r_i^{(1-a)}$ shows very little variation or is orthogonal to $L_i^{a} J_i^{(1-a)}$ or both, the putative Cobb-Douglas production function will once again give a very good fit to the data.

It should be noted that this argument is not just confined to the Cobb-Douglas production function. Simon (1979b) explicitly considers the CES production function given by $V = \gamma(\delta L^{-\rho} + (1-\delta)J^{-\rho})^{-(1/\rho)}$, where γ, ρ, and δ are parameters. He argues that if the true relationship were given by the accounting identity and we were mistakenly to estimate the CES production function, then if ρ goes to zero, the function becomes a Cobb-Douglas. He cites Jorgenson (1974) as suggesting that most estimates give ρ close to zero and so the argument still applies. However, more recent studies find that the putative aggregate elasticity of substitution is less than unity. But the argument is more general than Simon implies. If we were to express any production function of the form $V_i = f(L_i, J_i)$ in proportionate rates of change, we would find that $dV_i/V_i = c + \alpha_i dL_i/L_i + \beta_i\, dJ_i/J_i$ which is formally exactly equivalent to the accounting identity, provided $a_i dw_i/w_i + (1\text{-}a_i)dr_i/r$ is again roughly constant or orthogonal to $a_i dL_i/L_i + (1\text{-}a_i)\, dJ_i/J_i$. This may be seen from equation (8) from which it also follows that $\beta_i = (1\text{-}\alpha_i) = (1\text{-}a_i)$. If shares do vary, then we may be able find an explicit functional form that is more flexible than the Cobb-Douglas (such as the CES) that gives a good fit to the accounting identity; but, of course, this does not mean that the estimated coefficients can now be interpreted as technological parameters.[8] If $a_i dw_i/w_i + (1\text{-}a_i)dr_i/r$ does not meet the assumptions noted above, all this means is that the estimate functional form will be misspecified and the goodness of fit will be reduced (McCombie, 2000 and Felipe and McCombie, 2001).

Time-series data

The fact that the identity precludes interpreting the Cobb-Douglas or more flexible functional forms as unambiguously reflecting the underlying technology of the economy implicitly suggests that this is true of estimations using time-series data.[9] Nevertheless, the arguments, as Shaikh (1974, 1980) has shown, follow through in the case of time-series data. Differentiating the income identity with respect to time, we obtain

$$v_t = a_t \hat{w}_t + (1\text{-}a)_t \hat{r}_t + a_t \ell_t + (1\text{-}a)_t j_t \qquad (10)$$

where v, ℓ, j, $\hat{w}$, and $\hat{r}$ denote exponential growth rates. Assuming that factor shares are constant and integrating, we obtain:

$$V_t = B w_t^{a} r_t^{(1\text{-}a)} L_t^{a} J_t^{(1\text{-}a)} \qquad (11)$$

Let us assume that the growth of wages occurs at a roughly constant rate and the rate of profit shows no secular growth (both of which may be regarded as stylised facts). Consequently, $a_t \hat{w}_t + (1\text{-}a)_t \hat{r}_t \approx \lambda$, a constant, and so equation (11) becomes the familiar Cobb-Douglas with exogenous technical change, namely $V_t = A_o e^{\lambda t} L_t^{a} J_t^{(1\text{-}a)}$.

[8] The CES function may be regarded simply as a Box Cox transformation that is simply a mathematical transformation that attempts to find the best fit for the linear identity when *w* and *r* vary.

[9] Phelps Brown (1957) does consider time-series data, but only when there is no time trend included (i.e., the data is not de-trended). He shows that, in these circumstances, the estimated coefficients will merely reflect the historical growth rates of the various variables.

In fact, while the cross-section studies normally give a very good fit to the Cobb-Douglas (and other) production functions, the time-series estimations sometimes produce implausible estimates with, for example, the coefficient of capital being negative. The fact that the results are often so poor may ironically give the impression that the estimated equation is actually a behavioural equation. However, the failure to get a good fit will occur if either the factor shares are not sufficiently constant or the approximation $a_t \hat{w}_t + (1-a)_t \hat{r}_t \approx \lambda$ is not sufficiently accurate. In practice, the latter proves to be the case, as estimations of equation (11) with a variety of data sets produces well-determined estimates of the coefficients with low standard errors (notwithstanding the ever-present problem of multicollinearity). It transpires that the rate of profit has a pronounced cyclical component and so proxying the sum of the weighted logarithms of *w* and *r* by a linear time trend (or their growth rates by a constant) biases the estimated coefficients of *lnL* and *lnJ* (McCombie, 2000-2001, Felipe and Holz, 2001, and Felipe and McCombie, 2001b).

The conventional neoclassical approach, which is based on the maintained hypothesis that an aggregate production function is, in fact, being estimated, usually attributes a poor fit to the failure to adjust the growth of factor inputs for the changes in capacity utilisation. Since $\lambda_t \equiv a_t \hat{w}_t + (1-a)_t \hat{r}_t$ tends to vary procyclically, the inclusion of a capacity utilisation variable (or the adjustment of *k* and ℓ for changes in their utilisation rates) will tend to improve the goodness of fit and cause the estimated coefficients to approximate more closely the relevant factor shares. As Lucas (1970), commented: "...some investigators have obtained 'improved' empirical production functions (that is, have obtained labor elasticities closer to labor's share) by 'correcting' measured capital stock for variations in utilisation rates" (Lucas, 1970). An alternative procedure would be to introduce a sufficiently complex non-linear time trend more accurately to capture the variation of λ_t (Shaikh, 1980, Felipe and McCombie, 2001c). With sufficient ingenuity, we should be able eventually to approximate closely the underlying identity, increasing both the R^2 and the values of the t-statistics, and hence find a very good fit for the "production function". Generally, as we have noted above, it is this problem, rather than the change in factor shares, that is of greater empirical importance.

The Problems of Using Monetary Values at Constant Prices as Proxies for Quantities

The problem ultimately stems from the use of value data as a proxy for "quantity" or "volume" measures. The fact that the neoclassical production function should be theoretically specified in terms of physical quantities, and not in value measures, is not often explicitly stated, but an exception is Ferguson (1971). In his comment on Joan Robinson's review of his book, *The Neoclassical Theory of Production,* he argues, "I assume a production function relating physical output to the physical inputs of heterogeneous labour, heterogeneous machines and heterogeneous raw materials. As a first approximation, I further assume that the *definition* of the output required the various raw materials to be used in fixed proportions. Thus, attention was directed to the first two heterogeneous categories of inputs. Assuming variable proportions, each physical input has a well-defined marginal physical product. If profit maximisation is also assumed ... each entrepreneur will hire units of each physical input until the *value* of its marginal physical product is equal to its market determined and parametrically given input price" (emphasis in the original). He continues that "neoclassical theory deals with macroeconomic aggregates, usually by constructing the aggregate theory by analogy with the corresponding microeconomic concepts. Whether or not this is a useful concept is an empirical question to which I believe an empirical answer can be given. This is the 'faith' I have but which is not shared by Mrs Robinson. Perhaps it would be better to say that the aggregate analogies provide working hypotheses for econometricians." (Ferguson, 1971.) Thus, Ferguson is implicitly arguing that although the neoclassical production function should be specified in terms of physical quantities, these may be adequately proxied by deflated monetary values where this is necessary for aggregation. Jorgenson and Griliches (1967), for example, take a similar position, explicitly stating that physical output is taken as "real product as measured for the purposes of social accounting".

