

On Herbert Simon's criticisms of the Cobb–Douglas and the CES production functions

Abstract: Scott Carter (2011) reproduces and discusses correspondence between Simon and Solow in 1971, where Simon first outlined his critique that estimations of production functions merely capture an underlying accounting identity. This idea culminated in a paper published by Simon in 1979 in the *Scandinavian Journal of Economics*. We extend Simon's argument that production functions should ideally be estimated using physical data, and discuss the serious problems that arise when they are estimated using constant-price monetary data. Simon also suggested that the good statistical fits to production functions could be derived from a markup pricing model, but he did not follow this up. We show that this can indeed account for the very good statistical fits of the Cobb–Douglas and other production functions. We conclude by showing how estimates of cost functions suffer from the same problem.

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

The correspondence between Herbert Simon and Robert Solow in 1971 published in Scott Carter's article (this issue, pp. 255–273) 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 artifacts 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

Jesus Felipe is Lead Economist at the Asian Development Bank. John McCombie is Director of the Cambridge Centre for Economic and Public Policy, Department of Land Economy, University of Cambridge, UK. This article reflects solely the opinions of the authors and does not necessarily reflect those of the Asian Development Bank, Manila, Philippines, its Executive Directors, or those countries that they represent. The authors are grateful to Scott Carter for making available to them an advance copy of his article.

outputs and inputs must hold by definition. Estimates of aggregate production functions reflect an identity and are not behavioral relationships. While there were hints of this critique in some earlier literature (e.g., Bronfenbrenner, 1944; Marshak and Andrews, 1944), the starting point is really buried in the paper by Phelps Brown (1957). It was this argument as it pertains to cross-section data that was, in effect, subsequently formalized 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. The exchange between Simon and Solow unearthed by 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 Carter points out, there are now a number of papers on the problems posed by the accounting identity. Importantly, Shaikh (1974, 1980, 1987)¹ generalized the critique to time-series estimations of supposed production functions, but his 1974 paper was criticized by Solow (1974) who perhaps, not surprisingly, defended his original approach in Solow (1957). This criticism is not convincing (Shaikh, 1980). Solow (1987) returned 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 (see also Shaikh, 2005). The criticism was reexamined and extended by Felipe (2001a, 2001b), Felipe and Adams (2005), Felipe and Holz (2001), Felipe and McCombie (2001, 2002, 2003, 2005a, 2005b, 2006, 2007, 2009a, 2009b, 2010), McCombie (1987, 1998a, 1998b, 1999, 2000, 2000–1, 2001), McCombie and Dixon (1991), and McCombie and Thirlwall (1994, pp. 93–103). The critique as applied to cross-section data was also rediscovered by Samuelson (1979).

It is interesting that Simon relies on the principle of parsimony, or Occam's razor, to argue in favor of the interpretation of the data as reflecting the accounting identity, rather than a production function. (See, in particular, the discussion in Carter's, 2011, conclusions.) This is a weak methodological principle because it implies that if the world is more complicated than we think, then the aggregate production function may,

¹ It is interesting to note that although Simon (1979a) 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.

in fact, exist. This is despite that a simpler explanation may be available. Yet Simon was aware of the limitations of this principle:

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 refutation 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 argues that

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 labour cost plus capital cost) (Phelps Brown, 1957). The same comment applies to the SMAC [Solow, Minhas, Arrow, and Chenery; see Arrow et al. 1961] production function. (1979b, p. 497)

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 use of constant-price monetary data precludes the researcher from ever rejecting the hypothesis that an aggregate production function exists and prevents the interpretation of the estimated coefficients as reflecting technological 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 May 4, 1971 (see Carter, this issue, pp. 263-266), where

² It is true that there are some engineering studies that use physical data, but these are few and far between.

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 1979 paper (Simon, 1979a).

A second point is that it has often been asserted erroneously that the critique applies *only* to the Cobb–Douglas production function (Temple, 2006; see also Felipe and McCombie’s, 2010, reply). Simon’s argument (1979a) was that in estimations using cross-section data, the estimated parameters of the CES (constant elasticity of substitution production function) 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 applies *only* to where the Cobb–Douglas production function gives a reasonably good fit to the data, even though only as an approximation to a more flexible functional form. We show below that this is not the case.

