'Are There Laws of Production?': an assessment of the early criticisms of the Cobb–Douglas production function

J. S. L. McCOMBIE
Downing College, Cambridge CB1 2DQ, UK

This paper traces the development of the Cobb–Douglas production function from its inception in 1927 and critically assesses its early hostile reception. Further econometric evidence is also presented on these issues. Some of the criticisms were easily dealt with, but other more serious ones remained and, although equally relevant today, have been all but forgotten. The original regressions of Cobb and Douglas using time-series data produced some spectacularly good fits, with the estimates of the output elasticities being virtually identical to the relevant factor shares. (This was erroneously argued by Douglas, and others following him, as providing strong empirical support for the neoclassical marginal productivity theory of distribution.) It is shown that these results collapse once account is taken of the existence of either outliers or technical change, or both. There is some evidence that Douglas himself realised this and his emphasis subsequently shifted to cross-industry regressions. However, an important critique by Phelps Brown in 1957, formalised later by Simon & Levy, demonstrated that all that was being estimated was an accounting identity. This criticism was later generalised by Shaikh to time-series estimations. These critiques have been largely brushed aside and ignored in the literature. If they had not been, there would perhaps be a greater appreciation of just how flimsy is the theoretical basis of the production function.

1. Introduction

The year 1927 represents a landmark in the development of economics. It ranks alongside 1871 (the year of publication of Jevons' *Theory of Political Economy* and Carl Menger's *Grundsätze*), 1936 (the year of the *General Theory*) and 1961 (Muth's 'Rational expectations and the theory of price movements'). It was in 1927 that, at the annual meeting of the American Economic Association, Charles Cobb and Paul Douglas first promulgated the results of the estimation of their now famous aggregate production function. Here, for the first time, was supposedly empirical support for both the existence of a well-defined aggregate production function and the marginal productivity theory of distribution. Their controversial paper was published the following year in the *American Economic Review*.

The importance of some seminal works is immediately apparent on (or even

I am grateful to Geoff Harcourt, Tony Thirlwall and two anonymous referees for helpful comments. The usual disclaimer applies.

0953-8259/98/020141-33

prior to) their publication. The *General Theory* and the discussions of the Cambridge Circus come readily to mind. Some have never really lived up to their initial promise (for example, Joan Robinson's 1933 *Economics of Imperfect Competition*—as conceded by Robinson herself; see Dobb, 1973, p. 212, although there has been a revival of interest in this approach in recent years). With others, such as Walras's *Éléments* (1874) and Muth's (1961) article, it took years, if not decades, for the importance to become fully appreciated. But it would be difficult to find a path-breaking study that was received initially with such outright hostility and criticism as that of Cobb and Douglas. 'Our paper met with a very hostile reception, and the next few years were full of the most caustic criticism. I think no one said a good word about what we had tried to do' (Douglas, 1967, p. 17). What was especially discouraging for Douglas, in particular, was that from none was the criticism more vehement than his own senior colleagues at the University of Chicago. Eventual vindication and recognition, though putatively, came in December 1947 when Douglas gave the sixtieth presidential address to the American Economic Association at Chicago (published subsequently with the rhetorical title 'Are there laws of production?' in the 1948 *American Economic Review*).

The phrase 'laws of production' refers to a mathematical function that represents a technological relationship between the maximum value of output and the factor inputs (together with a time-trend, if time-series data are used), so that the factor inputs, usually labour and capital (and the time-trend, if appropriate), give a predicted value of output that is often very close to the observed value. Indeed, it could be argued that this is the nearest economics comes to a law comparable to those found in the physical sciences. However, it will be shown that there are no grounds for such a view. This is not to deny that there are 'laws of production' in the general sense that, for a particular individual production process, the volume of output is related in some systematic way to the physical quantities of inputs. What is denied is that empirical evidence using constant price value data can determine whether or not the aggregate production function actually exists.

Today, when the Cobb–Douglas production function and its subsequent generalisations such as the Constant Elasticity of Substitution [CES] and translog production functions are, rightly or wrongly, such indispensable tools for the majority of economists, it is easy to overlook the initial unflattering reception that Cobb and Douglas's work received.

In this paper, the development of the Cobb–Douglas production function is traced from its inception, and these early criticisms are assessed. This is an interesting exercise in the history of economic thought in its own right; but I also demonstrate the continuing relevance of some of these criticisms. With the renaissance of the aggregate production function after Solow's seminal paper in 1957, most of the unresolved criticisms were largely forgotten, occasionally to be rediscovered (sometimes more than once). To take just one example: Samuelson (1979), in his far from hagiographic tribute to Douglas on the latter's death, raised some perceptive and fundamental criticisms of the specification and estimation of, especially, the cross-industry production function regressions. He concluded that 'it is a late hour to raise these doubts about the Emperor's clothes, but not until undertaking the present assignment did this child give the matter of across-industry fitting the careful attention it deserves and does not appear to have received'. Samuelson seems to have been unaware that virtually identical criticisms had been raised in an important paper by Phelps Brown (1957), elegantly formalised by Simon & Levy (1963) and restated by Simon (1979). These papers were not published in obscure journals but in prestigious periodicals, which makes their neglect by not only Samuelson but also the profession at large all the more surprising.

It may be that the subsequent debates concerning the aggregate production function, most notably the Cambridge Capital Theory Controversies (Harcourt, 1972) and the stringent assumptions needed for aggregation (Fisher, 1969), overshadowed these earlier reservations. The production function is essentially a microeconomic concept and aggregation theory considers the conditions under which firms' production functions can be aggregated to give a well-defined macroeconomic production function. The work of Fisher (1992) has demonstrated that the conditions are so stringent as to make the aggregate production function a dubious concept.

The Cambridge Capital Theory Controversies are, to a large extent, a distinct, but related, problem (Harcourt, 1972, in his masterful summary, only briefly touches on the aggregation problem). The Capital Theory Controversies involve the problems that, *inter alia*, a value measure of capital cannot be defined independently of the rate of profit without circular reasoning and, under many conditions, there is no guarantee that the rate of profit will be uniquely inversely related to the capital–labour ratio. It is notable that the implications of these two critiques have been brushed aside, the questions raised largely unanswered. One reason stems from the original work of Cobb and Douglas. Estimations of the aggregate production function generally produce good statistical fits, and in the case of the Cobb–Douglas production function, the exponents are close to the expected factor shares. (Cobb & Douglas, 1928, and Douglas, 1934, using time series data for US manufacturing industry over the period 1890–1922 found correlation coefficients that were over 0.9 and output elasticities that were within one or two percentage points of the observed relevant factor shares.) Consequently, following Friedman's (1953) influential instrumental approach, it is inferred that these critiques, notwithstanding their logical correctness, are of no practical importance (McCombie, 1997a).

I shall not be concerned in this paper with the more recent issues raised by the consideration of the conditions for successful aggregation and by the Capital Theory Controversies, important though they are. Moreover, for reasons of space, I shall also not deal with econometric problems involved in estimating production functions, such as issues of identification and simultaneous equation bias (see Intriligator, 1978, and Wallis, 1973, for good introductions). Instead, I shall be concentrating primarily on the earlier critiques. Some of these were satisfactorily dealt with by Douglas and his collaborators, but others remain and are just as damaging as the better-known capital controversies and aggregation issues. In particular, it will be shown, following Phelps Brown (1957) and subsequent generalisations of his argument, that reliance on the instrumental position that aggregate production functions empirically 'work' is untenable. No less an authority as Joan Robinson (1970) commented that the aggregate production function 'must have needed an even tougher hide to survive Phelps Brown's (1957) article on 'The meaning of the fitted "Cobb–Douglas function" than to ward off Cambridge Criticism of the marginal productivity theory of distribution'.