If we are dealing with physical quanitities, then it is possible to estimate a production function and to test the marginal producivity theory of distribution. To see this, consider the neoclassical approach that uses a micro-production function specified in *physical* terms (assuming no technical change):

$$Q = f(L, K,) \qquad (12)$$

where Q and K are measured in physical or homogeneous units.

The first order conditions are, under the usual assumptions, $\partial Q/\partial L = f_L$ and $\partial Q/\partial K = f_K$, where f_L is the marginal product of labour (measured in physical units of output, say, widgets per worker) and f_K is the marginal product of capital measured in terms of widgets per "leet" (where the capital stock, after Joan Robinson, is measured as the number of leets).

Let us assume that the production function is homogeneous of degree one. By Euler's theorem which, of course, has no economic content, *per se*, we have:

$$Q = f_L L + f_K K \qquad (13)$$

It should be noted that equation (13) follows from equation (12) purely as a mathematical proposition. It is given economic content by assuming that factors are paid their marginal products. If we wish to express equation (13) in monetary terms, it is multiplied by the price of widgets so that:

$$pQ = mL + nK \qquad (14)$$

where m and n are the marginal products of labour and capital in monetary terms, i.e., $m = pf_L$ and $n = pf_K$. It should be emphasised that the physical quantities can always be simply recovered from the data by dividing by the price. Let us, for expositional purposes, assume that the marginal products are (roughly) constant and that equation (12) is a Cobb-Douglas production function. If $Q_t = AL_t{}^{\alpha}K_t{}^{(1-\alpha)}$ were to be estimated as $lnQ_t = c + b_1 lnL_t + b_2 lnK_t$, the estimates of b_1 and b_2 would be the relevant output elasticities. The accounting identity does not pose a problem in this case. If we were to estimate the linear equation (14) as:

$$pQ = b_3L + b_4K \qquad (15)$$

we would find the estimates of b_3 and b_4 would be pf_L and pf_K, or m and n.[10] If the output elasticities of labour and capital are 0.75 and 0.25, then $m/p = 0.75Q/L$ and $n/p = 0.25Q/L$.

Does the identity pose a problem for the interpretation of the production function? The answer is clearly no. Let us assume that factors are *not* paid their marginal products. We are dealing, say, with a command economy and labour only receives γ of its marginal product (where $0>\gamma>1$) while the state, the owner of capital, appropriates the remainder, plus the payments accruing to capital. Thus, we have a distribution equation $pQ = wL + rK$ where $w = \gamma m = \gamma pf_L$ and $r = (1-\gamma)pf_K$. There is of course an infinite number of combinations of w and r that could satisfy equation (15), but if we were to estimate it, then the coefficients would be m and n. Equation (14) can be interpreted as just a linear transformation of equation (12). This is an important point, because the coefficients of the estimated linear equation (15) are determined by the underlying production function and they will differ from the observed wages and rate of profit, if factors are not paid their marginal products. This is because Q is a physical measure and is independent of the distribution of the product.

Thus, in these circumstances, the discussions concerning the appropriate estimation procedures of the production function (whether it should be part of a simultaneous equation framework, etc.) become relevant.[11] Moreover, the marginal productivity theory of factor pricing may be tested by a comparison of the estimated output elasticities with the factor shares.

---

[10] We assume that there is sufficient variation in the data to provide estimates of b_3 and b_4 and abstract from problems of multicollinearity.

[11] Technical progress may be incorporated by including a correctly specified time trend.

Alternatively, it may be assumed that factors are paid their marginal products and the growth accounting approach may be adopted. There is a well-defined neoclassical production function $Q = f(L,K,t)$ which, when expressed in growth rates, becomes $q_t = \lambda_t + \alpha_t \ell_t + (1-\alpha_t)k_t$ where $\alpha_t = (f_L L/Q)$ and $(1-\alpha_t) = (f_K K/Q)$. Since factors are paid their marginal products, $pf_L = w$ and $pf_k = r$ and so the Solow residual (or the growth of total factor productivity) is given by $\lambda_t \equiv q_t - a_t\ell_t -(1-a_t)k_t$. Of course, we do not need to estimate a production function to obtain the Solow residual, but the whole procedure depends upon the existence of perfect competition and the marginal productivity theory of factor pricing and so ultimately requires that these assumptions are capable of being tested empirically. Alternatively, we may calculate the Solow residual from the "dual"as $\lambda_t \equiv a_t \hat{w} +(1-a)_t \hat{r}$. But it must be emphasised that all the usual neoclassical assumptions underlie this interpretation, including that there is a well-behaved production function and factors are paid their marginal products.

The neoclassical approach then moves almost seemlessly from the consideration of the production function in terms of physical quantities to the use of value data, where it assumed (erroneously) that all the arguments follow through in a straightforward manner. The proposition that the aggregate production function should be regarded simply as a "parable" crops up time and time again in the defence of the neoclassical approach, especially with regard to the Cambridge Capital Theory debates. The world is a complex place, so the argument goes, and any model necessarily abstracts from reality. It may be, for example, that the strict conditions for the aggregation of production functions are not met theoretically. But if the estimation of an aggregate production function gives a good statistical fit and plausible estimates of, for example, the output elasticities, we can have Ferguson's faith that the estimated relationship is telling us something about the underlying technology of the economy.

This is very much reminiscent of Friedman's (1953) instrumentalist approach to methodology. The realism or otherwise of the assumptions is irrelevant – what matters is the predictive power of the model.[12] Of course, it could be argued that considering whether or not the estimates are plausible rather begs the question. But we may absolve the argument from this charge at least, as we do have some indication of what is considered plausible. Fisher (1971) has noted that Solow has commented that if Douglas had found capital's share to have been three-quarters rather than one-quarter, we should not now be talking about production functions. But the problem that we encounter is that in moving from the micro- to the macro-level we need prices to aggregate the output.

---

[12] This is a very controversial methodological position and we shall not discuss its merits here, but rather note that it seems implicitly to underpin much of neoclassical modelling.

The difficulty is that with the use of value data, the underlying accounting identity does produce an insurmountable problem. Let us assume that firms pursue a mark-up pricing policy where the price is determined by a fixed mark-up on unit labour costs. (We make this assumption for simplicity as, in practice, firms mark-up on normal unit costs. See Lee (1999) for a detailed discussion.) Thus $p_i = (1+\pi_i)w_iL_i/Q_i$ where π is the mark-up. The value added is $V_i = p_iQ_i = (1+\pi_i)w_iL_i$ and for industry as a whole $V = \sum p_iQ_i = \sum(1+\pi_i)w_iL_i$, or approximately, $V = (1+\pi)wL$, where π is the average mark-up and w is the average wage rate. Labour's share is $a = 1/(1+\pi)$ and will be constant to the extent that the mark-up does not vary. In practice, it is likely to vary to the extent that the composition of firms with differing mark-ups alter and there are changes in the individual mark-ups, which may be temporary, as a result of the wage bargaining process. We also have the identity $V \equiv wL + rJ$ where rJ, the operating surplus, is equal to πwL and $(1-a) = \pi/(1+\pi)$. The identity now poses a major problem for the aggregate production function. Suppose that w and r do not change over time. If we were to estimate $V = b_6L + b_7J$, then the estimates of the b_6 and b_7 will always be w and r respectively. If factor shares are constant, then an approximation to the accounting identity will be given by $V = AL^aJ^{(1-a)}$, but the causation is from the identity to the multiplicative power function, not the other way around. The values of the putative elasticities are determined by the value of the mark-up and do not reflect any technological relationship. The fact that the Cobb-Douglas gives a good fit to the data does not imply that the aggregate elasticity of substitution is unity. Shares may vary, in which case a more flexible function than the Cobb-Douglas, such as the translog, will give better fit; but we still cannot be sure that the data is telling us anything about the underlying technology of the economy.