Finally, Simon (*ibid.*) cited Walters’s (1963) paper on production and cost functions. Walters reviewed a number of econometric studies that had estimated Cobb–Douglas production functions and cost curves. He found that the evidence on cost curves did not support the U-shaped long-run average cost curve consistent with neoclassical theory, although as Simon commented, this is “not so damning as to require the theory to be abandoned!” (Simon, 1979a, p. 470). In the current article, 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 function is unlikely to exist. There is a large technical literature on this problem that we do not discuss here (see Felipe and Fisher, 2003; Fisher, 1992), but to begin with, it is instructive to quote Fisher on the conclusions that can be drawn from this work.

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 labor 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. (2005, pp. 489–490)

However, in retrospect, these conclusions are hardly surprising. Take the manufacturing industry, for example. At the two-digit Standard Industrial Classification (SIC) level there are diverse industries such as tobacco products (SIC 21) and chemical and allied products (SIC 28). Does it make sense to aggregate the outputs of these two industries and their employment and capital stocks 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 the 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 (see Douglas, 1944, and references cited there). Furthermore, the estimated output elasticities are often very close to the factor shares obtained from the national income and product accounts, as predicted by the aggregate marginal productivity theory of factor pricing.³ 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 functions” (cited by Fisher, 1971, p. 305). But at the risk of getting ahead of ourselves,

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 utilization generally improves the results and gives plausible results. Furthermore, it is also possible to find a nonlinear 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.

we will show that Douglas could never have found the output elasticity of labor to be anything other than 75 percent.

The implications are far reaching. Douglas, 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. (1976, p. 914)

But the accounting identity renders this conclusion invalid.

The implicit defense of the use of the aggregate production function rests largely on a methodological instrumental argument. All economic 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, viewed the aggregate production function as an empirical law in its own right, capable of being statistically refuted, a view shared by Solow (1974). Ferguson explicitly made this instrumental defense in the context of the criticism of 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 in 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. (1969, p. xvii, 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 warning cited above that "one cannot escape the force of these results [of the aggregation literature] by arguing that aggregate production functions are only approximations" (2005, p. 490). In answering this question, we adopt a slightly different procedure to that adopted by Simon as reported in Carter (2011) and also used in 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

for 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, this issue, p. 264.) 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 that, 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

$$Q = f(K, L), \quad (1)$$

where 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 capital and labor. In other words,

$$pQ \equiv wL + \rho K, \quad (2)$$

where p is the price of widgets, w is the wage rate, and ρ is the rental price of capital of each machine. The variables p , w , and ρ are measured in monetary units. We can write the accounting identity in purely physical terms as

$$Q \equiv w'L + \rho'K, \quad (3)$$

where $w' = w/p$ and $\rho' = \rho/p$ are, respectively, the wage rate and rental price of capital expressed in terms of the numbers of widgets. It is often assumed that the wage rate and the rental price are determined in perfectly competitive markets.

As Q , K , and L are all measured in physical units, the production function is a behavioral equation and not an identity. Thus, it is possible to find no statistically significant relationship between Q , K , and L .

The marginal products of labor and capital are given by

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

as set out in Simon (1979a, p. 463). From Euler's theorem for homogeneous functions of degree one, $Q = f_L L + f_K K$.

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 of capital (α) and labor ($1 - \alpha$) equal their factor shares, that is, $\alpha = a$, $(1 - \alpha) = (1 - a)$. Thus, if the estimated coefficients of capital and labor equal the respective factor shares, then this shows that perfect competition prevails, or strictly speaking, the estimates do not refute the assumption that factor 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 erroneously 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 approximately $(\rho/i)K = J$, where i is the rate of discount so that $\rho K = rJ$, where r is the rate of profit and equals the discount factor. J is the constant price value of the capital stock. For generality, we now assume that ρ and r include any monopoly rents.⁵

Simon originally used the assumption of exogenous constant ratios (most notably, $Q = \theta K$, where θ is a constant) in the analysis in his May 4, 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 markup 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” (Carter, this issue, p. 264). In his letter of June 30, 1971, he was again willing “to take the markup of direct costs as exogenous à la Phelps Brown and others, and then discuss the stability of labor’s share directly—but somewhat inelegantly. *De gustibus . . .*” (ibid., p. 268). Nevertheless, surprisingly, he did not follow this up in his 1979 paper (Simon, 1979a).

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 $r = (V - wL)/J$, where V is value added.

