Abstract  Carter (2010) reproduces and discusses correspondence between Simon and Solow in 1971, where the former tentatively formulated his critique that estimates of aggregate production functions were merely capturing an underlying accounting identity. This came to fruition in a paper published by Simon in the 1979 edition of the Scandinavian Journal of Economics. This paper extends the comment of Simon that production functions should be estimated using physical data, and shows the serious problems that arise because they are estimated using constant-price monetary data. Simon also suggested that the good statistical fits to production functions could be derived from a mark-up pricing model, but he did not follow this up. The paper shows that this can indeed account for the very good statistical fits of the Cobb-Douglas and other productions functions. It concludes by showing how estimates of cost functions suffer from the same problem.

Key words. Simon, Solow, aggregate production functions, cost functions, accounting identity, mark-up.

JEL: B50, B30, E1, E23

The authors are Principal Economist at the Asian Development Bank and Director of the Centre for Economic and Public Policy, Department of Land Economy, University of Cambridge respectively. This paper reflects solely the opinions of the authors and does not necessarily reflect those of the Asian Development Bank, its Executive Directors, or those countries that they represent. We are grateful to Scott Cater for making available to us an advance copy of his paper.

Jesus Felipe, Asian Development Bank, Manila, Philippines (jfelipe@adb.org)

John McCombie, CCEPP, Department of Land Economy, University of Cambridge (jslm2@cam.ac.uk) Corresponding author.
Introduction

The correspondence between Herbert Simon and Robert Solow in 1971 published in Scott Carter’s (2011) paper gives an interesting insight into the development of the early criticisms of the aggregate production function. In particular, it shows how Simon’s idea that estimates of aggregate production functions are merely statistical artefacts developed over time. As the regression results of a putative aggregate production function reflect nothing more than a mathematical transformation of an accounting identity, a close relationship between outputs and inputs must hold by definition. Estimates of aggregate production functions reflect an identity and are not behavioural relationships. While there were hints of this critique in some earlier literature (e.g., Bronfenbrenner, 1944, and Marshak and Andrews, 1941), the starting point is really buried in the paper by Phelps Brown (1957). It was this argument as it pertains to cross-sectional data that was, in effect, subsequently formalised by Simon and Levy (1963). Nevertheless, given the nature of Phelps Brown’s argument, Simon and Levy were themselves not completely sure whether or not they had actually accomplished this. This present exchange between Simon and Solow unearthed by Scott Carter gives a fascinating insight into how Simon was developing his ideas in the early 1970s and which later came to fruition in a much more coherent form in his neglected, but incisive, Scandinavian Journal of Economics article published in 1979.

As Scott Carter points out there is now an extensive literature on the problems posed by the accounting identity, including rudimentary discussions to be found in Bronfenbrenner (1944) Marshak and Andrews (1944), and Phelps Brown (1957), noted above. Shaikh (1974, 1980, 1987) generalised the critique to time-series estimation of supposed production functions, but his 1974 paper was not criticised by Solow (1974) who perhaps, not surprisingly, defended his original approach (Solow, 1957). This criticism is not convincing (Shaikh, 1980). Solow returned (1987) to his criticism of Shaikh, but this is not compelling either (McCombie, 2001). Simon (1979a) considered the criticism in detail in the context of both cross-section and time-series data. The criticism was re-examined and extended by Felipe and Adams (2005), Felipe and McCombie (2001, 2002, 2003, 2005a&b, 2006, 2007, 2009a&b, 2010), Felipe (2001a&b),

1 It is interesting to note that although Simon (1979) acknowledges his intellectual debt to Solow, he was unaware at the time of Shaikh’s (1974) criticism of Solow (1957), and Solow’s (1974) dismissal of Shaikh’s argument. (We are grateful to Marc Lavoie for providing us with a copy of a letter from Simon to him acknowledging this fact.)