It is worth elaborating on the value added accounting identity. The estimate of r in the accounting identity, $V \equiv wL + rJ$, is the observed aggregate rate of profit or rate of return. It has been used, for example, in numerous studies concerned with the analysing the profit squeeze that occurred in the advanced countries during part of the post-war period. However, the impression may be given that somehow the critique depends upon this method of calculating r explicitly from the accounting identity. It clearly does not. Value added must, if it is to be accurately measured, be the sum of the total compensation of labour and capital (and land, but for expositional ease we shall ignore this). Since we have direct statistics on wages and employment, the rate of return is imputed from the data on total profits and the gross or net capital stock, the latter being calculated by the perpetual inventory method. However, the estimate of value added must equal the compensation of employees and the self-employed and the gross (or net) operating surplus. If we did have an independent measure of the rate of return that differed from the imputed value, then the statistical discrepancy would have to be resolved.[13, 14] The rate of profit calculated from the identity and the national accounts will contain any elements of monopoly profits that accrue to the firms, as will the estimate of value added. Hill (1979) contains a detailed discussion of the various types of profit measures, and identifies the conditions in which the net and gross rates of return will approximate to average realised internal rates of return.

A misunderstanding may arise because it is sometimes assumed that a firm's cost identity is given by $C = wL + r_cK$ where C is the total cost and r_c is the competitive rate of return or the competitive cost of capital, such that economic profits are zero. It is also usually assumed that the labour market is competitive so w is the competitive wage. r_c is sometimes calculated as the rental price of capital.[15] In this case, the rental price of capital will only equal the rate of return if the former has been calculated correctly and perfect competition prevails. If it is *assumed* that markets are competitive, then "given either an appropriate measure of the flow of capital services or a measure of its price, the other measure may be obtained from the value of income from capital" (Jorgensen and Griliches, 1967) -- although the rider should perhaps be added that the income from capital (and hence the measure of value added) should exclude the monopoly profits. However, because the conditions for producer equilibrium have been invoked, Jorgenson and Griliches continue that "the resulting quantity of capital may not be employed to test the marginal productivity theory of distribution, as Mrs Robinson and others have pointed out". Of course, this does not prevent the neoclassical economist from estimating an aggregate production function, but the "quantity" of output should, as we have noted, in these circumstances, be the constant price value of total costs.[16] However, even if the estimate of r_c is subject to serious measurement

---

[13] In practice there already is a (very small) statistical discrepancy in the estimates of GDP when calculated by the cost-structure method.

[14] Denison (1967 pp.142-144) warns that the levels of the rates of return, and their international differences, calculated from the national income statistics, must be interpreted with caution. This is because they are calculated using the *levels* of the net or gross capital stock. Both measures are sensitive to the assumed service lives, while more confidence can be had in their growth rates. The implication is that one can also be more confident about the rate of change of the rate of return, as opposed to its level. Hsieh (1999) used three alternative measures of the rate of return (the curb loan rate, the bank discount rate and the bank deposit rate) to measure the marginal product of capital and concluded that their rates of change in Korea in 1960s was not greatly different from that of the national accounts measure. However, in the case of Singapore, the alternative measures did not fall over time nearly as much as that implied by the national accounts. Hsieh concludes that the national accounts measure is inaccurate (notably because of measurement errors in the estimates of the capital stock). However, the fact that there may be measurement errors in r and J does not affect the theoretical basis of the critique.

[15] It is taken to be the real rate of interest rate plus the depreciation rate multiplied by the price of capital.

[16] This rationale for this may be also seen as follows. Using the measure of value added in the national accounts, the growth of value added is $v \equiv \lambda'_t + a_t\ell + (b_t + c_t)k$ where $\lambda'_t = a_t\hat{w} + b_t\hat{r}_c + c_t\hat{r}_m$ where $\hat{r}_c$ and $\hat{r}_m$ are the growth rates of the competitive and the monopoly element of the rate of return. a, b, and c are the shares of wages, competitive profits and profits due to the monopoly component in total revenue ($a + b + c = 1$). It can be seen that the residual contains an element due to the rate of

errors, if this procedure is followed we should still find that that the output elasticities equal the factor shares so long as the total cost identity holds by construct. It should be noted, however, that the national accounts data for value added are not constructed on the basis of any of these neoclassical assumptions, but from actual magnitudes. Hence while estimating production functions is a tautology, it is not a tautology generated by constructing the data under neoclassical assumptions.

Micro-production Functions and the Aggregate Cobb-Douglas "Production Function"

As we have seen, so long as factor shares are constant, we will obtain a good statistical fit to the supposed Cobb-Douglas production function even though it does not represent the underlying technology. There have been a number of studies that have explicitly illustrated how this can occur under a variety of assumptions.

Houthakker (1955-6) developed a model where, although each industry was subject to a fixed coefficients technology, the aggregate data behaved as if it were a Cobb-Douglas production function. (A simple explanation of this model may be found in Heathfield and Wibe (1987, pp.150-152).) He assumes that there are machines in existence that can produce output for every fixed pair of input coefficients and that the input-output ratios are distributed in ascending order according to a Pareto distribution. He shows that these micro-production functions can be aggregated into an industry-wide Cobb-Douglas production function with decreasing returns to scale. This result clearly shows how an aggregate production function may give the appearance of a technology that has an elasticity of substitution of, in this case, unity, whereas at the micro level there is no possibility of smoothly substituting between inputs. The implications of this have been succinctly stated by Blaug (1974). "It is well known that the competitive theory of factor pricing does not stand or fall on the existence of continuous and differential production functions: we can handle fixed coefficients simply by writing our equilibrium conditions as inequalities rather than equalities. True, but the ability to fit an aggregate production of the Cobb-Douglas form may throw no light on the underlying technology of the firm and hence on the process by which competitive pressures in individual markets impute prices to factors of production, and that is the question at issue." Of course, it may be argued that the fact that the firm sizes have to be distributed in a particular

---

change of the component of monopoly profits which, *under neoclassical assumptions*, has nothing to do with technical change. Consequently, if one were to interpret this in growth accounting terms, the measure of value added should be net of r_mK, and labour's and capital's share will the ratio of the total compensation of labour and capital to total costs, rather than to total revenue. One neoclassical approach is to use labour and capital's share in total costs (rather than revenue) as proxies for the output elasticities, but this does not obviate the problem.

way, which is unlikely to occur in reality, makes this merely a curiosity. Nevertheless, it does serve to demonstrate theoretically how a good fit to an aggregate production function may give a completely misleading picture of the underlying technology of the economy.

Fisher (1971) has likewise raised serious doubts concerning the aggregate production function. His approach was to construct a hypothetical economy where it was known in advance that the conditions for successful aggregation were not met. To this end, he postulated that firms produced a homogeneous output with homogeneous labour and capital that was specific to the firm and could not be reallocated to other firms. The micro-productions were the Cobb-Douglas and labour was allocated optimally to ensure that output was maximised. The hypothetical economy consisted of 2, 4 or 8 firms. The method was to simulate the growth of these firms over 20 time periods. The total labour force, the firms' technology and their capital stocks were assumed to grow at a constant rate (with a small random term to reduce multicollinearity in the subsequent regression analysis). In certain of the experiments, some of these growth rates were set equal to zero and the growth of the capital stock was allowed to vary between firms.