Some neoclassical economists see this as a criticism of the accounting identity critique as they argue neoclassical production theory assumes optimization in perfectly competitive markets. But we could deduct monopoly profits from V so that ρ becomes the competitive rental price of capital and the argument carries through unaffected. See Felipe and McCombie (2007) for a detailed discussion; they 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 markup on unit labor costs.⁶

$$p = (1 + \pi)wL/Q. \quad (5)$$

The underlying accounting identity is given by

$$pQ \equiv wL + rJ \quad (6a)$$

or

$$V \equiv wL + rJ. \quad (6b)$$

In reality, as noted above, we have to move away from the simple model because there are heterogeneous output and capital goods that are aggregated using constant-price values. However, for simplicity, we remain in a one-good world, but assume that only V (value added) and J are known to the researcher, and not Q and K . Factor shares are constant, given by $a = \pi/(1 + \pi)$ (for capital) and $(1 - a) = 1/(1 + \pi)$ (for labor). Let us assume that unrealistically there exists an underlying production function and that this is given by a Cobb–Douglas. Through this production function, K and L determine Q . However, we assume, as is the case in practice, that factor shares of capital and labor 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 opposite of what is usually considered to be plausible.*

If we totally differentiate the accounting identity, Equation (6b), and then integrate it, we obtain

$$V \equiv a^{-a} (1 - a)^{-(1-a)} r^a w^{(1-a)} J^a L^{(1-a)}. \quad (7)$$

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 competition or the form of the production function. Indeed, we have not even assumed that production functions exist in well-defined forms.

If we were to estimate $V = AJ^\alpha L^{(1-\alpha)}$, either using cross-section data where the expression $a^{-a}(1 - a)^{-(1-a)}r^a w^{(1-a)}$ in Equation (7) is approximately constant or using time-series data where $r^a w^{(1-a)}$ 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.

⁶ Note for the accounting identity to give rise to the Cobb–Douglas we only need factor shares to be roughly constant. A constant markup will give this. Kaldor's (1955–56) distribution theory is an alternative explanation as to why this may be the case.

It is worth emphasizing that this is the case even when markets are imperfectly competitive. Consequently, the following hold:

$$\alpha = 0.75 \neq \hat{\alpha} = a = 0.25 \quad (\text{capital})$$

$$(1 - \alpha) = 0.25 \neq (1 - \hat{\alpha}) = (1 - a) = 0.75 \quad (\text{labor}),$$

where the hat over the parameter indicates that it is a statistical estimate. In other words, the estimates of the “output elasticities” always equal the value of the factor shares and not the true values.

What happens if the “true” production functions are characterized by increasing returns to scale? The use of value data will still give the estimates of the “output elasticities” equal to the factor shares. The intercepts will differ depending on whether there are actually constant returns to scale or increasing returns to scale, but the estimated coefficients of labor and capital *must* sum to unity (Felipe and McCombie, 2006).

Most importantly, suppose that because of factors such as differences in *x*-efficiency, differences in technology, and, in particular, aggregation problems, and the like, there is *no* well-defined relationship between the inputs and the outputs. The markup and the accounting identity will still result in a perfect fit to the Cobb–Douglas production function estimated using *V*, *J*, and *L*, with output elasticities again equal to the factor shares and the estimates indicating constant returns to scale. It is worth emphasizing this point. Even though there is *no* well-defined production function at either the micro- or the macrolevel, in that there is no stable mathematical relationship that explains output (*Q*) in terms of capital (*K*) and labor (*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, *ibid.*, using simulation analysis.)

Let us suppose now that 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 equal to their respective factor shares, although this is because of the use of value data. We may show this as follows. Differentiating the accounting identity with respect to time gives:

$$\dot{V}_t \equiv a_t \dot{r}_t + (1 - a_t) \dot{w}_t + a_t \dot{J}_t + (1 - a_t) \dot{L}_t, \quad (8)$$

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:⁷

⁷ Assuming biased technical change does not make any significant difference to the argument (McCombie and Dixon, 1991).

$$\dot{V}_t \equiv \lambda_t + \alpha_t \dot{J}_t + \beta_t \dot{L}_t, \quad (9)$$

where λ is the rate of technical progress.