The original good fits of Cobb & Douglas (1928) seemed to be subsequently confirmed by further time-series results for Massachusetts and New South Wales. But a reworking of Douglas's original data by Mendershausen (1938) and by the author (inspired by Mendershausen's and certain of the other early criticisms) show that the estimates are not robust and, in fact, offer no support for the marginal productivity theory, as Douglas argued, even by his own criteria. Indeed, these revised results would lead one to doubt the empirical existence of an aggregate production function. There is exegetical evidence that Douglas (1934, 1948) began to realise that his results using time-series data would not bear the interpretation he originally placed on them, although not surprisingly he played down his reservations. The outcome was that he increasingly shifted the emphasis to regression analysis using cross-industry data, especially as the results appeared to be considerably more stable. Indeed, for the two decades of the 1930s and 1940s there was an impressive amount of replication of the estimation of the Cobb–Douglas production function, using a variety of different cross-industry data sets. (See, *inter alia*, Handsaker & Douglas, 1937; Bronfenbrenner & Douglas, 1939; Gunn & Douglas, 1941, 1942; Daly & Douglas, 1943; and Daly *et al.*, 1943.) Nevertheless, as I have noted above and will show below, these results are not immune to very serious criticisms. It will become clear that many of these early criticisms, and their subsequent elaborations, have never been satisfactorily answered and are as relevant today as they were several decades ago.

2. The Initial Development and Estimation of the Aggregate Production Function

Before these and other criticisms are pursued further, it is useful to consider the original study of Cobb and Douglas. It is difficult to do better than to quote Douglas (1948) himself as to how they came to settle on the Cobb–Douglas relationship.

> It was twenty years ago last spring that, having computed indexes for American manufacturing of the numbers of workers employed by years from 1899 to 1922, as well as indexes of the amounts of fixed capital in manufacturing deflated to dollars of approximately constant purchasing power, and then plotting these on a log scale together with the Day index of physical production for manufacturing, I observed that the product curve lay consistently between the two curves for the factors of production and tended to be approximately a quarter of the relative distance between the curve index for labor, which showed the least increase over the period and that of an index of capital, which showed the most. Since I was lecturing at Amherst College at the time, I suggested to my friend, Charles W. Cobb, that we seek to develop a formula which would measure the relative effect of labor and capital upon

product during this period. We were both familiar with the Wicksteed analysis and Cobb was, of course, well versed in the history of the Euler theorem. At his suggestion, therefore, the sum of exponents was tentatively made equal to unity in the formula

$$P = bL^kC^{1-k} \quad (1)$$

[where P, L and C are output, labour and capital].

Here it was only necessary to find the values of b and k. This was done by the method of least squares and the value of k was found to be .75. This was almost precisely what we had expected to find because of the relative distance of the product curve from those of the two factors. The value of the capital exponent, or $1-k$, was, of course, then taken as .25. Using these values, we then computed indexes of what we would theoretically have expected the product to be in each of the years had it conformed precisely to the formula. We found that the divergencies between the actual and theoretical product were not great since in only one year did they amount to more than 11 per cent, and that except for two years, the deviation of the differences was precisely what we would expect from the imperfect nature of the indexes of capital and labor.

The Cobb–Douglas production function also putatively provides a joint test of the aggregate marginal productivity theory of factor pricing and perfect competition. It will be recalled that, in Cobb and Douglas's notation, the production function is given by $P = bL^kC^{1-k}$. If factors are paid their marginal products then the real wage is given by $w = \partial P/\partial L = kP/L$. Similarly, the rate of profit is given by $r = \partial P/\partial C = (1-k)P/C$. Given the underlying accounting definition of value added, namely, $P \equiv wL + rC$, it can be seen that, by substituting in the marginal productivity conditions, the product will be exactly exhausted by the factor payments. Moreover, from the condition that $w = kP/L$, an equation for the elasticity of output with respect to labour may be obtained as $k = wL/P$, which equals labour's share in value added. Likewise, it may be simply shown that capital's output elasticity will equal capital's share in output.

The studies of the *National Bureau of Economic Research* on income were available in 1927 and these showed that labour's share in value added over the period 1909–19 was almost constant and the average was 74.1%. This was virtually identical to the estimate of k, a fact which convinced Douglas that he was on the right track. A major rationale for his work on the aggregate production function was explicitly to test and quantify the neoclassical marginal productivity theory of distribution. For example, Douglas complained, in his 1948 address, that he had observed an 'experienced instructor' drawing marginal productivity curves on the blackboard without the faintest idea of what the slope of the curve should look like. Now, for the first time, there was some empirical evidence. The estimates of the production function implied an elasticity of demand for capital of -4 and for labour of $-1\frac{1}{3}$. As Samuelson (1979) points out, Douglas was reassured to find the former was close to Pigou's (1933) 'deductive estimate' of -3 (see Douglas, 1934, *Addendum*, pp. xvii–xviii). All these findings gave Cobb and Douglas the confidence to proceed and present their paper to the December 1927 meeting of the American Economic Association. In 1976, looking back over his work, Douglas concluded that 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' (Douglas, 1976, p. 914). (I shall show below that there is no justification for this view.)

The impetus for Douglas's research may be traced back to 1888 when John Bates Clark effectively rediscovered Von Thünen's principle of marginal productivity (first outlined in the latter's *Der Isolierte Staat*, 1826). Clark's definitive statement of the principle appeared in his book *The Distribution of Wealth* published in 1899. At about this time, Wicksteed and Wicksell were also elaborating on distribution theory. Douglas, in fact, mistakenly gave priority for the discovery of the Cobb–Douglas production function to Wicksteed (rather than Wicksell) citing Wicksteed's *Essay on the Coordination of the Laws of Distribution* (1894). In choosing the multiplicative specification of the production function, Cobb & Douglas (1928) also referred to the theory as being 'due to J. B. Clark, Wicksteed *et al.*'

Wicksteed, it is true, did extend the marginal principle from the utility function and the determination of the pricing of commodities to the pricing of factors of production. Furthermore, he discussed the linear homogeneous production function in general terms. 'If we have $\pi = \phi(A,K)$ then we also have $m\pi = \phi(mA, mK)$.' From here, he demonstrated Euler's theorem 'without knowing that Euler had done it more than a hundred years before' (Sandelin, 1976, p. 118). Wicksteed, however, unsuccessfully tried to demonstrate the 'adding-up problem'; namely, that if there is perfect competition, then only if the production function is homogeneous and of the first degree will the payment to factors of their marginal products exactly exhaust total output. It was left to Flux (1894) in his review of Wicksteed to 'give an elementary but elegant proof of Wicksteed's contention' (Blaug, 1978, p. 463). Moreover, Wicksteed did not explicitly specify a Cobb–Douglas production function. In fact, he describes a production function where, as more and more labour is added to a fixed amount of land, the volume of total output eventually actually diminishes. 'That is, the marginal product of labour switches from positive to negative, a property which excludes the Cobb–Douglas function as a possible description of the "laws of production", at any rate in the neighbourhood of the switching point' (Sandelin, 1976, p. 119).