Fisher finds that the aggregate production function gives a good fit to the resulting aggregated data and that the wages predicted from the aggregate production function are very close to the actual wages. These results occur even though it can be shown unequivocally that the "underlying technical relationships do not look anything like an aggregate Cobb-Douglas (or indeed *any* aggregate production function) in any sense". The implications are far reaching. "The point of our results, however, is not that any aggregate Cobb-Douglas fails to work well when labor's share ceases to be roughly constant, but that an aggregate Cobb-Douglas will continue to work well so long as labor's share continues to be roughly constant, *even though that rough constancy is not itself a consequence of the economy having a technology that is truly summarized by an aggregate Cobb-Douglas*" (Fisher, 1971, emphasis added). While Fisher considers that the predictive power of the aggregate production function is a "statistical artefact", he contends it is not "an obvious one". Nevertheless, Shaikh (1980) has shown that the reason may be attributed to the underlying accounting identity.

Intuitively, Fisher's result occurs roughly for the following reason. As has been seen, the homogeneous output of the *i*th individual firm is defined as $Q_i \equiv w_i L_i + r_i K_i$. If the aggregate factor shares are constant, the arithmetically summed data of the various firms should give a good fit to the aggregate Cobb-Douglas production function, with the output elasticities being the weighted average of the individual factor shares. (This summation requires an index of the capital stock J in homogenous units to be constructed.) Thus, the converse also holds if we start with micro-production functions, as does Fisher. These will have corresponding linear accounting identities where the data vary over time in such a way that the factor shares are constant. These identities can be summed arithmetically and, provided that the aggregate shares are constant, a good statistical fit to the aggregate Cobb-Douglas production function will be found. This is not withstanding the fact that there are severe theoretical aggregation problems if we try to aggregate the micro-production functions directly. The position is actually a little more complicated than this as in Fisher's simulations K_i is specific to each firm and, as noted above, it is necessary to construct a measure of the total capital stock. Fisher does this by weighting the K_{it} by $\bar{r}_i$, the average rate of return over the 20-year period (and hence a constant), before summing the series to gain a measure of the aggregate capital stock. Thus $J_{it} = \sum \bar{r}_i K_{it}$. Shaikh (1980), however, has shown that the bias induced by this aggregation procedure is not large.

Fisher (1971, p.319-20), himself, realises that the reason why the aggregate production function gives a good prediction of wages is because "the organizing principle" is that the aggregate shares are constant. He hesitates to make this argument "lest the reader mistake it for a proof and believe all the experiments were unnecessary". In the light of the above arguments, this may be legitimately questioned.

The final example we shall consider is the evolutionary model of Nelson and Winter (1982). They develop a simulation model where a hypothetical economy is made up of a large number of firms producing a homogeneous good. The technology available to each firm is fixed coefficients, with a large number of possible ways of producing the good given by different input coefficients (a_L, a_K). However, the firm does not know all the possible combinations of the input output coefficients, but only learns about them by engaging in a search procedure. The firms are not profit maximisers, but are satisficers and will only engage in search if the actual rate of profit falls below a certain satisfactory minimum. There are two ways by which the firm may learn of other fixed coefficients techniques. First, it engages in a localised search in the input coefficient space. It is assumed that the probability of identifying a technique new to the particular firm is a declining function of the "distance" between the two techniques in terms of their efficiency. (The efficiency of a technique h' compared with h is a weighted average of $ln(a_K^h / a_K^{h'})$ and $ln(a_L^h / a_L^{h'})$.) Secondly, the firm may discover the existence of a technique because other firms are already using it. It is assumed that the probability of discovering this technique is positively related to the share of output produced by other firms using this technique. The probability of finding a new technique h' is thus a weighted average of the probability of finding the technique by local search and by imitation. The firm will adopt h' only if it gives a higher rate of profit than that obtained by the existing technique. It is also possible for the firm to misjudge the input coefficients of an alternative technique. The wage rate is endogenous to the model and is determined in each time period by reference to a labour supply curve. The prevailing wage rate affects the profitability of each firm, given the technique it is using. The behaviour of the industry as a whole also has some affect on the wage rate. The labour force is assumed to grow in each period.

In the simulations, it is possible to vary the degree of localness of the search and the ease of discovering more efficient techniques that are either labour or capital saving. It is likewise possible to vary the degree of searching for imitation. The model was simulated with a view to comparing the outcome with Solow's (1957) results from fitting an aggregate production for US data. To achieve this, the input coefficient pairs space was derived from Solow's historical data.

The simulations show that the increase in wages has the effect of moving firms towards techniques that are relatively capital intensive. When the firms check the profitability of technique, when there is a higher wage rate, it will be the more capital-intensive techniques that will pass the test. The rising wage rate will make all techniques less profitable, but those that are labour-intensive will be relatively even less profitable. However, as Nelson and Winter (1982) point out, "while the explanation has a neoclassical ring, it is not based on neoclassical premises". The firms are not maximising profits. "The observed constellations of inputs and outputs cannot be regarded as optimal in the Paretian sense: there are always better techniques not being used because they have not yet been found and always laggard firms using technologies less economical than current best practice."

The simulation results produce industry data very similar to Solow's historical data. Indeed, if an aggregate Cobb-Douglas production is fitted to the data generated by the model using Solow's procedure, very good fits are obtained with the R^2s often over 0.99 and the estimated aggregate "output elasticity with respect to capital" (which, in fact, does not exist) very close to capital's share. As Nelson and Winter (1982) observe, " the fact that there is no production function in the simulated economy is clearly no barrier to a high degree of success in using such a function to describe the aggregate series it generates." For our purposes it is worth emphasising that the macroeconomic data suggests an economy characterised by factors being paid their marginal products and an elasticity of substitution of unity, even though every firm is subject to a fixed coefficients technology. The reason why the good fit to the Cobb-Douglas production function is found is once again because the factor shares produced by the simulation are relatively constant.

Finally, Shaikh (1987) has also demonstrated that when shares are constant "*even a fixed proportion technology undergoing Harrod neutral technical change is perfectly consistent with an aggregate pseudo-production function* [that is, Cobb-Douglas in form]" (emphasis in the original).

Why Has the Critique been so Widely Ignored?

There is little doubt that the critique is either largely ignored or generally seems to be unknown. From a search of the literature, we can find only a handful of references to the problem posed by the accounting identity and these largely relate to the Phelps Brown critique of cross-section data. Given the potential importance of the criticism in undermining the empirical basis of the aggregate production function, it is interesting to speculate why this should be the case. Undoubtedly, there is a reluctance to abandon what is seen as a powerful explanatory concept; indeed, if the aggregate production function and the marginal productivity theory of factor pricing were to be abandoned, there would perhaps be little left of neoclassical macroeconomics.

Of course, whether something is well known or not is a rather subjective matter. One commentator on this debate disingenuously commented that the Cambridge Capital Theory Controversies were, in fact, widely debated in the most prestigious journals and so are well known, conveniently overlooking the fact that this critique is logically independent of the theoretical issues raised in the Capital Theory Controversies (and, indeed, was never quoted in the latter).[17] He further argued that the references that have been cited by the authors in other papers to the three now dated textbooks, namely, Cramer (1969), Wallis (1979) and Intriligator (1978); the passing remarks (literally a few sentences) in Robinson (1970) and Harcourt (1982); and the brief passage in Lavoie (1992) all constitute the argument being well-known. (The last three authors, however, did fully understand the damaging implications of the critique for the concept of the aggregate production function.)