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

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 that 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:

$$Y^{(\eta)} = \begin{cases} \frac{Y^\eta - 1}{\eta} & \eta \neq 0 \\ \ln Y & \eta = 0. \end{cases} \quad (10)$$

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

$$V^{(\eta)} = c + b_1 J^{(\eta)} + b_2 L^{(\eta)}. \quad (11)$$

If $\eta = 1$, and the regression goes through the origin, we have the linear accounting identity. If $\eta = 0$, we have the familiar Cobb-Douglas. What happens if $\eta \leq 1$? Consider the CES "production function":

$$V = \gamma \left[\delta J^{-\rho} + (1 - \delta) L^{-\rho} \right]^{-1/\rho} \quad (\gamma > 0; 1 > \delta > 0; \infty \geq \rho \geq -1), \quad (12)$$

where γ is interpreted as an efficiency parameter, δ is a distributional parameter, and the elasticity of substitution is given by $\sigma = 1/(1 + \rho)$. It is assumed that there are constant returns to scale, but we have seen that the data and accounting identity, which must always be satisfied, implies this, so it is not an arbitrary assumption.

Equation (12) may be expressed as

$$V^{-\rho} = \left\{ \gamma^{-\rho} \delta \right\} J^{-\rho} + \left\{ \gamma^{-\rho} (1 - \delta) \right\} L^{-\rho}. \quad (13)$$

Compare this with the extended Box-Cox transformation when $\eta \leq 1$ and the constant term is constrained to equal zero:

$$V^\eta = b_3 J^\eta + b_4 L^\eta. \quad (14)$$

It follows that $\eta = -\rho$. 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 markup varies over time, either randomly or because of changes in the market power of firms and the bargaining power of labor. 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, data for one year, Equation (7) is not an approximation of Equation (6b), 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 emphasized that the causation runs from the accounting identity to the “production function” and not the other way around (Simon and Levy, 1963), as the theoretical literature demonstrates that the latter does not exist, except under the 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 microproduction 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 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, that is, $C = f(Q)$, where C is total costs. This is derived on the assumption that the firm chooses the optimum combination of the factors of production given relative factor prices. It is easiest to demonstrate our argument with the Cobb–Douglas case.

As any standard microeconomics textbook shows, the total cost function is obtained by maximizing output given by the production function (in the case of the Cobb–Douglas production function):

$$Q = AK^\alpha L^\beta \quad (15)$$

subject to the cost equation or accounting identity:

$$C = wL + \rho K, \quad (16)$$

where C is assumed to be constant, that is, 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 maximization problem and setting them equal to zero, it may be shown through some straightforward algebra that the cost equation is given by

$$C = A^{-1/(\alpha+\beta)} \left[\left(\frac{\beta}{\alpha} \right)^\alpha + \left(\frac{\alpha}{\beta} \right)^\beta \right]^{1/(\alpha+\beta)} \rho^{\alpha/(\alpha+\beta)} w^{\beta/(\alpha+\beta)} Q^{1/(\alpha+\beta)}. \quad (17)$$

Thus, total costs depend on the volume of output; the production parameters, α , β , and A ; and the prices of the factors of production, namely, w and ρ .⁸ Equation (17) is interpreted as a behavioral relationship as it can supposedly 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 $\alpha + \beta = 1$. (As was shown above, the data and the accounting identity will always imply “constant returns to scale.”) Under these circumstances, the expression in square brackets in Equation (17), namely, $[(\beta/\alpha)^\alpha + (\alpha/\beta)^\beta]$, is equal to $\alpha^{-\alpha}(1-\alpha)^{-(1-\alpha)}$, and thus, Equation (17) may be rewritten using value data as

$$C = \alpha^{-\alpha} (1-\alpha)^{-(1-\alpha)} r^\alpha w^{(1-\alpha)} V/A. \quad (18)$$

⁸ As we have noted, neoclassical production theory often assumes that the rental price of capital (and the wage rate) is set in competitive markets. This does not affect the argument for the reasons set out in note 5.

As $V/A \equiv L^\alpha J^{(1-\alpha)}$, it follows that Equation (18) is

$$C = \alpha^{-\alpha} (1 - \alpha)^{-(1-\alpha)} r^\alpha w^{(1-\alpha)} J^\alpha L^{(1-\alpha)}, \quad (19)$$

which is nothing more than the accounting identity given by Equation (7), where $\alpha = a$, $(1 - \alpha) = (1 - a)$, and $C = V$ when measured at the same price level. Thus, if we use value data, Equation (19) is definitionally true and does not need to be derived by the optimizing procedure outlined above. The reason Equation (19), in logarithmic form, is seen as a behavioral equation is that if $r^\alpha w^{(1-\alpha)}$ is proxied by a linear time trend, the statistical fit may be poor, for reasons set out earlier in this article.