It is interesting that Simon relies on the principle of “parsimony” or, Occam’s razor, to argue in favour of the interpretation of the data as reflecting the accounting identity, rather than a production function. (See, in particular, the discussion in Scott Carter’s (2011) conclusions.) This is too weak a methodology because it implies that if the world is more complicated than we think, then it may be that the aggregate production does, in fact, exist. This is in spite of a simpler explanation being available. Yet, Simon himself was aware of the limitations of this analogy. “Occam’s Razor has a double edge. Succinctness of statement is not only the measure of a theory’s simplicity. Occam understood his rule as recommending theories that make no more assumptions than necessary to account for the phenomena. … In whichever way we interpret Occam’s principle, parsimony can only be a secondary consideration in choosing between theories, unless those theories make identical predictions” (Simon, 1979b, p.495).

The criticism of the aggregate production function does not rely simply on parsimony because, as we shall see, there are not two competing theories that give identical predictions. The existence of an accounting identity actually precludes the testing, or estimation, of an aggregate production function. It is a critique based on logical grounds and is not, for example, a statistical identification problem. This is why it is so surprising that its validity has been generally ignored. The argument is either logically correct or incorrect and the papers cited above, to our mind, have established that the answer is unequivocally the former. For example, Simon (1979b, p. 497) argues “fitted Cobb-Douglas functions are homogeneous, generally of degree close to unity and with a labor exponent of about the right magnitude [i.e. close to its factor share]. These findings, however, cannot be taken as strong evidence for the [neo]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 goods equals labor cost plus capital cost) (Phelps Brown, 1957). The same comment applies to the SMAC [CES] production function”.

The production function, which is a technological relationship, should be estimated using physical data. The essence of the problem stems from the fact that, in practice, to estimate a production function constant-price monetary values have to be used for output and capital.² But the constant-price monetary data precludes us from ever testing whether or not a true

² It is true that there are some engineering studies that use physical data but these are few and far between.
aggregate production function exists or, if this is the case (however implausible this might be), from estimating its true parameters. It is this distinction between physical and monetary data that is crucial. Simon nearly hit on this line of reasoning in his letter of 4th May 1971, where he correctly specified the production function in physical terms. But he then specified the accounting identity in such a way that it could also be transformed into physical data, which vitiates this strand of the criticism. Furthermore, he did not subsequently develop this line of reasoning in his 1979a paper.

A second point is that it has often been asserted erroneously that the critique only applies to the Cobb-Douglas production function (Temple, 2006, Felipe and McCombie, 2010). Simon’s argument (1979a) was that in estimations using cross-sectional data, the parameters of the CES did not differ significantly from those implied by the Cobb-Douglas. Hence, the critique of the accounting identity follows through. From this and similar arguments, it has sometimes erroneously been inferred that the critique only applies to where the Cobb-Douglas production function gives a reasonably good fit to the data, even although only as an approximation to a more flexible functional form. We show below is not the case.

Finally, in his 1979a paper, Simon cites Walters’s (1963) paper on production and cost functions with respect to cost curves. Walters reviews a number of econometric studies from fitted Cobb-Douglas production functions and estimated cost curves. He finds the data on cost curves were generally not the U-shape generally supportive of neoclassical theory, although as Simon commented “not so damning as to require the theory to be abandoned!” (Simon, 1979a, p.470). In this paper, we show that the estimation of cost curves also suffers from exactly the same problems as estimating the Cobb Douglas production function, which is perhaps not surprising as the former can be derived from the latter.

The Aggregation Problem

It is now comprehensively established that on theoretical grounds the aggregate production is unlikely to exist. There is a large technical literature on this problem which we shall not discuss here (See, Fisher, 1992 and Felipe and Fisher, 2003), but to begin with is instructive to quote Fisher (2005, pp.489-90) on the conclusions that can be drawn from this literature.

Briefly, an examination of the conditions required for aggregation yields results such as:

- Except under constant returns, aggregate production functions are unlikely to exist at all.
• Even under constant returns, the conditions for aggregation are so very stringent as to make the existence of aggregate production functions in real economies a non-event. This is true not only for the existence of an aggregate capital stock but also for the existence of such constructs as aggregate labour or even aggregate output.

• One cannot escape the force of these results by arguing that aggregate production functions are only approximations. While, over some restricted range of the data, approximations may appear to fit, good approximations to the true underlying technical relations require close approximation to the stringent aggregation conditions, and this is not a sensible thing to suppose.