As Sandelin (1976) and Samuelson (1979) point out, it is clear that the credit should have gone to Wicksell instead. In his classic paper 'Marginal productivity as the basis of distribution in economics', published in *Ekonomisk Tidskrift* in 1900, Wicksell also considered the adding-up problem and it was here that the function used by Cobb and Douglas was alluded to for the first time.[1]

Wicksell (1900) presented the argument so succinctly that it is worth quoting him:

> The matter is quite simple from the mathematical viewpoint. If we consider the

[1] Sandelin (1976, p. 119) has also noted that the origins of the Cobb–Douglas production function can be traced back even further, namely to Wicksell (1895). However, the production function is implicit, rather than explicit, in this work.

product P as a function of the number of workers, a, and the number of acres of land, b, [these being the only two factors of production] the marginal productivities are the partial differential coefficients of P with respect to a and b, so that we have

$$a\frac{\delta P}{\delta a} + b\frac{\delta P}{\delta b} = P.$$

The general solution to this equation is

$$P = a.f(\frac{b}{a})$$

where $f(b/a)$ is an arbitrary function. In other words, P must be a *homogeneous* and *linear* function of a and b. Among the infinite number of functions with these properties we may select: $P = a^{\alpha}b^{\beta}$, where α and β are two constant fractions whose sum is 1.

It was perhaps natural that Cobb should suggest fitting the data to such a function as $P = a^{\alpha}b^{\beta}$, not least that it already had recognition as 'a well-known theory' (Cobb & Douglas, 1928, p. 151). As Samuelson notes (1979), Wicksell seems to have discovered the relationship by backing into it, 'beginning with the simplest square-root examples such as $bL^{\frac{1}{2}}$.[2]

3. The First Reaction to the Cobb–Douglas Production Function

We have noted that much of the early reaction to Douglas's first paper was hostile. Reflecting on this early denigration (which is not too strong a word) of his work, Douglas (1948) considered the critics could be divided into three camps.

First, there were the 'institutionalists' who decried any type of statistical or econometric work. It should be remembered that econometrics was in its infancy and was still far from being generally accepted as a useful tool. While the Econometric Society had been founded in 1930, the major impetus for statistical analysis can be traced to the work of the Cowles Commission in the 1940s and, in particular, to Haavelmo's 'methodological manifesto', 'The Probability Approach in Econometrics' (1944) (de Marchi & Gilbert, 1989, p. 2).

Secondly, there were the pure neoclassical theorists who believed it was fruitless to try to assign numerical values to the parameters of a theory that they regarded as intrinsically unquantifiable. The 1930s saw the attempt to establish a quantitative revolution in economics. 'The econometricians of the 1930s had a strong sense that it was part of their mission to make economics "operational".... However, the extended conflict over the desirability of mathematical economics, while vociferous, probably involved philosophical preju-

[2] Douglas was well aware of other possible specifications. One of his associates, Sidney Wilcox, suggested as early as 1926 the relationship $P = bR^{(1-k-j)}L^kC^j$, where R is defined as the combination of inputs given by $(L^2 + C^2)^{1/2}$. When $j = 1 - k$, this specification reduces to the Cobb–Douglas function. Douglas found the estimation of this relationship gave the result $P = 1.063R^{-0.146}L^{0.788}C^{0.358}$ (Douglas, 1934). There are, however, two disadvantages to this formulation. First, the estimate of the coefficient of R is not independent of the units of measurement and, secondly, the production function does not everywhere exhibit convex isoquants when $(1 - k - j) \neq 0$ (Samuelson, 1979).

dices less deep-seated than the proposition that statistical methods are applicable in economics' (de Marchi & Gilbert, 1989, pp. 1–2). Thus, it is not so strange to find Douglas having to justify his attempts at quantification, *per se*. As he wrote in 1934 (Douglas, 1934, p. 106):

> Any inductive study dealing with the problems of distribution or of value is almost invariably either brushed aside or attacked by the devotees of 'pure' theory on the ground that since statistical analysis is necessarily based on comparisons between time or space its units can never be identical with those timeless concepts which characterize 'pure' theory. ... When statistical series dealing with time sequences or even relative distributions in space are brought forward, the armchair theorists brush these aside on the ground that they may include either shiftings of the curves or different curves. These series are then dismissed as being merely historical or empirical.

Yet, on reflection, this reaction to his work is perhaps not so strange. Similar sentiments were also expressed by no less an authority than Keynes (1939) in his review of Tinbergen's (1939) econometric study, although the latter is now seen as one of the pioneering attempts at econometric model building. Keynes held that Tinbergen's work was merely 'a piece of historical curve fitting and description', so at least Douglas was in good company. (Morgan, 1990, has convincingly argued that not only was Keynes unaware of the new developments in econometrics that were used by Tinbergen, but had also clearly not read the book carefully!)

Douglas (1934, p. 106) felt it necessary to ask, even if there were some force to the criticism mentioned above,

> should we abandon all efforts at the inductive determination of economic theory and remain in the ivory tower of 'pure' theory? If this is what is done, we may as well abandon all hope of further developing the science of economics and content ourselves with merely the elaboration of hypothetical assumptions which will be of little aid in solving problems since we will not know the values. Or shall we try to make economics a progressive science?

Finally, in the third camp, and to be taken the most seriously, were those econometricians who disputed the interpretation of the results, primarily for statistical reasons. Most notable of these were the distinguished pioneering econometrician Ragnar Frisch and his former pupil, Horst Mendershausen. The latter considered the whole study so specious that 'all past work should be torn up and consigned to the wastepaper basket' (quoted by Douglas, 1976, p. 905). This bitter comment must have left a deep impression on Douglas, as he also had mentioned it in his earlier work when he also remarked: 'My friends thought the better part of valor was to ignore the whole subject and never mention it, but others were not so kind' (Douglas, 1967, p. 18).

Consequently, it is not surprising that Douglas was so discouraged and disheartened by such criticism that, at one stage, he had thought of giving up the effort entirely (Douglas, 1967, p. 18 and 1976, p. 905). Nevertheless, he persevered and undertook, in conjunction with a number of assistants and collaborators, more and more estimations of production functions, with increasing emphasis on the use of cross-industry data which were more plentiful, especially statistics for the capital stock, e.g. for the Australian states. (1931 was the last year until the postwar period for which the United States statistical authorities and Congress, on the advice of American economists and statisticians, collected United States capital stock statistics. It was felt that the data were too unreliable to be of any use.) Douglas (1934) incorporated the original 1928 study into his book *The Theory of Wages*, along with two further time-series studies (of Massachusetts and New South Wales) which seemed to confirm the early results. As has been noted, Douglas's 1947 presidential address was a lengthy survey and defence of his procedure for estimating production functions. But it was not until 1957, with the publication of Solow's classic paper on technical change and the aggregate production function, that Douglas's work received universal recognition. Indeed Samuelson, one of Douglas's pupils, considers that Douglas never fully appreciated the ultimate impact which the Cobb–Douglas production function (and the subsequent generalisations such as the Constant Elasticity of Substitution and translog production functions) had on the profession.[3]