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 single regression analysis before it has been run. The only additional information that is sometimes required to interpret what some researchers have done 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 [Organisation for Economic Co-operation and Development] sample), 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 elasticity of physical capital more closely approximated its factor share. However, as Felipe and McCombie (2005b) show, the results merely capture 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 labor 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 $\ln A \equiv a \ln r + (1 - a) \ln w$, 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. (1992) find a very good fit to the aggregate production function using data from both the advanced and the developing countries is hardly surprising.

Oulton and O'Mahony (1994) tested the hypothesis that "capital is special," that is, whether or not 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, they found that the estimated elasticities of the factor inputs, including capital, do not differ significantly from their respective factor shares. They concluded that "these results therefore provide no support for the view that the role of capital has been understated" (*ibid.*, p. 162). However, as all Oulton and O'Mahony are, in effect, estimating is essentially the identity, it is hardly surprising that they get this result (Felipe and McCombie, 2009b).

Feder (1983) developed a dual-sector model that purported to show the important externality effect that 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) showed that the results are entirely driven by the fact that the estimated model is, in effect, a hybrid of the accounting identity and the sectoral identity, where the latter shows that 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 labor 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 labor 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, 2009a).

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 labor-capital ratio is statistically significantly

greater than labor's share. Specifically, he estimated the equation $(V_t - J_t) = c + \mu((1 - a_t)(L_t - J_t))$ and tested whether μ exceeds unity. If this is the case, then Hall concludes that market power exists, as this result shows that price exceeds marginal cost. Hall found that the coefficient did exceed unity and so he 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 μ of unity. Felipe and McCombie (2002) showed that the sole reason for Hall's results was the upward bias induced by assuming that the rate of "technical progress" is constant. In reality, as it merely measures the weighted growth of the factor prices, it has a pronounced procyclical variation, and this biases the coefficient upward. 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 also 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, 2001b; McCombie, 2000–1); and the role of infrastructure in economic growth (Felipe, 2001a).

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 that much of their academic work may be literally meaningless. Fisher, in the conclusion to his article in a symposium on this critique, ended with the warning, "[d]on't interfere with fairytales if you want to live happily ever after" (2005, p. 491). A salutary warning to Post Keynesians everywhere?

REFERENCES

- Arrow, K.; Chenery, H.B.; Minhas, S.; and Solow, R.M. "Capital-Labor Substitution and Economic Efficiency." *Review of Economics and Statistics*, 1961, 43 (3), 225–250.
- Bronfenbrenner, M. "Production Functions: Cobb–Douglas, Interfirm, Intrafirm," *Econometrica*, 1944, 12 (1), 35–44.
- Carter, S. "On the Cobb–Douglas and All That . . . : The Solow–Simon Correspondence over the Aggregate Neoclassical Production Function." *Journal of Post Keynesian Economics*, Winter 2011–12, 34 (2), 255–273.
- Cohen, A., and Harcourt, G.C. "Whatever Happened to the Cambridge Capital Theory Controversies?" *Journal of Economic Perspectives*, 2003, 17 (1), 199–214.