However, in retrospect, these conclusions are hardly surprising. Take manufacturing industry, for example. At the two-digit SIC level there are such diverse industries as Tobacco Products (SIC 21) and Petroleum Refining and Related Industries (SIC 27). Does it make sense to aggregate the outputs of these two industries and their employment and capital stock and estimate an “aggregate” production function that putatively represents the technology of this new industry? If it does not, then how much more dubious is it to aggregate all 19 diverse two-digit SIC industries and estimate an “aggregate” elasticity of substitution of total manufacturing that in all probability does not exist, or, indeed, to estimate one for the whole economy?

Thus, this theoretical literature, together with the Cambridge capital theory controversies (Cohen and Harcourt, 2003), suggest that the aggregate production function cannot theoretically exist. So why does the aggregate production function continue to be so widely and uncritically used? The answer lies in the fact that it gives good statistical results with plausible estimates of the parameters. This has been true ever since Douglas’s work in the 1920s with Cobb and subsequently in the 1930s with other colleagues. Furthermore, the estimated output elasticities are often very close to the factor shares obtained from the national accounts, as predicted by the aggregate marginal productivity theory of factor pricing. As Solow once remarked to Fisher, “had Douglas found labor’s share to be 25 percent and capital’s 75 per cent instead of the other way around, we would not now be discussing aggregate production function” (cited by Fisher, 1971, p.305). But at the risk of

---

3 Time-series data do not always give good statistical fits to the putative aggregate production function, although adjusting the capital stock for the level of capacity utilisation generally improves the results and gives plausible results. Furthermore, it is also possible to find a non-linear time-trend that will always give the output elasticities equal to the factor shares. This time trend in traditional estimations of aggregate production functions is supposed to capture the rate of technical progress.
getting ahead of ourselves, we shall show that Douglas could never have found the output elasticity of labour to be anything other than equal to capital’s share of 25 percent.

The implications are far reaching. Douglas (1976, p.914), in reviewing his studies on the aggregate production function commented, “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 competitive theory of distribution and disproves the Marxian”.

The implicit defence of the use of the aggregate production function rests largely on a methodological instrumental argument. All models involve unrealistic assumptions; after all, as Joan Robinson once remarked, a map on a scale of one-to-one is of no use to anyone. What matters is the explanatory power of the model, which is taken to be synonymous with its predictive power – the symmetry thesis (Friedman, 1953). Wan (1971, p.71), for example, views the aggregate production function as an empirical law in its own right which is capable of statistical refutation, a view shared by Solow (1974). Ferguson (1969, p.xvii) explicitly made this instrumental defence with respect to the criticism about the measurement of capital as a single index in Cambridge capital theory controversies. “Its validity is unquestionable, but its importance is an empirical or an econometric matter that depends upon the amount of substitution there is the system. Until the econometricians have the answer for us, placing reliance upon [aggregate] neoclassical economic theory is a matter of faith. I personally have faith” (emphasis added).

But all this does not explain why aggregate production functions generally give such good statistical results, especially in the light of Fisher’s (2005, p.490) warning cited above that the “one cannot escape the force of these results [of the aggregation literature] by arguing that aggregate production functions are only approximations.” In answering this question, we adopt a slightly different tack to that adopted by Simon in Carter (2011) and Simon (1979a).

**On Accounting Identities and Aggregate Production Functions**

As we noted above, it is interesting that Simon almost stumbled on the most serious problem posed by the accounting identity for the production function. This is that the production function should use physical data in its estimation, while in practice value data for output and the capital stock have to be used. The accounting identity is also expressed in value data, and any estimation of the production function consequently merely reflects the former. (See
Simon’s equations (1), (1’) and (2) in Carter (2011). It is this equivalence that makes the estimation of production functions a meaningless proposition, in the literal sense.