One of the authors was told that this critique is part of an "oral tradition" criticising production functions, but has been completely discounted, although for reasons that are left unreported in the literature. Of course, to contend that a criticism of any argument is well known is a convenient rhetorical device as it implies that if there were anything in it, it would not have been ignored. However, the fact remains that in the plethora of published production function studies and related papers we have found only one that attempts to refute the critique and that is Solow (1987).[18] (This is discussed below.) We agree with Sylos Labani (1995)

[17] It is surprising just how often the defence, which we noted and criticised in the pervious section, that the aggregate production function is merely a parable is put forward and the critique advanced in this paper is erroneously dismissed as just an attempt to resurrect the Cambridge Capital Controversies.
[18] Fisher (1971), in a footnote, comments that "Phelps Brown [1957] simply dismisses the time series results as poor or implausible, largely because of their failure to allow for technical change. In the light of Solow's seminal paper [1957] and its successors, this can no longer be done. Phelps Brown's arguments as to cross-section estimates explain nothing about the time series results, nor do they show why a cross-sectionally estimated production function should give reasonable wage predictions for years far from the original cross-section". Two comments are in order here. First, it is true that the first part of Phelps Brown's paper is addressed to criticising time series estimations of supposed production functions without a time trend (and had nothing to do with the accounting identity), which was common with the early studies. (See McCombie, (1998), and Felipe and Adams (2001) for discussions of Douglas's original studies.) But we have seen that his critique concerning the accounting identity can be simply generalised to the case of time series estimations. Secondly, we do not understand Fisher's second comment. If shares remain roughly constant over time as well as between industries,

when he writes that "it is worth recalling these criticisms, since an increasing number of young and talented economists do not know them, or do not take them seriously, and continue to work out variants of the aggregate production function and include, in addition to technical progress other phenomena, including human capital".

The fact that the critique has been largely overlooked stems to some extent from the fact that the early discussions of the Phelps Brown critique did not take its implication to its logical conclusion, namely that it is not possible to *test* for the existence of the aggregate production function using value data. Let us take a few examples. In Walters's (1963) early, but influential, study surveying production and cost functions, the Phelps Brown criticism is there, but buried in a short paragraph on page 37. "The early commentators pointed out that the data may be explained by what Bronfenbrenner called the interfirm function [$pQ = wL + rK$]. Evidence has been adduced by Phelps Brown to show that the scatter of observations of Australia in 1909 can be explained in terms of this simple linear relationship. Thus, in fitting a Cobb Douglas function (with $\alpha+\beta=1$), we merely measure the share of wages in the value added. The result does not provide a test of the marginal productivity law."[19] This would seem to be pretty conclusive, but in the very next paragraph, Walters goes on to argue:

> The inter-industry results give, I think the most unsatisfactory estimates of the production function. But aggregate industry data have been used with considerable success in *interstate* (or *international*) studies of the SMAC (or the CES, as it is now more commonly known) function. The authors used observations of the same industry in different countries to estimate the parameters. Given that the industry has the same production function, the different ratios of factor prices will generate observations which should trace out the production function. (emphasis in the original)

It is difficult to reconcile these arguments. True, inter-industry estimations of putative production functions are likely to be suspect for other reasons, such as we do not really expect each different industry to have the same "production function" parameters. There may also be little variation in factor prices because of competition, so all the data will be simply observations for one particular capital-labour ratio. But the Phelps Brown critique, although originally addressed to Douglas's cross-industry study applies, of course, equally to estimates for the same industry but using interregional or international data.[20] Hildebrand and Liu

then estimating two "production functions" using cross-section data that are several years apart will give estimates of the "output elasticities" that are the same for both regressions.

[19] The notation has been changed to make in consistent with that in this paper.

[20] It may be that Walters implicitly assumes that for the accounting identity to pose a problem, wages and the rate of profit must be constant so that $VA = b_1L + b_2K$ gives a good statistical fit. This is violated in the data used by ACMS and so perhaps Walters assumes that there is now no problem. As we have shown, this assumption is not required for the Phelps Brown critique.

(1965), for example, is an early study that uses regional data to estimate production functions for the same industry.

Intriligator's (1978, chapter 8) more recent textbook treatment of the issue displays a similar ambivalence. After discussing the Simon and Levy's (1963) interpretation of the Phelps Brown critique, he concludes that "assuming only small variations in output and inputs, the form of the production function and the equality of the values of output and income imply that the production function exhibits approximately constant returns to scale and that factor shares are approximately the elasticities". But again, there is no mention that this undermines the very possibility of testing the production function. Instead, Intriligator goes on to discuss other specifications of production functions, including those estimated by time-series data.[21]

Wallis (1979, chapter 10) also accepts that "the equation as estimated by Douglas and his co-workers is a close approximation to this [accounting] identity and there is very little point in attempting to rediscover it. If all revenue is paid to either capital or labour, it is difficult [or, rather, impossible] to distinguish between this accounting identity and the estimated equation." Again, we infer that Wallis considers that this is only a problem if wages and the rate of profit are constant.

The critique has been occasionally reinvented.[22] Samuelson (1979), for example, in reviewing Douglas's academic contribution on the latter's death, became yet another to discover, to his evident surprise, that the Emperor had no clothes. He noted the fact that that there is an underlying accounting identity and that all that is being estimated is the identity because of the tautology induced by computing r as a residual by defining it as $(V\text{-}wL)/J$. As he put it: "No one can prevent us from labelling this last vector as (rJ), as J.B. Clark's model would permit – even though we have no warrant for believing that noncompetitive industries have a common profit rate r and use leets capital ...in proportion to the $(V - wL)$ elements!"

Commenting on the cross-sectional results, he commented: "Should I not concede that, at least, these cross-sectional investigations have tested – and verified triumphantly - the hypothesis that the C-D exponents do sum to unity to a good approximation as the neoclassical marginal-productivity wants them to do? On examination I find, when one specifies $V = AL^{\alpha}J^{\beta}$ and lets the cross-sectional data decide whether $\alpha + \beta = 1$, that results tend to follow purely as a cross sectional *tautology* based on the residual computation of the non-wage share. […] Profit and wages add up to total V along any fixed ray not because Euler's theorem is revealed to apply on that ray but rather because of the accounting identity involved in the residual definition of profit."

---

[21] A possibility is that he considers that the critique only applies to the Cobb-Douglas production function or the use of cross section data or both.

[22] It is interesting to note that Simon (1979b) was unaware of Shaikh's (1974) extension of the critique to time-series data, although Simon himself considered this aspect in his article.

Samuelson gives an example of why Douglas's estimates will always give constant returns to scale using hypothetical data and estimating $V_i = lnA + \alpha ln(wL_i) + \beta ln(V_i\text{-}wL_i)$. This is somewhat puzzling, as his previous discussion had been in terms of the Cobb-Douglas production function, $V_i = AL^{\alpha}J^{\beta}$; it is not clear why he did not deal with the problems of estimating this function directly. However, as we have seen above, the sum of the coefficients in his example must equal unity because of the underlying identity.

Samuelson also discusses time-series estimation of production functions and raises the question as to whether "Kaldor and the neo-Keynesians are right in suggesting that the Cobb-Douglas results are cooked-up forgone conclusion from the nature of the methodology!" While he concedes that this is a possibility, he does not emphasis the underlying accounting identity in this case. He does note, however, that so long as factor shares are constant, "no Clarkian can get a good fit with a function far away from Cobb-Douglas so long as shifts in the *K(t)/L(t)* ratio are not serendipitously just offset by labor-saving rather than capital-saving biases in technical change".