4. Criticisms of the Initial Time-series Studies

In the 1928 paper, Cobb & Douglas reported neither the standard errors nor the R^2. The goodness of fit was determined by a comparison of the predicted and the observed value of $\ln P$. Consequently, I have re-estimated here the production function with Douglas's original data (except I used his revised 1934 employment series) and this gives the following result for American manufacturing over the period 1899–1922, where the figures in parentheses under the estimated coefficients are *t*-values[4]:

$$\underset{}{\ln(P/L)} = \underset{(0.50)}{0.010} + \underset{(5.92)}{0.251}\ \ln(C/L) \qquad \bar{R}^2 = 0.597 \quad \text{SER} = 0.0584 \quad \text{D.W.} = 1.572$$

$$F_{SC}(1, 21) = 0.178\ [0.68] \qquad F_{HET(1,22)} = 12.068\ [0.00]$$
$$F_{FF}(1, 21) = 0.111\ [0.74] \qquad \chi^2_N(1) = 1.145\ [0.56]$$

---

[3] Douglas's main academic interest was in labour economics, especially the history of wage theories, the effect of the elasticity of supply on wages, and occupational and geographical differences in wages. His magnum opus was *The Theory of Wages* (1934) and it was as part of this study that Douglas became interested in production theory and the marginal productivity theory. 'It began as an analysis of the relative elasticities of supply of both labor and capital and the effect of varying rates of change in these upon the distribution of product. But without an adequate theory of production, elasticities of supply did not themselves explain much' (Douglas, 1972, p. 46). A first draft of his book won a prestigious prize in 1927, but extensive revisions and elaborations meant that its publication was delayed until 1934. Douglas went on to have a distinguished role in public life, being elected to the US Senate in 1948. What is perhaps surprising is how little emphasis he placed, in later life, on his academic work. In his autobiography of over 600 pages, *In the Fullness of Time* (1972), barely two pages are devoted to the Cobb–Douglas production function and only one short chapter documents his academic career.

[4] F denotes the LM F statistic. F_{SC} is the Lagrangian multiplier test of serial correlation; F_{FF} is the Ramsey RESET test of functional form using the square of the fitted values; F_{HET} is a test of heteroscedascity based on a regression of squared residuals on squared fitted values and χ^2_N is the Bera–Jarque test of the normality of the residuals. I have used the LM F statistics in preference to the Lagrangian Multiplier statistics as generally they are preferable for small samples (Pesaran & Pesaran, 1991, p. 66). The figures in square brackets are the probability values.

Since there is evidence of heteroscedasticity, the t-values based on White's heteroscedastic-consistent standard errors were estimated. They are 0.53 and 4.25 for the intercept and slope, respectively. Of course, it should be emphasised that I have the benefit of hindsight since these diagnostic tests were not, of course, available to Cobb & Douglas. Nevertheless, they are not without interest, if only to confirm the validity of the original results.

$(1-k)$ and k take values of 0.251 and 0.749. Durand (1937) and Mendershausen (1938) both pointed out that a better procedure would be to leave the exponents of L and C unconstrained and estimate $P=bL^kC^j$. This was easily remedied by Douglas and the outcome reported in his presidential address (published in 1948). It was found that the sum of the coefficients was close to unity, although there was no test to determine whether or not the difference was statistically significant. Re-estimating the specification by OLS and not constraining the coefficients gives $k=0.766$ (t-value $=5.32$) and $j=0.245$ (3.82), where j is capital's output elasticity. The sum of the output elasticities is 1.011 with a standard error of 0.09. Consequently, the hypothesis that constant returns to scale prevails cannot be rejected at the 99% confidence level. These results, of course, merely serve to confirm those of Cobb & Douglas (1928) and Douglas (1934).

Criticisms of the results and specification of the model, from those who did not simply dismiss the whole endeavour out of hand, were not slow in forthcoming. Several issues were raised by J. M. Clark (1928) in a comment published at the same time as Cobb & Douglas's original paper. Another major criticism that deserves explicit mention was Horst Mendershausen's (1938) predominantly econometric critique.

It is possible to identify three main strands in these early criticisms, some of which were easily disposed of by Douglas, but others were not. First, it was held that the original results were plagued by the problem of serious multicollinearity and the undue influence of outliers. Consequently, the results were spurious, notwithstanding the fact that the estimates took plausible values, with the output elasticities being close to the relevant factor shares. Secondly, there was the complete absence of any allowance for technical change, a fact which shocked Schumpeter (Samuelson, 1979). Indeed, the remarkable goodness-of-fit suggested that, in fact, there was very little left for technical change to explain. Thirdly, and indeed related to the second problem, was the objection that the estimates were only capturing the historical trend growth rates of output, labour and capital which had no implications for the form of the production function. Each of these criticisms will be considered in turn.

5. The Problem of Multicollinearity and Outliers

Mendershausen (1938) argued that there was very nearly perfect multicollinearity between the three variables $\ln P$, $\ln L$ and $\ln C$ with partial correlation coefficients between pairs of the variables of over 0.8. He demonstrated this by performing two additional multiple regressions, minimising the sum of squares in the direction of $\ln L$ and $\ln C$. He then calculated the estimates of k and j from each of these two additional regressions and compared them with those obtained by the original Cobb–Douglas specification. 'By minimizing—in the three set—in the three different directions one gets very different intercoefficients, and k and [j] show correspondingly different values if taken from three different elementary regression equations.' I was unable to replicate Mendershausen's exact results, but his conclusions still apply.[5]

In spite of the good fits, the three regressions give very disparate results, not only for the coefficients of k and j but also for the implied degree of returns to scale. The regressions with either $\ln L$ or $\ln C$ as the regressands give most implausible values for the degree of returns to scale. The reason for these discrepancies is the presence of severe multicollinearity plus the excessive leverage from three observations, namely those for the years 1908, 1921 and 1922. 1920–21 saw a fall in output of just under 30% and 1921–22 a recovery of a similar magnitude. From a modern perspective, Mendershausen's statistical procedure is unusual, since if we assume that $\ln L$ and $\ln C$ are the independent variables, then the other specifications violate the assumptions of OLS. Moreover, at first glance, the multicollinearity in Cobb & Douglas's original specification (i.e. with $\ln P$ as the regressand) does not seem to be a problem; the partial correlation between $\ln C$ and $\ln L$ is less than the multiple R^2 and the standard errors of the coefficients are low. While these findings are suggestive rather than conclusive, $k+j$ can be estimated with precision. As has been be seen, this is not statistically significantly different from unity at the 95% (or, indeed, the 99%) confidence level. Thus, as $j=(1-k)$, it follows that $\ln P/L = \ln b + (1-k)\ln(C/L)$ is an appropriate transformation and this does not suffer from the problem of multicollinearity. Estimating this equation, as we have seen, gives values for k and $(1-k)$, or j, of 0.749 and 0.251.