To see this, let us take the simplest neoclassical case that assumes away all the aggregation problems and the problems of measuring capital. This is the case where it is meaningful to estimate a production function which, for the moment we unrealistically assume exists. (We consider the case where it does not below.) Abstracting from technical change for expositional ease, this is given by:

(1) \[ Q = f(K,L) \]

\( Q \) is a single homogeneous quantity, say, identical widgets, \( K \) is the number of identical machines, and \( L \) is the number of identical workers. The value of output must by definition equal the remuneration of the capital and labour. In other words:

(2) \[ pQ \equiv wL + \rho K \]

where \( p \) is the price of widgets, \( w \) is the wage rate and \( \rho \) is the rental price of capital of each machine. \( p, w \) and \( \rho \) are measured in $’s. We can write the accounting identity in purely physical terms as:

(3) \[ Q \equiv wL + \rho K \]

where \( w’ = w/p \) and \( \rho’ = \rho/p \) are the wage rate and rental price of capital expressed in terms of the numbers of widgets. Neoclassical production theory assumes that both are determined in perfectly competitive markets.

As \( Q, K, \) and \( L \) are all measured in physical units, the production function is a behavioural equation and not an identity. Thus, it is quite possible to find no statistically significant relationship between \( Q, K \) and \( L \). The marginal products of labour and capital are given by:

(4) \[ \frac{\partial Q}{\partial L} = f_L, \quad \frac{\partial Q}{\partial K} = f_K \]

as set out in Simon (1979a, pp. 463).
From Euler’s theorem for functions of homogeneity of degree one, \( Q = f_L + f_K K \).

If we assume that the true underlying function is a Cobb-Douglas with constant returns to scale, namely, \( Q = AK^\alpha L^{(1-\alpha)} \) and factors are paid their marginal products, then \( w = pf_L, \rho = pf_K \) and \( pQ \equiv wL + \rho K \). The output elasticities equal the factor shares, viz., \( \alpha = a, (1-\alpha) = (1-a) \). Thus, if the estimated coefficients of capital and labour equal the respective factor shares, then this shows that perfect competition prevails, or, strictly speaking, the estimates do not refute the assumption that markets are perfectly competitive.

This is all familiar elementary microeconomics textbook material and is based on output and capital measured in homogeneous physical units. But the problem is that for reasons of heterogeneity, we do not have physical measures of the quantities of \( Q \) or \( K \), but only constant price monetary values, sometimes confusing referred to as “quantities” or “volumes”.\(^4\) For example, in practice, we do not have the physical number of machines but just the constant price value of the capital stock. This is given by \( (\rho*/i)K = J \) where \( i \) is the rate of discount so that \( \rho*K = rJ \) where \( r \) the rate of profit and equals the discount factor. For generality, we now assume that \( \rho^* \) and \( r \) include any monopoly rents.\(^5,6\)

Simon originally used the assumption of exogenous constant ratios (most notably \( Q = aK \)) in the analysis his 4th May 1971 letter. In the same place, he wrote “Let me restate my goal: to construct a believable model of an economy that is based on the assumption of a mark-up or average cost pricing rather than maximization, but that explains the relative stability of labor’s share in the face of technological change that may or may not be neutral”. In his letter of June 30th 1971 he was again willing “to take the markup of direct costs as exogenous a la Phelps Brown and others, and then discuss the stability of labor’s share directly – but somewhat inelegantly. De gustibus…..”. Nevertheless, surprisingly, he did not follow this up in his 1979a paper.

\(^4\) \( L \) is also heterogeneous, but we ignore this problem.

\(^5\) Note that in practice the capital stock is calculated by cumulating the constant-price value of net investment by the perpetual inventory method. The ex post rate of profit is often calculated as \( r = (V - wL)/J \).