This lack of emphasis on the role of the identity in time series regressions is surprising because equation (11) can be expressed as:

$$lnV_t = lnB + aln(w_t L_t) + (1\text{-}a)ln(r_t J_t) \qquad (16)$$

which is the same expression as the identity that Samuelson derived using cross-sectional data except that *w* and *r* now vary.

The fact that the critique applies to the estimation of time-series data has not permeated down to the textbooks with the exception of a short reference in Heathfield and Wibe (1987). They mention Shaikh's (1974) critique, but seemingly dismiss it on the basis of Solow's (1974) one-page rejoinder, which rejects Shaikh's (1974) paper as simply wrong.

However, there were two themes in Shaikh's paper, only one of which was really discussed by Solow. The first was the one outlined in the section above, demonstrating the equivalence of the accounting identity and the Cobb Douglas production function. The second was the tautological nature of the procedure involved in Solow's (1957) procedure for estimating various specifications of the production function, and it was this Solow (1974) addressed.

Solow (1957) first calculated the growth of total factor productivity or the "residual", using annual data, as $\lambda_t \equiv v_t - a_t(j_t - \ell_t)$ where a_t is capital's share calculated from the national accounts. (Thus, Solow assumes the existence of a well-behaved production function, perfect competition and the neo-classical marginal productivity theory of factor pricing.) The index *A(t)* is then calculated by setting the base year equal to unity and then using λ_t to calculate A_t.

The production function is "deflated" to remove the effect of technical change (which is what the residual quickly becomes in the paper, after a brief acknowledgement that it will also capture all sorts of measurement errors). Hence, various specifications of $V/A = f(L, J)$ are estimated, of which the Cobb-Douglas gives (marginally) the best fit. But as shares are roughly constant, Shaikh argues that this *must* be the case by virtue of the way $A(t)$ is constructed. Moreover, he shows that any data set (even one where the plot of productivity on capital per worker traces out the word HUMBUG) will give a good fit to Cobb-Douglas function so long as factor shares are constant.

Solow's rejoinder was that he was not *testing* the neo-classical production function at all, as he had already assumed that it existed. He had, after all, assumed the marginal productivity theory of factor pricing in constructing $A(t)$. Nevertheless, Solow does consider that one can, in principle, "test" the production function. "When someone claims that aggregate production functions work, he means that (a) that they give a good fit to input-output data without the intervention of factor shares; and (b) that the function so fitted has partial derivatives that closely mimic observed factor shares." He then delivers his supposed *coup de grâce* by freely estimating the Cobb- Douglas function using the Humbug data with a linear time trend to capture technical change and finds no statistically significant relationship. Solow takes this as showing conclusively that the artificial data do not capture even a hypothetical production function. "If this were the typical outcome, we would not now be having this discussion" (Solow, 1974).

Ironically, freely estimating a Cobb-Douglas production function with Solow's (1957) own data also produces such poor results that if Solow, himself, had undertaken this, by his own criterion, he would have been forced to concede that the data rejected the existence as the production function (McCombie, 2000-2001). However, the poor fit is simply due to the fact that the linear time trend does not provide a good fit to the weighted logarithm of wages and the rate of profit. There is, of course, nothing to say that "technical change" should be a linear function of time. Shaikh (1980) finds the Humbug data give an excellent fit to the Cobb-Douglas, provided that a complex time trend is used.

Solow (1987) returns to the critique as outlined by Shaikh, arguing that Shaikh's analysis would hold even when there is a well-defined micro-production function estimated using physical units. As, it is argued, it cannot hold in this case, then the implication is that it equally cannot apply to value data, although this is a non sequitur. Let us consider the argument only as it applies to value data. Solow's contention is that from the accounting identity we may derive the expression $V = Bw^{a}r^{(1-a)}L^{a}J^{(1-a)}$, as we have seen above. But we know that $w = aV/L$ and $r = (1-a)V/J$. Let us substitute these expressions into the former equation. This gives us

$$V= (Ba^{a}(1-a)^{(1-a)}\ V/L^{a}J^{(1-a)})L^{a}J^{(1-a)} \qquad (17)$$

$$= (V/\ L^{a}J^{(1-a)})L^{a}J^{(1-a)} \qquad (18)$$

and, assuming no technical change:

$$V = (f(\ L,\ J)/\ L^{a}J^{(1-a)})L^{a}J^{(1-a)} =\ f(L,\ J) \qquad (19)$$

"What Shaikh has discovered, in other words is that any production function can be written as the product of a Cobb-Douglas and something else; and the something else is the production function divided by the Cobb-Douglas" (Solow, 1987). However, this argument presupposes that we have data from which the "true" production function can be estimated. In the case of value data, all that this procedure has done, of course, is to derive the approximation given by equation (19) to the identity. As we have seen, equation (19) will give a good fit to the data, even though there is no well-defined aggregate production function.

Another misinterpretation of the critique is that it applies merely to steady state data. The argument goes along the following lines. The income identity will give a good approximation to a constant returns to scale Cobb-Douglas production function as long as factor shares are approximately constant. The question arises of why factor shares will be constant. This will occur when outputs and inputs grow at constant exponential rates. But the use of steady state aggregate data, it is argued, will not lead to the unwitting estimation of the Cobb-Douglas production function as it will be readily apparent that there is not enough variation in the data to estimate the production function.

This argument is perfectly correct. In steady state, the factor shares must be constant (regardless of the exact underlying form of the production function, assuming it exists) and it will not be possible to estimate the production function (there will be perfect multicollinearity). But the critique outlined above has nothing to do with whether or not the economy is in steady state. Assume that firms pursue a constant mark-up pricing policy so factor shares are constant, then even though value added, capital (measured in constant price value terms) and the labour input shows a great deal of variation (indeed, even if all three series are randomly generated), the accounting identity will ensure that the Cobb-Douglas gives a good fit to the data. Moreover, even if shares do vary, as we have shown, other specifications of the putative production function (such as the translog) are simple better approximations to the identity. The counter-argument also goes on to claim that one can estimate a production function by the instrumental variable approach by finding instruments that are uncorrelated with technical change and cites the work of Hall (1988, 1990). Of course, if all one is estimating an identity then the problem of endogeneity becomes irrelevant and indeed Hall's procedure, not surprisingly, amounts to nothing more than the mere estimation of a (biased) identity (Felipe and McCombie, 2002a).

Finally, another view is merely to dismiss the whole exercise on the grounds that it is well known that under neoclassical assumptions the growth of total factor productivity is equal to the weighted growth of wages and the rate of profit. However, this overlooks the rather fundamental fact that the identity shows that this will be true even if we do not make the usual neoclassical assumptions. Related to this is the view that one does not need to estimate the production function to know the rate of technical change - under the assumption of constant returns and perfect competition, one just computes the Solow residual. But as we have shown, the Solow residual is not a measure of technical progress. Of course, if one can never test whether the underlying production function exists or is well behaved, then this procedure becomes tautological.