Mendershausen's discussion opens an interesting window on the methodological debates in econometrics that were going on in the 1930s. One school of thought, which is implicit in Mendershausen's argument, assumed that statistical specifications were exact relationships, especially after the data had been pre-adjusted to take account of other influences. The reason that perfect fits were not obtained was the presence of measurement errors and the object was to choose the method of normalisation that would minimise the impact of these errors. The other view was the one that is generally accepted today, namely that the direction of causation is important and that the residuals capture the effect of the inevitable missing variables, as well as measurement errors and errors

[5] Mendershausen's (1938, p.147) estimates were

Direction of minimisation	$\hat{k}$	$\hat{j}$
$\ln P$	0.76	0.25
$\ln L$	− 1.06	− 1.14
$\ln C$	2.23	− 0.34

Table 1. Estimates of k, j and $(1-k)$ from a rolling regression, American manufacturing, 1899–1908 to 1913–22

Sample	k	(t-value)	j	(t-value)	$k+j$	$(1-k)$	(t-value)
1899/1908	1.476	(5.16)	−0.051	(−0.36)	1.425	0.180	(1.31)
1900/1909	1.450	(5.00)	−0.065	(−0.46)	1.385	0.161	(1.22)
1901/1910	1.381	(4.76)	−0.076	(−0.56)	1.305	0.078	(0.70)
1902/1911	1.314	(4.11)	−0.112	(−0.75)	1.202	−0.144	(−0.13)
1903/1912	1.422	(4.37)	−0.154	(−0.91)	1.268	−0.007	(−0.05)
1904/1913	1.652	(5.09)	−0.291	(−1.58)	1.361	−0.053	(−0.31)
1905/1914	1.588	(4.80)	−0.250	(−1.46)	1.338	−0.901	(−0.58)
1906/1915	1.453	(4.13)	0.004	(0.02)	1.457	0.168	(0.94)
1907/1916	1.270	(4.93)	0.181	(1.13)	1.451	0.432	(2.12)
1908/1917	0.978	(2.52)	0.268	(0.99)	1.246	0.535	(2,30)
1909/1918	0.701	(1.24)	0.290	(0.85)	0.991	0.280	(1.51)
1910/1919	1.000	(1.86)	0.054	(0.17)	1.054	0.109	(0.64)
1911/1920	0.940	(2.05)	0.080	(0.32)	1.020	0.096	(0.67)
1912/1921	0.839	(6.52)	0.062	(0.89)	0.901	0.062	(0.89)
1913/1922	0.612	(2.81)	0.216	(1.79)	0.828	0.237	(2.04)

Notes: k and j are the unconstrained output elasticities of labour and capital and $k+j$ is the degree of returns to scale. $(1-k)$ is the output elasticity of capital when constant returns to scale are imposed.
Data: Douglas (1934).

induced by averaging or aggregating the data (Morgan, 1990). (The former approach is now largely forgotten, although a vestige remains in the method of indirect least squares.)

If the conventional approach is adopted, and it is assumed that the appropriate specification is to regress $\ln P$ on $\ln L$ and $\ln C$, then Mendershausen's critique becomes obviated. Nevertheless, further analysis of Cobb & Douglas's specification, i.e. with $\ln P$ as the regressand, shows that the results are extremely unstable. I performed a rolling regression with a window of size 10 and the results are reported in Table 1. The size of the window is, to a certain extent, arbitrary and the value of 10 was chosen as providing an acceptable trade-off between the degrees of freedom and the number of regressions. The instability of the coefficients is readily apparent from the table and imposing the assumption of constant returns to scale (and hence avoiding the problem of multicollinearity) does not rescue the results.

The importance of the outliers 1921 and 1922 in producing anything like plausible significant values for k and j is readily apparent. It is only in the regression using the sample from 1913 to 1922 that the estimated coefficients are anywhere near their corresponding factor shares and are statistically significant. Moreover, the excessive leverage of the last two observations is confirmed by simply omitting them from the full sample. The results in this case are:

$$\ln P = -1.100 + 1.243 \ln L - 0.002 \ln C \qquad \bar{R}^2 = 0.975 \quad \text{SER} = 0.0418$$
$$(-2.04) \quad (5.63) \quad (-0.02) \qquad \text{D.W.} = 1.913$$

$(k + j) = 1.241$, t-ratio $= 10.75$

(Estimated by Newton–Raphson iterative method; errors AR(2).)

The sum of the coefficients k and j differ from unity at the 95% confidence level, and it should be noted that the statistical insignificance of the coefficient of $\ln C$ cannot be attributed to multicollinearity.

It also transpires that the other data sets (for Massachusetts and New South Wales) which Cobb & Douglas used putatively to support their conjectures also suffer from similar problems of instability. Thus, it is difficult not to agree with Mendershausen's (1938, p. 152) conclusion, although for different reasons, that 'it is now obvious that the empirical relation found between the coefficients k and j cannot be taken as a verification of the *pari passu* law. ... The nature of the production law cannot be ascertained at all from this set of variates.' To be fair, Douglas (1948) himself does allude, although understandably briefly, to the problems posed by omitting a few of the terminal years of his data, but does not report the regression results.

6. The Absence of Technical Progress

From a modern perspective, a glaring omission from Cobb & Douglas's specification is the assumption of a constant technology (b in the equation $P = bL^kC^j$ does not vary over time). Their critics were quick to seize on this point. Mendershausen (1938, p. 145) commented that 'these assumptions are manifestly in contradiction to all that economists know about the industrial development during this period'. Clark (1928, p. 463), likewise, expressed concern that 'one of the striking things in this study as presented is the fact that it seems to allow no room for the natural effect of advances in the "state of the arts". To one accustomed to crediting our increase in per capita income to triumphs of inventive genius, it must be a rude shock to see the whole increase calmly attributed to increased capital; while even on this basis the share of capital is only one-fourth of the whole, which seems too modest to leave room for any deductions. What, then, has become of our boasted progress?' Phelps Brown (1957, p. 550) also pointed out that an implication of Douglas's estimates was that the marginal product of capital in American manufacturing fell by one-half over the period 1899–1922, which seems implausible.[6] Douglas (1934, pp. 209–216) did discuss the problems posed by technical progress in some depth, but, as we shall see, did not come up with a satisfactory solution.

To a modern reader, with the benefit of the hindsight of Solow's (1957) classic paper (where technical progress explained over 85% of the growth of

[6] This was based on the observation that the rate of change of the rate of profit is equal to the growth of the capital–output ratio, as the marginal product of capital is given by $j(P/C)$. Hence, its rate of growth is $p - c$, where the lower case denotes a proportionate growth rate. From 1899 to 1922, the marginal product of capital was declining at an exponential rate of 2.88% per annum, which implies a fall of 48% over the period.

GNP per head in the US private non-farm sector from 1909 to 1949) and the experience of the plethora of neoclassical 'growth accounting' studies that have attempted with limited success to explain this 'residual', such criticisms of Cobb & Douglas's specification are not only pertinent, but also have a certain irony. Clark (1928, p. 464) argued that growth, *per se*, together with rising labour costs, induced technical progress. In a statement that has a modern ring to it and is reminiscent of the theory of induced innovations, he argued that 'it is typical of present-day methods of management to set a research department to work definitely on the problem created by changing cost conditions. The result is that any such changes will call forth a crop of new devices or cause others to be quickly developed which would otherwise have been very slow in getting past the experimental stage.' Clark, however, does not suggest how an allowance for technical progress could be included in the regression analysis. Rather, he comes close to advocating a growth accounting procedure: 'In inquiring whether these figures offer any evidence of the existence of "pure progress", the only available method seems to be to make all reasonable adjustments in the direction which would tend to indicate such progress, and then to see if the resulting trend of product is higher, relative to those of capital and labor, than can be plausibly explained by the actions of labor and capital alone'. In particular, he suggested that the growth of the labour input may have been overestimated in Cobb & Douglas's (1928) earliest study through the failure to adjust for the 16% fall in the average weekly hours worked. This would lead to an overstatement of the contribution of the growth of the labour input and, hence, to a reduction in the contribution of technical progress. However, this decline was exactly offset by an error in the data introduced by the failure to include the rapid growth of clerical workers in the initial labour series. The index of the first series for the volume of labour in 1922 was 161, with 1899 = 100 (Cobb & Douglas, 1928, p. 148, Table III). The value in their revised statistics, which made an allowance for both the change in hours worked and the growth in clerical workers was almost identical at 160.5 (Douglas, 1934, p. 126, Table 9). If we undertake a simple growth accounting exercise, it is possible to obtain an estimate of the growth of total factor productivity (or the residual), λ, from the expression

$$\lambda = p - 0.75l - 0.25c \tag{2}$$

where the lower case denotes the rate of growth per annum.