\(^6\) Some neoclassical economists see this as a criticism of the accounting identity critique as they argue neoclassical production theory refers to optimisation in perfectly competitive markets. But we could deduct both monopoly profits from \( \rho^* \) and from \( V \) and the argument carries through unaffected. See Felipe and McCombie (2007) who term this neoclassical identity, the “virtual identity”
Consequently, following Simon’s suggestion, let us assume the price of the widgets is given by a constant mark-up on unit labour costs.7

\[(5) \quad p = (1+\pi)wL/Q\]

The underlying accounting identity given by:

\[(6a) \quad pQ \equiv wL + rJ\]

or

\[(6b) \quad V \equiv wL + rJ\]

Suppose that the researcher only has data on \(V, L\) and \(J\). (In reality, as we noted above, we have to move away from the simple model because there are heterogeneous output and capital goods that are aggregated in constant price values. But for simplicity, we remain in a one-good world, but assume that only \(V\) and \(J\) are known to the researcher, and not \(Q\) and \(K\).) Factor shares are constant and for capital and labour are given by \(a = \pi/(1+\pi)\) and \(1-a) = 1/(1+\pi)\) respectively. For the moment, we continue to assume unrealistically that there is an underlying production function and that this is given by the Cobb-Douglas. Through this production function, \(L\) and \(K\) determine \(Q\). However, let us assume, as is the case in practice, that factor shares of capital and labour are \(a = 0.25\) and \(1-a) = 0.75\), but the “true” output elasticities are \(\alpha = 0.75\) and \((1-\alpha) = 0.25\). In other words, the true values of the output elasticities are the converse of what is usually considered to be plausible.

If we totally differentiated the accounting identity, equation (6b), and then integrate it we obtain:

\[(7) \quad V = a^{-a}(1-a)^{-(1-a)}p^{a}w^{(1-a)}J^{a}L^{(1-a)}\]

Note that in deriving equation (7) from equation (6b), we have neither made use of the marginal productivity conditions, nor have we assumed anything about the state of

---

7 Note for the accounting identity to give rise to the Cobb-Douglas we only need factor shares to be roughly constant. A constant mark-up will give this. Kaldor’s (1956) distribution theory is an alternative explanation as to why this may be the case.
competition or the form of the production function. Indeed, we have not even assumed that it exists.

If we were to estimate \( Q = A^\alpha J^{(1-\alpha)} \) either using cross-sectional data where the expression \( a^{-a}(1-a)^{-(1-a)}r^aw^{(1-a)} \) in equation (7) is approximately constant or using time-series data where \( r^aw^{(1-a)} \) can be accurately approximated by a time-trend, then the estimated “output elasticities” must by virtue of the accounting identity always equal the factor shares. It is worth emphasising that this is the case even when markets are imperfectly competitive. In other words the following hold:

\[
\alpha = 0.75 = \hat{\alpha} = a = 0.25 \quad \text{(capital)} \\
(1-\alpha) = 0.25 = (1-\hat{\alpha}) = (1-a) = 0.75 \quad \text{(labour)}
\]

where \( \hat{a} \) over the parameter indicates that it is a statistical estimate.

What happens if the “true” production function is characterised by increasing returns to scale? The estimates using value data will still give the estimates of the “output elasticities” equal to the factor shares. The intercepts will differ depending upon whether there are actually constant returns to scale or increasing returns to scale, but the estimates of the production function must give the sum of the output elasticities equal to unity.

Most importantly, suppose that because of such factors as differences in x-efficiency, technology, and, in particular, aggregation problems, etc., there is no well-defined relationship between the inputs and the outputs. The mark-up and the accounting identity will still result in a perfect fit to the Cobb-Douglas production function estimated using \( V, J, \) and \( L \) complete with output elasticities equalling factor shares and constant returns to scale. It is worth emphasising this point. Even though there is no well-defined production function at either the micro or the macro level, in that there is no stable mathematical relationship that explains output \( (Q) \) in terms of capital \( (K) \) and labour \( (L) \), the accounting identity can always be transformed to give the misleading impression that it does exist. (These three cases have been confirmed by Felipe and McCombie (2006) using simulation analysis.)