Conclusions

This paper has revisited some problems that the use of value data poses for the estimation of production functions. It is shown that the estimated coefficients of the supposed production function may be doing no more than capturing an underlying income identity from which data used in the estimation of the production function are drawn. This criticism has its origins in a paper by Phelps Brown (1957), although it was anticipated to a certain extent by Bronfenbrenner (1944), and was applied originally to the cross-section studies of Douglas. However, the critique was generalised by Shaikh (1974) and Simon (1979b) to the use of time series data. Other studies also have shown that the data may give a good fit to the aggregate Cobb-Douglas even though it is clear that either the aggregation conditions are violated, or the underlying relations are not Cobb-Douglas, or there is no neoclassical production function at all. All these arguments suggest that the results of regressions purporting to estimate an aggregate production function (whether it is a Cobb-Douglas or a more flexible functional form) must be treated with caution.

As Hahn (1972, cited by Blaug, 1974) puts it. "It has often been the case that a neo-classical theory has been attempted in terms of aggregate production functions and aggregates like capital. Except under absurdly unrealistic assumptions such an aggregate theory cannot be shown to follow from the proper theory and in general is therefore open to severe logical criticisms … On purely theoretical grounds there is nothing to be said in its favour. The view that that nonetheless it "may work in practice" sounds a little bogus and in any case the onus of proof is on those who maintain this." This may be impossible. We are content to leave the last word to Simon (1979b). "An examination of the data suggests instead that the observed good fit of these functions to data, the near equality of the labor exponent with the labor share of value added, and the first degree of homogeneity of the function are very likely statistical artefacts. The data show no more than that the value of product is approximately equal to the wage bill plus the cost of capital services."

REFERENCES

Arrow, K.J. (1962), "Economic Welfare and the Allocation of Resources of Invention." In R.R.Nelson (ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton NJ: NBER & Princeton University Press.

Blaug, M. (1974), *The Cambridge Revolution. Success or Failure? A Critical Analysis of Cambridge Theories of Value and Distribution*. Eastbourne: Institute of Economic Affairs.

Bronfenbrenner, M. (1944), "Production Functions: Cobb-Douglas, Interfirm, Intrafirm." *Econometrica,* vol. 12: 35-44.

Cobb, C.W. and P.H. Douglas. (1928). "A Theory of Production." *American Economic Review, Supplement*. vol. 18, pp.139-165.

Cramer, J.S. (1969), *Empirical Econometrics*. Amsterdam: North-Holland.

Denison, E. (1967), *Why Growth Rates Differ: Postwar Experience in Nine Western Countries*. Washington D.C.: The Brookings Institution.

Douglas, P.H. (1944), "Are There Laws of Production?", *American Economic Review*, vol. 38, pp.1-41.

Douglas, P.H. (1976), "The Cobb-Douglas Production Function Once Again: Its History, Its Testing, and Some Empirical Values." *Journal of Political Economy*, vol. 84, pp.903-115.

Felipe, J. (2001a), "Endogenous Growth, Increasing Returns, and Externalities: An Alternative Interpretation of the Evidence", *Metroeconomica*. (forthcoming).

Felipe, J. (2001b), "Aggregate Production Functions and the Measurement of Infrastructure Productivity: A Reassessment", *Eastern Economic Journal* (forthcoming)

Felipe, J. and Adams, G. F. (2001), " 'A Theory of Production'. The Estimation of the Cobb-Douglas Function: A Retrospective View", Georgia Institute of Technology and Northeastern University (mimeo).

Felipe, J. and Holz, C. (2001), "Why do Production Functions Work? Fisher's Simulations, Shaikh's Identity and Some New Results", *International Review of Applied Economics* (forthcoming)

Felipe, J and McCombie, J.S.L. (2001), "The CES Production Function, the Accounting Identity and Occam's Razor", *Applied Economics*, Vol.33. pp.1221-1232.

Felipe, J and McCombie J.S.L., (2001c), “Can Solow’s Growth Model be Tested? Some Methodological Concerns”, Georgia Institute of Technology & the University of Cambridge *(mimeo).*

Felipe, J and McCombie, J.S.L. (2002a), “A Problem with Some Recent Estimations and Interpretations of the Mark-up in Manufacturing Industry”, *International Review of Applied Economics,* (forthcoming)

Felipe, J. and McCombie, J.S.L. (2002b), “Methodological Problems with the Neoclassical Analyses of the East Asian Economic Miracle”, *Cambridge Journal of Economics*, (forthcoming)

Ferguson, C.E. (1969), *The Neoclassical Theory of Production and Distribution*. Cambridge: Cambridge University Press (revised edition 1972).

Ferguson, C.E. (1971), Capital Theory up to Date: A Comment on Mrs Robinson’s Article”, *Canadian Journal of Economics*, vol. IV, pp.250-254.

Fisher, F.M. (1971), "Aggregate Production Functions and the Explanation of Wages: A Simulation Experiment." *The Review of Economics and Statistics*, vol. LIII, pp. 305-25.

Fisher, F.M. (1987). "Aggregation Problems." In J. Eatwell, M. Milgate, and P. Newman (eds.), *The New Palgrave. A Dictionary of Economics*, pp.53-5. Basingstoke: Macmillan,

Fisher, F.M. (1992), *Aggregation. Aggregate Production Functions and Related Topics*. (Monz, J., ed.). London: Harvester Wheatsheaf.

Friedman, M. (1953), "The Methodology of Positive Economics." In Milton Friedman (ed.), *Essays in Positive Economics*. Chicago: Chicago University Press.

Hahn, F. (1972), *The Share of Wages in the National Income: An Enquiry into Distribution Theory*, London: Weidenfeld and Nicholson.

Hall R.E. (1988), “The Relation between Price and Marginal Cost in U.S. Industry”, *Journal of Political Economy*, vol. 96, pp. 921-947.

Hall, R.E. (1990), “Invariance Properties of Solow's Productivity Residual”, in P. Diamond (ed.) *Growth/Productivity/Employment*, Cambridge, MA :MIT Press.

Harcourt. G.C. (1972), *Some Cambridge Controversies in the Theory of Capital.* Cambridge: Cambridge University Press.

Harcourt, G.C. (1982), *The Social Science Imperialist.* (P. Kerr, ed.) London: Routledge and Kegan Paul.

Heathfield, D.F. and Wibe S. (1987), *An Introduction to Cost and Production Functions.* Basingstoke: Macmillan.

Hildebrand, G. and Liu, T.C. (1965*), Manufacturing Production Functions in the United States*, *1957*, Ithica, New York: Cornell University Press.

Hill T.P. (1979), *Profits and Rates of Return,* Paris: OECD.

Houthakker H.S. (1955-56), “The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis”, *Review of Economic Studies,* vol.23, pp.27-31.

Hsieh, C-T (1999), Productivity Growth and Factor Prices in East Asia, *American Economic Review, Papers and Proceedings*, Vol. 89, pp.133-138.

Intriligator, M.D. (1978), *Econometric Models, Techniques and Applications*. Englewood Cliffs, NJ: Prentice Hall.

Jorgenson, D.W. (1974), “Investment and Production: A Review”, in M.D. Intriligator and D.A. Kendrick (eds*), Frontiers of Quantitative Economic*s, Vol II, Amsterdam: North Holland.

Jorgenson D.W. and Griliches, Z. (1967), “The Explanation of Productivity change*”, Review of Economic Studies*, vol. 34, pp.249-83.

Lavoie, M. (1992), *Foundations of Post-Keynesian Economic Analysis,* Aldershot, Edgar Elgar.

Lee, F.S. (1999), *Post Keynesian Price Theory*, Cambridge: Cambridge University Press.

Lucas, R. E. (1970) “Capacity, Overtime, and Empirical Production Functions*”, American Economic Review. Papers and Proceedings,* vol. 60, pp. 23-27.