Over the period 1899–1922 the exponential growth rates of the variables are $p = 3.81\%$ per annum, $l = 2.06\%$ and $c = 6.35\%$, suggesting the residual is about 0.68% or less than 18% of total output growth. If the trend growth rates estimated by fitting a logarithmic time-trend are used, it is found that the residual falls to 0.03% per annum.[7]

[7] The trend rates of growth obtained by OLS by estimating, for example, $\ln P = c + bt + u$, where u is the error term are:
$p = 3.59\%$ p.a. (t-ratio = 13.37; $\bar{r}^2 = 0.885$),
$l = 2.59\%$ p.a. (t-ratio = 9.35; $\bar{r}^2 = 0.790$),
$c = 6.47\%$ p.a. (t-ratio = 58.81; $\bar{r}^2 = 0.994$).

Douglas (1934, p. 209) certainly conceded that the omission of any allowance for technical change was 'disconcerting'. As he pointed out, the growth of productivity in his formula was almost entirely explained by the growth of the tangible factors of production. 'But this is not really progress in any dynamic sense. It is a mere accumulation of greater quantities of the factors rather than a greater effectiveness of each unit.' How can the approach allow for technical change? Douglas made some not very convincing suggestions. First, he argued that a greater role for technical change may be found in the sub-periods and especially during the boom years of 1921–26 (although much of the rapid increase in productivity was likely to have been the result of the economy coming out of a severe recession and, hence, the effect of a greater utilisation of the factor inputs). Secondly, the reduction in the average weekly hours of manual workers could be attributed to technical progress, as Clark suggested. The problem with this line of argument is that presumably the increase in clerical workers would have to be due to an offsetting technical regress. Thirdly, he had sympathy with Clark's suggestion, anticipating the new growth theory by several decades, that 'the product apparently attributable to capital alone is also in a sense attributable as well to progress.' One possibility, Douglas advanced, was that an improvement in the quality of the capital stock is likely to have been matched by an increase in the quality of the labour force. 'If this be the case, then the improvement in the quality of the workers has served to balance the qualitative improvement of the capital instruments with the result that while "progress" would have affected the joint product through each of the factors it will not be reflected in the formula.'

Douglas's interpretation seems to be something along the following lines. The production function may be written in intensive form as

$$P/L = b(C/L)^{(1-k)} \tag{3}$$

Define the quality-adjusted capital and skill-adjusted labour as $C'_t = C_t \exp(\lambda_C t)$ and $L'_t = L_t \exp(\lambda_L t)$, where λ_C and λ_L are the rates of improvement in the efficiency of capital and labour. Let us assume that C'_t/L'_t is approximately equal to C_t/L_t, so that $\lambda_C \simeq \lambda_L$. Consequently, if equation (3) is expressed as (dropping the time subscripts once again)

$$P/L = b(C'/L')^{(1-k)} \tag{4}$$

the contribution of technical progress is being captured by the *observed* capital–labour ratio. The difficulty with this line of reasoning is that it is only correct if the level of productivity is also measured in terms of efficiency labour units, i.e. P/L', which is not the case. Improvements in the quality of labour and capital are not offsetting but rather are reinforcing since Equation (4) should be written as

$$P/L = be^{\lambda t}(C/L)^{(1-k)} \tag{5}$$

where $\lambda = k\lambda_L + (1-k)\lambda_C$.

Moreover, even if the whole of the residual is attributed to technical progress, as we have seen above, it has quantitatively a small role to play.

Nevertheless, Douglas made no explicit justification for a constant, or time invariant, production function except by pointing to the good fit such a function gave to the data. Mendershausen (1938, p. 145) dismissed this as a '*petitio principii*, since only if this hypothesis [of a constant production function] is justified can the function claim to be taken as production function.'

One solution to the problem of the absence of technical progress is the inclusion of a time-trend in the regression to capture the exogenous shift of the production function. As Brown (1966) notes, one of the first economists to undertake this was Tinbergen (1942). Tinbergen, in fact, constrained the coefficients, a priori, to be 0.75 and 0.25 on the basis of Douglas's findings, but as Brown points out, this is an unnecessary restriction. 'In retrospect, Professor Tinbergen's introduction of a trend term appears so obvious that one wonders why it was not done before. The obviousness of the innovation should not detract from its importance: it provides an operational means of quantifying neutral changes in the production process' (Brown, 1966, p. 112). In fact, a hint as to how to proceed was given to Douglas by Copeland who, in correspondence, informed him that he (Copeland) had assumed that the whole of the growth of productivity was the result of technical progress and had consequently attempted to explain the growth of P/L solely in terms of a log-linear time-trend. He found that the fit was as good as that achieved by Cobb & Douglas. It is a short step to combine the two methods and to incorporate a time-trend as in Equation (5); but for some reason Douglas never took it.

Including a time-trend to capture exogenous technical change also has the advantage that it detrends the data (the Frisch–Waugh theorem). This brings us to the next criticism, namely that the estimates of the regression were merely reflecting historical trend growth rates and consequently had no implications for the form of the production function.

7. Are the Estimates Merely Picking Up Historical Growth Rates?

As Samuelson (1979) points out, the fact that the relative trend rates of growth of P, L and C took their declared values may also have suggested the Cobb–Douglas relationship to Cobb. However, Mendershausen (1938) (and, following him, Phelps Brown, 1957) saw this as a major criticism of the fitting of the relationship.

The data for P, L, and C are all strongly trended, yet Cobb & Douglas, as we have noted, did not include a time-trend in their specification of the production function. Let us assume that the data for the three variables can be represented by the expressions $P_t = P_0 e^{pt}$, $L_t = L_0 e^{lt}$ and $C_t = C_0 e^{ct}$, where the subscript o denotes that the variable refers to the base year with an index equal to 100. The difference in the growth rates of P and L, together with the difference between the growth of C and L, is given by:

$$\ln P_t - \ln L_t = (p - l)T \tag{6}$$

and

$$\ln C_t - \ln L_t = (c - l)T \tag{7}$$

where T is the length of time from the base year to time t.

It follows that we would be almost certain to get a very good fit if the following relationship were to be estimated:

$$\ln(P_t/L_t) = \ln b + (1 - k)\ln(C_t/L_t) \tag{8}$$

The reason is that, as the data are expressed as indexes, $\ln b$ will approximate to 0. Furthermore, the values of the coefficients k and $(1 - k)$ will be determined by the trend rates of growth of the three variables and more explicitly, from equations (6) and (7), they will be given by:

$$k = \frac{(c - p)}{(c - l)} \text{ and } (1 - k) = \frac{(p - l)}{(c - l)} \tag{9}$$

The trend growth rates reported in footnote 7 confirm that all the indexes do, in fact, exhibit very pronounced trends. Thus, using these results together with Equation (9), I obtained the results that $k = 0.74$ and $(1 - k) = 0.26$, which are almost identical to the OLS estimates.

How does this affect the interpretation of the results? Phelps Brown (1957, p. 550) had no doubts. He argued that these rates of growth 'are historical. The differences between them will not directly have the significance of exponents in a production function.' Douglas was well aware of this problem, which had obviously been pointed out by others. 'These critics have alleged that equally good results could be obtained by comparing the relative movement of hogs in Iowa, hens in Wisconsin, and product in manufacturing. Such critics have, to be sure, not submitted the data to justify this contention, but their implication has been that the correlation between [the predicted value of P] and P was nonsensical' (Douglas, 1934, p. 141).