Let us suppose now we are again neoclassical economists and we believe that the true production function is a CES or a translog. A good statistical fit is found with once again the putative output elasticities equalling their respective factor shares, but this is because of the
use of value data. We may show this as follows. Differentiating the accounting identity, for example, with respect to time gives:

\[ \dot{V}_t = a_t \dot{r}_t + (1 - a_t) \dot{w}_t + a_t \dot{J}_t + (1 - a_t) \dot{L}_t \]

(8)

where a dot over a variable denotes a growth rate and the factor shares can change over time. Differentiating the general form of the hypothetical production function with respect to time gives:\(^8\)

\[ \dot{V}_t = \lambda_t + \alpha_t \dot{J}_t + \beta_t \dot{L}_t \]

(9)

It can be seen that for the general case of the “production function” by virtue of the accounting identity, \( \alpha_t \equiv a_t \) and \( \beta_t \equiv (1 - a_t) \) and \( a_t \dot{r}_t + (1 - a_t) \dot{w}_t = \lambda_t \).

Of course, an aggregate “production function” is theoretically merely a mathematical function that relates inputs to outputs. It is actually nothing more than a mathematical transformation that satisfies the accounting identity and which accurately tracks the path of the factor shares. Suppose, like Douglas, we have data on \( V, L \) and \( J \). Cobb suggested to Douglas that he try estimating the familiar multiplicative power relationship which we now know as the Cobb-Douglas production function. But he might have suggested using the less restrictive Box-Cox transformation. Consider the transformation of a variable:

\[ Y(\eta) = \frac{Y^\eta - 1}{\eta}, \quad \eta \neq 0 \]

\[ = \ln Y, \quad \eta = 0 \]

The extended Box-Cox transformation of the accounting identity is therefore

\[ V(\eta) = c + b_1 L(\eta) + b_2 J(\eta) \]

(11)

If \( \eta = 1 \), and the constant is constrained to go through the origin, we have the linear accounting identity. If \( \eta = 0 \) we have the familiar Cobb-Douglas. What happens if \( 0 < \eta < 1 \)? Consider the CES “production function”:

---

\(^8\) Assuming biased technical change does not make any significant difference to the argument (McCombie and Dixon, 1991).
where \( \gamma \) is interpreted as an efficiency parameter, \( \delta \) is a distributional parameter, and the elasticity of substitution is given by \( \sigma = 1/(1+\rho) \). But equation (12) may be expressed as:

(13) \[
V^{-\rho} = \{ \gamma^{-\rho} \delta \} L^{-\rho} + \{ \gamma^{-\rho} (1-\delta) \} J^{-\rho}
\]

Compare this with the extended Box-Cox transformation when \( \eta \neq 0 \) and the constant is constrained to go through the intercept:

(14) \[
V^\eta = b_1 L^\eta + b_2 J^\eta
\]

It can be seen that the CES is nothing more than a Box-Cox transformation of the linear accounting identity and will give a better fit than the former if \( w \) and \( r \) and/or factor shares vary. This could be because the mark-up varies over time, either randomly or because of changes in the market power of firms and the bargaining power of labour. Thus the data may suggest the existence of a CES production function, even though it does not exist. (See McCombie and Dixon (1991) for an alternative demonstration of why the CES is only an approximation to the accounting identity.)

It should be noted that for, say, annual data, equation (7) is not an approximation of equation (6), but both equations are formally equivalent or isomorphic. This follows purely for mathematical reasons without recourse to any economic assumptions or arguments. As factor shares change only slowly over time without any pronounced secular trend, this explains intuitively why the Cobb-Douglas production function can often give such a good statistical fit.

In all these examples, it must be emphasised that the causation runs from the accounting identity to the “production function” and not the other way around, as the theoretical literature demonstrates that the latter does not exist, except under most implausible circumstances. There are, of course, some aggregation problems in summing industry accounting identities to obtain a single economy-wide accounting identity. However, these are considerably less severe than those involving aggregating micro-production functions. Ironically, Solow (1958) has shown that under plausible assumptions, and empirically, the aggregate factor shares will
show less variability than the individual shares. Thus, the Cobb-Douglas should give a better fit to the data at the higher levels of aggregation!

**Cost Functions and Accounting Identities**

We have seen how the existence of the underlying accounting identity means that we can always get a good fit to an aggregate production function. It is not surprising that the same applies to the cost function. The cost function shows how total costs vary with output in the light of fixed factor prices, i.e., \( C = f(Q) \) where \( C \) is total costs. This is derived on the assumption that the firm chooses the optimum combination of the factors of production given the relative factor prices. It is perhaps easiest to demonstrate our argument for the Cobb-Douglas case.