- Douglas, P.H. "Are There Laws of Production?" *American Economic Review*, 1944, 38 (1), i-ii, 1-41.
- . "The Cobb-Douglas Production Function Once Again: Its History, Its Testing, and Some Empirical Values." *Journal of Political Economy*, 1976, 84 (5), 903-915.
- Easterly, W., and Levine, R. "What Have We Learned from a Decade of Empirical Research on Growth? It's Not Factor Accumulation: Stylized Facts and Growth Models." *World Bank Economic Review*, 2001, 15 (2), 177-219.
- Feder, G. "On Exports and Economic Growth." *Journal of Development Economics*, 1983, 12 (1-2), 59-73.
- Felipe, J. "Aggregate Production Functions and the Measurement of Infrastructure Productivity: A Reassessment." *Eastern Economic Journal*, 2001a, 27 (3), 323-344.
- . "Endogenous Growth, Increasing Returns, and Externalities: An Alternative Interpretation of the Evidence." *Metroeconomica*, 2001b, 52 (4), 391-427.
- Felipe, J., and Adams, G.F. "A Theory of Production': The Estimation of the Cobb-Douglas Function: A Retrospective View." *Eastern Economic Journal*, 2005, 31 (3), 427-445.
- Felipe, J., and Fisher, F.M. "Aggregation in Production Functions: What Applied Economists Should Know." *Metroeconomica*, 2003, 54 (2-3), 208-262.
- Felipe, J., and Holz, C.A. "Why do Aggregate Production Functions Work? Fisher's Simulations, Shaikh's Identity and Some New Results." *International Review of Applied Economics*, 2001, 15 (3), 261-285.
- Felipe, J., and McCombie, J.S.L. "The CES Production Function, the Accounting Identity and Occam's Razor." *Applied Economics*, 2001, 33 (10), 1221-1232.
- . "A Problem with Some Recent Estimations and Interpretations of the Mark-Up in Manufacturing Industry." *International Review of Applied Economics*, 2002, 16 (2), 187-215.
- . "Some Methodological Problems with the Neoclassical Analysis of the East Asian Economic Miracle." *Cambridge Journal of Economics*, 2003, 27 (5), 695-721.
- . "How Sound Are the Foundations of the Aggregate Production Function?" *Eastern Economic Journal*, 2005a, 31 (3), 467-488.
- . "Why Are Some Countries Richer than Others? A Skeptical View of Mankiw-Romer-Weil's Test of the Neoclassical Growth Model." *Metroeconomica*, 2005b, 56 (3), 360-392.
- . "The Tyranny of the Identity: Growth Accounting Revisited." *International Review of Applied Economics*, 2006, 20 (3), 283-299.
- . "On the Rental Price of Capital and the Profit Rate: The Perils and Pitfalls of Total Factor Productivity Growth." *Review of Political Economy*, 2007, 19 (3), 317-345.
- . "Are Estimates of Labour Demand Functions Mere Statistical Artefacts?" *International Review of Applied Economics*, 2009a, 23 (2), 147-168.
- . "Why the Data Tell Us Nothing About Returns to Scale and Externalities to Capital in Economic Growth." *Economia e Sociedade*, December 2009b, 17, 655-677.
- . "What Is Wrong with Aggregate Production Functions: On Temple's 'Aggregate Production Functions and Growth Economics.'" *International Review of Applied Economics*, 2010, 24 (6), 665-684.
- Ferguson, C.E. *The Neoclassical Theory of Production and Distribution*. Cambridge: Cambridge University Press, 1969.
- Fisher, F.M. "Aggregate Production Functions and the Explanation of Wages: A Simulation Experiment." *Review of Economics and Statistics*, 1971, 53 (4), 305-325.

———. *Aggregation: Aggregate Production Functions and Related Topics*. Edited by J. Monz. London: Harvester Wheatsheaf, 1992.

———. "Aggregate Production Functions—A Pervasive, but Unpersuasive, Fairytale." *Eastern Economic Journal*, 2005, 31 (3), 489–491.

Friedman, M. "The Methodology of Positive Economics." In M. Friedman (ed.), *Essays in Positive Economics*. Chicago: Chicago University Press, 1953, pp. 3–43.

Hall, R.E. "The Relation Between Price and Marginal Cost in U.S. Industry." *Journal of Political Economy*, 1988, 96 (5), 921–947.

———. "Invariance Properties of Solow's Productivity Residual." In P. Diamond (ed.), *Growth/Productivity/Employment*. Cambridge: MIT Press, 1990, pp. 71–122.

Hamermesh, D.S. *Labor Demand*. Princeton: Princeton University Press, 1993.

Hansen, L.P., and Sargent, T.J. "Recursive Linear Models of Dynamic Economies." Working Paper Series no. 3479, National Bureau of Economic Research, Cambridge, MA, 1990.

Hartley, J.E. "Does the Solow Residual Actually Measure Changes in Technology?" *Review of Political Economy*, 2000, 12 (1), 27–44.

Kaldor, N. "Alternative Theories of Distribution." *Review of Economic Studies*, 1955–56, 23 (2), 83–100.

Mankiw, N.G.; Romer, D.; and Weil, D.N. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics*, 1992, 107 (2), 407–437.

Marshall, J., and Andrews, W.H. "Random Simultaneous Equations and the Theory of Production." *Econometrica*, 1944, 12 (3–4), 143–205.