One must have a certain sympathy with Douglas when he points out that he, at least, had a theoretical justification for the relationship between P, L and C which is absent from the above mentioned 'attempted *reductio ad absurdum*' (Douglas, 1934). If there is an underlying production function given by Equation (8), then, of course, the growth rates of P, L and C are not independent since Equation (8) implies that $p = kl + (1 - k)c$.[8] The growth rates do have significance for the exponents of the production function. It is possible, of course, that p, l and c could have taken the values they do by chance to give the expected value of k, but the replication of this result in other studies (Massachusetts and New South Wales) led Douglas (1934) to argue that this was unlikely. Douglas (1948) also pointed to the many similar results from cross-industry

[8] Phelps Brown (1957, p. 551) also argues that 'we can know from the start that some time-series will not yield acceptable values of k, but will imply that the share of labor is negative, or greater than the whole product. For we have seen that k will lie between 0 and 1 only if the capital coefficient (stock of capital per unit of annual product) is progressively changing: but the evidence indicates that there have been periods, in the United States and the United Kingdom, in which this has not been so'. However, if the data is capturing an underlying production function of the form $p = kl + (1 - k)c$, then if $p = c$, it follows that $p = l$ and k is undefined. There is no a priori reason why k should not yield acceptable results. Whether or not it does is an empirical question.

regressions. (However, these results are not devoid of problems, which are discussed below.)

8. Detrending the Data

In the light of Mendershausen's and Phelps Brown's criticism, it is clearly necessary to detrend the data. In *The Theory of Wages*, Douglas reported that two of his students had estimated the Cobb–Douglas function using detrended data. They somewhat unconventionally fitted a log-log time-trend to C (i.e. $\ln C = a_1 + b_1 \ln t$) and linear trends to L and P (e.g. $L = a_2 + b_2 t$), rather than using the more usual log-linear relationships. They found that k took a value of 0.84 and b, the constant, was to all intents unity when the detrended data were used. Douglas considered that the results were more or less in accord with the original results. 'The equation of trend ratios can be treated as $P' = L^{0.84}K^{0.16}$ [where P' is the predicted value of P]. The value of k is only 9 points or 12 per cent more than the value of .75 as computed from the original data' (Douglas, 1934, p. 144). This is misleading and is not the whole story. I re-estimated the production function with the coefficients unconstrained. The estimate of k was 0.864 (with a t-ratio of 5.72), but j took a value of -0.464 (-1.31). The sum of k and j is 0.400 with a large standard error of 0.309. This would tend to refute Douglas's contention, not lend it support.

As we noted above, one procedure is to include an explicit time-trend, which will both detrend the data and provide an estimate of the rate of technical change. When a time-trend is included, the following results are obtained for US manufacturing for 1899–1922 using Douglas's (1934) data and OLS:

$$\underset{(1.88)}{\ln P = 2.728} + \underset{(1.98)}{0.043t} + \underset{(6.01)}{0.848 \ln L} - \underset{(-1.26)}{0.449 \ln C} \qquad \bar{R}^2 = 0.963 \quad \text{SER} = 0.056 \quad \text{D.W.} = 1.63$$

$$F_{SC}(1, 21) = 0.045\ [0.84] \qquad F_{HET}(1, 22) = 0.624\ [0.44]$$
$$F_{FF}(1, 21) = 0.054\ [0.82] \qquad \chi^2_N(1) = 1.885\ [0.39]$$

($k + j = 0.399$ with a t-value of 1.25)

The insignificance of the capital stock as an explanatory variable is somewhat disconcerting for those who would interpret these results as reflecting a production function. These results are also, of course, very similar to those obtained using Douglas's students' detrended data as reported above. (The alternative specification, namely constraining the coefficients to sum to unity, fares no better. The coefficient of $\ln(C/L)$ is not significantly different from zero.)

Recent developments in econometrics mean that tests for the stationarity and, if necessary, co-integration of the data would now be routinely undertaken. Using the Dickey–Fuller and the Phillips–Perron tests, it is found that $\ln L$ is $I(1)$, while both $\ln P$ and $\ln C$ are $I(0)$. This rules out the possibility of the variables being co-integrated. This is because the regressors must both be of the same order of integration, and this order must be equal to, or greater than, that of the regressand. Overall, this makes the interpretation of the regression results even more problematical. One can adopt a pragmatic approach and, given the smallness of the sample, assume that the results are the best we can do. (If all the data were $I(1)$ and not co-integrated, the appropriate procedure would be to use first differences. This does not produce an improvement in the results from Douglas's point of view.) On the other hand, by the Sargan–Hendry approach, the non-stationarity means that the production function estimates are meaningless.

Ex post, it may not be too difficult to explain away the poor results. As Douglas points out, neither the labour nor the capital inputs have been adjusted for changes in the rate of utilisation. This is likely to be an especially serious problem with the use of the capital stock estimates, as the series represents an estimate of the maximum potential capital stock, and values for a number of years were calculated by interpolation. Hence, it shows very little variation about trend. Ideally, the series should be adjusted for changes in capacity utilisation so that it more accurately proxies the flow of actual capital services. However, the estimate of the regression coefficient is likely to be sensitive to the precise way in which this is done. The employment index is likely to be less misleading as a proxy for the flow of labour services over the cycle, but even here there may be errors introduced by labour-hoarding. This may also explain the problems of non-stationarity.

Moreover, the series for the capital stock is very highly correlated with time and so multicollinearity is a further problem. Consequently, *pace* Douglas, on these grounds alone there are severe problems with the time-series estimation of the production function and the interpretation of the results is extremely problematical. As Phelps Brown (1957) noted, this problem plagued other early studies as, when a time-trend had been included in the studies of, for example, Wall (1948) and Leser (1954), 'the results have not been acceptable'.

9. The Cross Sectional Studies

Douglas's reaction to all this hostile criticism was, in collaboration with a number of colleagues, to undertake even more statistical analysis using American and Australian data. There was obviously a limit to the amount of replication that could be achieved using time-series data given the limited data then available. Moreover, the anomalies were beginning to mount up, as we have noted. 'One persistent area of difficulty in these last months has been the Massachusetts time-series. We tried to improve on Professor Cobb's series of capital and product with the result that the more we refined the basic series, the more nonsensical the results became. ... Secondly, it is disconcerting to observe if we shorten our time periods by dropping off a number of terminal years, we appreciably alter our results. We observed this fact earlier, as did Professor Williams in New Zealand, but this paradox has been most manifest when we omit the war years from 1916 on in our United States time series' (Douglas, 1948, p. 21). Consequently, with the help especially of Grace Gunn, Douglas turned his attention to estimating production functions using inter-industry data, the results of which, he considered, supported his earlier time series results.

Douglas (1948, Table I, p. 12), for example, reports cross-industry regression results for 6 years between 1889 and 1919 using US manufacturing data. The smallest number of observations was 258 (for 1909) and the average estimate of k was 0.63, while that for j was 0.34.

The subsequent multitude of results certainly removed the objections that the first estimates were either merely the consequence of coincidental historical growth rates or plagued by multicollinearity, or both. As Douglas (1948, pp. 40–41) argued 'it is hard to believe that these results can be purely accidental, as some critics have maintained. ... The deviations of the actual or observed values from those which we would theoretically expect to prevail under the formula are not large and indeed are slightly less than we would expect under the random distribution of errors and of measurement. It is submitted that the total number of observations, namely over 3500, is sufficiently large so that if the results had been purely accidental, this degree of agreement would not have occurred.'