As any standard microeconomics textbook shows, the total cost function is obtained by maximising output given by the production function, namely in the case of the Cobb-Douglas production function:

\[
Q = AK^{\alpha}L^{\beta}
\]  

subject to the cost equation or accounting identity:

\[
C = wL + \rho K
\]

where \( C \) is assumed to be constant, i.e., the firm has a fixed budget to spend on both factors of production. This procedure is not seen as tautological because \( Q \) is assumed to be a homogeneous quantity, independent of the costs of production, although, of course, \( pQ = C \) where \( p \) is the price per unit of output \( Q \). Obtaining the first order conditions from the constrained maximisation problem and setting them equal to zero, it may shown through some straightforward algebra that the cost equation is given by:

\[
C = A^{-1/(\alpha+\beta)}\left[\left(\frac{\alpha}{\beta}\right)^{\beta} + \left(\frac{\beta}{\alpha}\right)^{\alpha}\right]^{1/(\alpha+\beta)}w^{\alpha/(\alpha+\beta)}\rho^{\beta/(\alpha+\beta)}Q^{1/(\alpha+\beta)}
\]
Thus, total costs depend upon the volume of output, the production coefficients, $\alpha$, $\beta$, and $A$ and the prices of the factors of production, $w$ and $\rho$. Equation (17) is interpreted as a behavioural relationship as it supposedly can be used to estimate the degree of returns to scale $(\alpha + \beta)$. Moreover, it is also seen as a testable hypothesis because, according to this interpretation, if firms were not productively efficient, even if the production function were a Cobb-Douglas, the estimation would give a very poor statistical fit.

However, if value data are used for output, then it is straightforward to show that we have a tautology again. To see this, let us assume constant returns to scale so that $\alpha + \beta = 1$. (But as we have seen the above the data will always imply “constant returns to scale” so this is not a restrictive assumption.) Under these circumstances, the expression is square brackets in equation (17), namely, $\left[(\alpha / \beta)^\beta + (\beta / \alpha)^\alpha\right]$, is equal to $\alpha^{-\alpha}(1-\alpha)^{-\alpha}$ and thus equation (17) may be written using value data as:

$$(18) \quad C = \alpha^\alpha (I - \alpha)^{-(I - \alpha)} w^\rho p^{(I - \alpha)} V / A$$

As $V / A \equiv L^\alpha J^{(I - \alpha)}$, it follows that equation (18) is:

$$(19) \quad C = \alpha^\alpha (I - \alpha)^{-(I - \alpha)} w^\rho p^{(I - \alpha)} L^\alpha J^{(I - \alpha)}$$

which is nothing more than the accounting identity given by equation (7), where $\alpha = a$ and $(I - \alpha) = (I - a)$ and $C = V$. Thus, if we use value data, equation (19) is definitionally true and does not need to be derived by the optimising procedure outlined above. The reason why equation (19), in logarithmic form, is seen as a behavioural equation is that if $w^\rho p^{(I - \alpha)}$ is proxied by a linear time trend, the statistical fit may be poor for reasons set out earlier in this paper.

The Implications of the Critique

Once it is appreciated that an estimated aggregate production function is simply capturing a (sometimes misspecified) accounting identity, it is often possible to determine the outcome of a regression analysis before a single regression has been run. The only additional information that is sometimes required are the “stylized facts” that factor shares are roughly constant, the capital-output ratio does not significantly change over time and the weighted growth of wages and the rate of profit is also constant. None of these assumptions is dependent on the
existence of a production function. There is not the space to go into all these studies in detail so we merely mention some of the more important ones.