McCombie J.S.L. (1987). “Does the Aggregate Production Function Imply Anything about the Laws of Production? A Note on the Simon and Shaikh Critiques.” *Applied Economics*, 19: 1121-36.

McCombie, J.S.L. (1997), “Rhetoric, Paradigms, and the Relevance of the Aggregate Production Function”, in P. Arestis and M.C. Sawyer (eds) *Method, Theory and Policy in Keynes. Essays in Honour of Paul Davidson, Vol III,* Aldershot: Edward Elgar.

McCombie, J.S.L. (1998), “ ‘Are There Laws of Production?: An Assessment of the Early Criticisms of the Cobb-Douglas Production Function”, *Review of Political Economy*, vol.10, pp.141-173.

McCombie, J.S.L. (2000) “ The Regional Production and the Accounting Identity: A Problem of Interpretation” *Australasian Journal of Regional Studies*, Vol 6,

McCombie, J.S.L. (2000-2001). “The Solow Residual, Technical Change and Aggregate Production Functions ”, *Journal of Post Keynesian Economics*, vol.23, pp. 267-297 (errata Vol 23(3) p.544).

McCombie, J.S.L. (2001), “What does the Aggregate Production Function Tell Us? Second Thoughts on Solow’s ‘Second Thoughts on Growth Theory’ ”, *Journal of Post Keynesian Economics* (forthcoming)

McCombie, J.S.L., and Dixon., R. (1991), “Estimating Technical Change in Aggregate Production Functions: A Critique”, *International Review of Applied Economics,* vol. 4,pp. 24-46.

McCombie, J.S.L., and Thirlwall, A.P. (1994), *Economic Growth and the Balance-of-Payments Constraint*, Basingstoke: Macmillan.

Marshak, J. and Andrews, W.H. (1944), "Random Simultaneous Equations and the Theory of Production", *Econometrica*, vol. 12, pp.143-205.

Nelson, C. and Kang, H. (1984), "Pitfalls in the Use of Time as an Explanatory Variable in Regressions", *Journal of Business and Economic Statistics*, Vol. 2: 73-82.

Nelson R.R and Winter S.G. (1982), *An Evolutionary Theory of Economic Change*, Cambridge MA: Harvard University Press.

Pasinetti, L.L. (1994), "The Structure of Long-Term Development: Concluding Comments", in Pasinetti, L.L. and Solow, R.M. (eds) *Economic Growth and the Structure of Long-Term Development*, Basingstoke: Macmillan.

Phelps Brown, E.H. (1957), "The Meaning of the Fitted Cobb-Douglas Function", *Quarterly Journal of Economics*, vol. 71. pp. 546-60.

Robinson, J. V. (1970), "Capital Theory up to Date", *Canadian Journal of Economics*, vol.3: 309-17.

Samuelson, P.A. (1979), " Paul Douglas's Measurement of Production Functions and Marginal Productivities", *Journal of Political Economy*, vol. 87, pp. 923-939.

Shaikh, A. (1974), "Laws of Production and Laws of Algebra: The Humbug Production Function", *Review of Economics and Statistics* Vol. LVI, pp. 115-20.

Shaikh, A (1980), "Laws of Production and Laws of Algebra: Humbug II", In Edward J. Nell (ed.), *Growth, Profits and Property, Essays in the Revival of Political Economy.* pp. 80-95, Cambridge: Cambridge University Press.

Shaikh, A. (1987), "Humbug Production Function", in Eatwell, J., Milgate, M. and Newman, P. (eds) *The New Palgrave. A Dictionary of Economic Theory and Doctrine.* London: Macmillan.

Simon, H. A. (1979a), "Rational Decision Making in Business Organizations", *American Economic Review*, vol.69, pp. 493-513.

Simon, H.A. (1979b), "On Parsimonious Explanation of Production Relations", *Scandinavian Journal of Economics*, vol. 81, pp. 459-74.

Simon, H. A. and Levy, F.K. (1963), "A Note on the Cobb-Douglas Function", *Review of Economic Studies*. vol. 30, pp. 93-4.

Solow, R.M. (1957), "Technical Change and the Aggregate Production Function", *Review of Economics and Statistics*, vol. 39, pp. 312-20.

Solow, R.M. (1966), "Review of *Capital and Growth*", *American Economic Review,* vol. 56, pp.1257-60.

Solow, R.M. (1974), "Laws of Production and Laws of Algebra: The Humbug Production Function: A Comment", *Review of Economics and Statistics*, vol. 64, p.121.

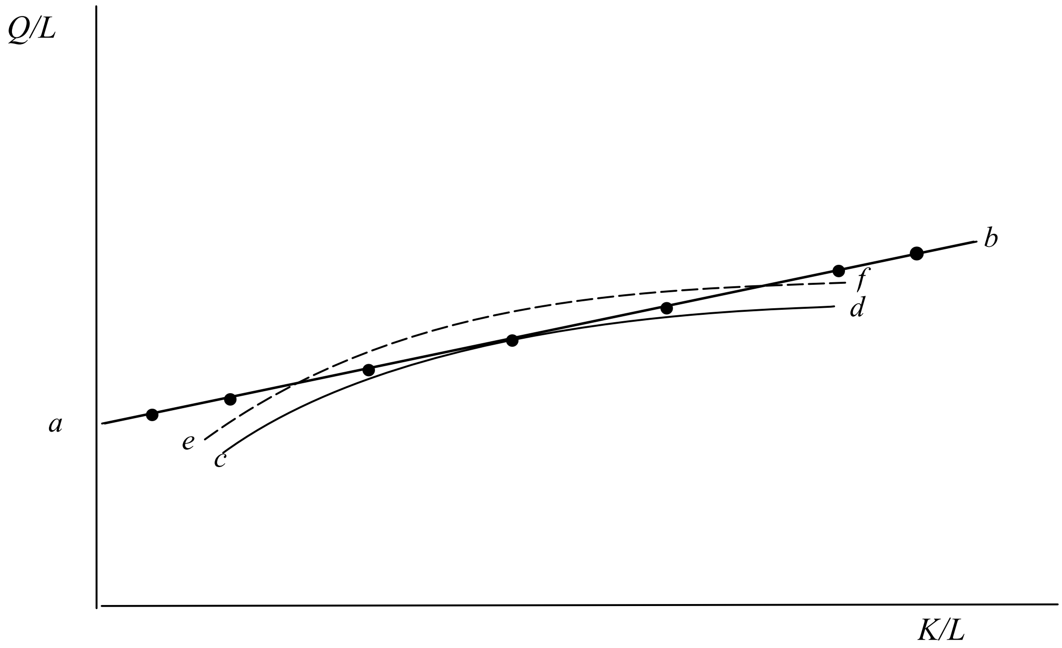
Solow, R.M. (1987), "Second Thoughts on Growth Theory", in A. Steinherr and D. Weiserbs (eds), *Employment and Growth: Issues for the 1980s*, Dordrecht: Martinus Nijhoff Publishers.

Sylos Labani, P. (1995), " Why the Interpretation of Cobb-Douglas Production Function Must be Radically Changed." *Structural Change and Economic Dynamics*, vol.6, pp.485-504.

Wallis, K. F. (1979), *Topics in Applied Econometrics*, London: Gray-Mills Publishing

Walters, A.A. (1963), "Production and Cost Functions", *Econometrica*, vol.31, pp. 1-66.

Wan, H.Y. (1971), *Economic Growth*, New York: Harcourt Brace Jovanovich.



The Approximation to the Linear Accounting Identity by the Cobb-Douglas Relationship