McCombie J.S.L. "Does the Aggregate Production Function Imply Anything About the Laws of Production? A Note on the Simon and Shaikh Critiques." *Applied Economics*, 1987, 19 (8), 1121–1136.

———. "Are There Laws of Production? An Assessment of the Early Criticisms of the Cobb–Douglas Production Function." *Review of Political Economy*, 1998a, 10 (2), 141–173.

———. "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, UK: Edward Elgar, 1998b, pp. 44–68.

———. "A Problem with the Empirical Neoclassical Analysis of Economic Growth." In M. Setterfield (ed.), *Growth, Employment and Inflation: Essays in Honour of John Cornwall*. Basingstoke, UK: Macmillan, 1999, pp. 127–148.

———. "Regional Production Functions and the Accounting Identity: A Problem of Interpretation." *Australasian Journal of Regional Studies*, 2000, 6 (2), 133–155.

———. "The Solow Residual, Technical Change, and Aggregate Production Functions." *Journal of Post Keynesian Economics*, 2000–1, 23 (2), 267–297. (Errata, vol. 23, no. 3, 544.)

———. "What Does the Aggregate Production Function Tell Us? Second Thoughts on Solow's 'Second Thoughts on Growth Theory.'" *Journal of Post Keynesian Economics*, 2001, 23 (4), 589–615.

McCombie, J.S.L., and Dixon, R. "Estimating Technical Change in Aggregate Production Functions: A Critique." *International Review of Applied Economics*, 1991, 5 (1), 24–46.

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

Oulton, N., and O'Mahony, M. *Productivity and Growth: A Study of British Industry, 1954–1986*. Cambridge: Cambridge University Press, 1994.

- Phelps Brown, E.H. "The Meaning of the Fitted Cobb-Douglas Function." *Quarterly Journal of Economics*, 1957, 71 (4), 546-560.
- Ram, R. "Government Size and Economic Growth: A New Framework and Some Evidence from Cross-Section and Time-Series Data." *American Economic Review*, 1986, 76 (1), 191-203.
- Samuelson, P.A. "Paul Douglas's Measurement of Production Functions and Marginal Productivities." *Journal of Political Economy*, 1979, 87 (5), 923-939.
- Shaikh, A. "Laws of Production and Laws of Algebra: The Humbug Production Function." *Review of Economics and Statistics*, 1974, 56 (1), 115-120.
- . "Laws of Production and Laws of Algebra: Humbug II." In E.J. Nell (ed.), *Growth, Profits and Property, Essays in the Revival of Political Economy*. Cambridge: Cambridge University Press, 1980, pp. 80-95.
- . "Humbug Production Function." In J. Eatwell, M. Milgate, and P. Newman (eds.), *The New Palgrave: A Dictionary of Economic Theory and Doctrine*. London: Macmillan, 1987, pp. 690-692.
- . "Nonlinear Dynamics and Pseudo-Production Functions." *Eastern Economic Journal*, 2005, 31 (3), 447-466.
- Simon, H.A. "On Parsimonious Explanation of Production Relations." *Scandinavian Journal of Economics*, 1979a, 81 (4), 459-474.
- . "Rational Decision Making in Business Organizations." *American Economic Review*, 1979b, 69 (4), 493-513.
- Simon, H.A., and Levy, F.K. "A Note on the Cobb-Douglas Function." *Review of Economic Studies*, 1963, 30 (2), 93-96.
- Solow, R.M. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics*, 1957, 39 (3), 312-320.
- . "A Skeptical Note on the Constancy of Relative Shares." *American Economic Review*, 1958, 48 (4), 618-631.
- . "Laws of Production and Laws of Algebra: The Humbug Production Function: A Comment." *Review of Economics and Statistics*, 1974, 56 (1), 121.
- . "Second Thoughts on Growth Theory." In A. Steinherr and D. Weiserbs (eds.), *Employment and Growth: Issues for the 1980s*. Dordrecht: Martinus Nijhoff Publishers, 1987, pp. 13-28.
- Temple, J.R.W. "Aggregate Production Functions and Growth Economics." *International Review of Applied Economics*, 2006, 20 (3), 301-317.
- Walters, A.A. "Production and Cost Functions: An Econometric Survey." *Econometrica*, 1963, 31 (1-2), 1-66.
- Wan, H.Y. *Economic Growth*. New York: Harcourt Brace Jovanovich, 1971.