Mankiw et al., (1992) estimated the Solow model, augmented by human capital, using cross-country data for both the developed and less developed countries. While they generally found a good statistical fit (apart from the OECD countries), nevertheless the estimate of the output elasticity of capital was greater than its factor share, when human capital was not included. Once they included human capital, the output of physical capital more closely approximated to its factor share. However, once again the results are merely capturing the underlying identity. Initially, this is misspecified by the assumption that the logarithm of the “level of technology” is common for all countries. But once dummies are introduced to allow for differences in “technology”, a better fit is obtained and the output elasticities of capital and labour are close to their factor shares, without the need to include human capital (Easterly and Levine, 2001). The explanation for this is straightforward. From the identity we know that $\ln A = a \ln w + (1-a) \ln r$ which shows considerable international variation. Consequently, the introduction of regional (continental) dummies, which largely capture this variation, gives a better estimate of the accounting identity. Therefore, the fact that Mankiw et al., find a very good fit to the aggregate production function using data from both the advanced and the developing countries is hardly surprising (Felipe and McCombie, 2005).

Oulton and O’Mahony (1994) test whether “capital is special”, i.e., whether the growth of capital contributes more to the growth of total factor productivity than that implied by its factor share. Using cross-industry growth data for 120 UK manufacturing industries and regression analysis, they find that the estimated elasticities of the factor inputs, including capital, do not differ significantly from their respective factor shares. They therefore conclude that “these results therefore provide no support for the view that the role of capital has been understated” (p.162). However, as all they are, in effect, estimating is essentially the identity, it is hardly surprising that they get this result (Felipe and McCombie, 2009a).

Feder (1983) developed a dual-sector model which purported to show the important externality effect the growth of exports has on the growth of output. This approach was adopted by Ram (1986) to examine the role of the growth of government expenditure, which likewise was found to impart an important externality effect. However, McCombie (1999) shows that the results are entirely driven by the estimated model being in effect a hybrid of the accounting identity and the sectoral identity. The latter is where the growth of output is definitionally equal to the growth of exports and /or government expenditure and the rest of the economy, each weighted by its share in output.
Estimates of the labour demand function almost always find a statistically significant negative coefficient on the logarithm of real wages and a wage elasticity with respect to the demand for labour of about -0.30 (Hamermesh, 1993). However, it can be shown that these results are entirely driven by the accounting identity, even when the rental price of capital (rather than the ex post rate of profit) is used as capital’s factor price. Consequently, no policy conclusions regarding, for example, the effect of a minimum wage, should be drawn from these results (Felipe and McCombie, 2009b).

In an influential paper, Hall (1988) attempted to estimate the degree of market power in U.S. manufacturing industry. (See also Hall, 1990.) This was undertaken by estimating a production function in growth rate form using time-series data and determining whether or not the coefficient on the growth of the labour–capital ratio exceed the value of labour’s share. He estimates $\mu$ in the equation $\dot{V} - \dot{J} = c + \mu(\dot{L} - \dot{J})$ and if it exceeds unity, market power is assumed to exist, as it shows that price exceeds marginal cost. With value added data, the coefficient did exceed unity and so Hall inferred that industries operated with market-power. However, we know from the identity that any estimate of the coefficients of the inputs of the putative production function should be equal to the factor shares and Hall should have found a coefficient of unity. Felipe and McCombie (2002) show that the sole reason for his contrary results was the upward bias induced by assuming the rate of “technical progress” is constant. In reality, as it merely measures the weighted growth of the factor prices, it has a pronounced pro-cyclical variation and this biases the coefficient upwards. The data cannot show the existence or otherwise of market power.

Other examples where the identity drives the regression results includes the calibrated real business cycle analysis of Hansen and Sargent (1990) (see the discussion by Hartley, 2000), the putative importance of the externality effect of total output growth when considering industry time-series productivity growth rate estimates (Felipe, 2001a, McCombie, 2000-2001) and the role of infrastructure in economic growth (Felipe, 2001b).

Given the importance of this critique, it is surprising that it has been almost totally ignored, misinterpreted or even greeted with outright hostility within the mainstream economics profession. But perhaps on reflection it is not all that surprising. Few people are willing to concede much of their academic work may be literally meaningless. Fisher (2005, p. 491) in the conclusion to his paper in a symposium on this critique ended with the warning “Don’t
interfere with fairytales if you want to live happily ever after”. A salutary warning to Post Keynesians everywhere?

References