What may be regarded as the seminal paper using cross-industry results was written by Douglas in collaboration with Grace Gunn and published in 1941. Here, they threw down the challenge—'we invite critics to study these relationships [of the Cobb–Douglas production function] and we shall be glad to hear of any other interpretations'. It was some 16 years later before Henry Phelps Brown picked up the gauntlet and, using their data, launched his devastating critique.

Gunn & Douglas's (1941) study used data for 85 or 87 (depending upon the year) Australian manufacturing industries for 1912, 1922/23, 1926/27 and 1936/37. The major aim of Gunn & Douglas was once again explicitly to test the marginal productivity theory of factor pricing by comparing the estimated output elasticities with the relevant factor shares. They commenced their paper by outlining the simple profit maximising model where first the firm is perfectly competitive and is a price-taker in all markets and, secondly, where the firm can influence its product price but not factor prices. Using the Cobb–Douglas production function, they showed that in the first case labour's share will be equal to k, while in the second case it will equal $k(1 - 1/\eta)$ where η is the price elasticity of demand for output. Estimates for k and j were initially obtained by estimating the production function using, first, the industry averages per plant (i.e. the total industry values of P, C and L divided by the number of firms in that industry) and, secondly, the total industry values for the variables. The results showed that the estimates of k and j were well-determined with t-values of about 10 in both cases. The estimate of k varied from 0.61 to 0.50 using the averages per plant and 0.59 to 0.49 using industry totals. Gunn & Douglas concluded that 'on the whole, the evidence in this study, as in others we have made, seems to indicate an approximately linear homogeneous function for production. There is, however, a slight indication of increasing returns in 1922–23 and of decreasing returns in 1926–27' (Gunn & Douglas, 1941, p. 118). However, they did not test whether or not these values were significantly different from unity. Gunn & Douglas then proceeded to compare the estimated elasticities of output with respect to labour with labour's factor share, namely W/P, where W is labour's total compensation. Averaging over the four years, they found that labour's share was 0.54 which compared with an estimate of k of 0.55 (using averages per plant) and 0.53 (using industry totals). The results were seen as a strong vindication for the argument that factors were paid their marginal products under conditions approximating to perfect competition. 'The close agreement between k and W/P, the theoretical and actual shares received by labor, is truly astonishing. ... These results might be taken as an indication that for manufacturing industries in these particular years the elasticity of demand for their products was very high or, stated in terms of price, that the price flexibility was very close to zero. This would mean that the industries were operating under conditions which were approximately those of perfect competition in the markets for the factors' (Gunn & Douglas, 1941, p. 127).

But once again, objections were not slow in forthcoming. The consensus today is that the estimation of inter-industry production functions makes little theoretical sense. First, it is highly unlikely that each industry is subject to the same production function and, if this is the case, the regression analyses are likely to suffer substantial mis-specification errors. Or, to put this another way, there is only one observation for each potential production function and the data are drawn from as many different production functions as there were observations. Of course, nothing about a production function can be inferred from only one observation. (See Bronfenbrenner, 1944, who interpreted the aggregate production function as the envelope of the differing micro-production functions of the various firms.) Secondly, even if there were a common production function, the fact that, in equilibrium, firms would face the same factor prices means that all the observations would be on the same point on the production function. Thus, if we were to estimate $P/L = b(C/L)^{1-k}$, we should find that there was no systematic variation in the capital–labour ratio (since this is a function of relative factor prices, which are constant). The only variation in the data would come from the disturbance term. Hence, the results could not be interpreted with confidence as reflecting the parameters of the production function. This point was illustrated by Phelps Brown with the help of a three-dimensional diagram of the scatter of observations of output, capital and labour (similar to a photograph of a three-dimensional physical model which Gunn & Douglas published in their paper). He argued that the scatter showed the tendency of P, C and L to change in much the same proportion between industries 'so that industries that are small, judged by their labour-forces, use much the same amount of capital per head, as the big ones do' and so C/L is likely to be roughly constant. However, to the extent that the coefficients are well-determined, then Gunn & Douglas's contention that 'there is, we believe, sufficient scatter to determine a plane' could be seen as well-founded. Phelps Brown conceded that the scatter could be sufficient for k not to be indeterminate, but his assessment of the pattern of the scatter is, at least on first reading, somewhat obscure.

His argument proceeded as follows. Using Gunn & Douglas's data, he calculated for each industry the level of output that would have occurred if the industry had the average output–capital ratio. He terms this the 'expected' output. This value is then subtracted from the 'actual' output and the difference expressed as a percentage of the 'expected' output. In other words, the percentage excess of actual over expected output is defined as

$$\left[P_i - \left\{\frac{\bar{P}}{C}\right\}C_i\right] \Big/ \left\{\frac{\bar{P}}{C}\right\}C_i \tag{10}$$

multiplied by 100, where $\left\{\frac{\bar{P}}{C}\right\}$ is the average output-capital ratio and $\left\{\frac{\bar{P}}{C}\right\}C_i$ is the 'expected' level of output for firm i.

He then adopted the same procedure to calculate the percentage excess of the 'actual' over the 'expected' volume of labour.

The former variable (for output) was plotted against the latter (for employment) (Phelps Brown, 1957, Fig. IV) and Phelps Brown concluded that 'here clearly there is much evidence for a systematic relation: the points lie fairly closely about a line of regression, whose slope tells us that, between one industry and another, a 1 per cent increase in the intake of labor per unit of capital generally went with an increase of 0.53 of 1 per cent in the corresponding net value product. ... The slope just calculated *may* be said to correspond with the Cobb–Douglas coefficient k, in that both measure the ratio of a proportionate difference in product, between one situation and another, to the associated proportionate difference in labor intake per unit of capital; and the value of 0.53 may be compared with that of 0.52 obtained by Gunn & Douglas for k when they fitted the same census data. Whether the two coefficients have the same economic meaning, however, is another question' (emphasis added).

The difficulty with the Phelps Brown argument, as it stands, is that all he has, in fact, accomplished (without seeming to realise it) is the replication Gunn & Douglas's OLS regression procedure. Hence, there is little wonder that the two estimates are virtually identical. This may be shown as follows. Expressing the data in logarithmic form for industry i in terms of deviations from the mean, we obtain for the output-capital ratio:

$$\ln(P_i/C_i) - \ln\overline{(P/C)} \tag{11}$$

But for any variable Y, the approximation $\ln(Y_i/\bar{Y}) \simeq (Y_i - \bar{Y})/\bar{Y}$ holds, where $\bar{Y}$ is the mean value.[9] Consequently, expression (11) may be expressed as

$$\left[\frac{P_i}{C_i} - \left(\frac{\bar{P}}{C}\right)\right] \Big/ \left(\frac{\bar{P}}{C}\right) \tag{12}$$

which, after rearrangement, is identical to expression (10). Thus Phelps Brown's procedure is analogous to that of Gunn & Douglas when they regress $\ln(P_i/C_i)$ on $\ln(L_i/C_i)$. One should not be surprised that the slope coefficients are virtually identical.

Phelps Brown is correct when he states that this relationship is merely capturing differences between industries and does 'not necessarily throw any

[9] $\ln Y$ may be approximated by a Taylor series expansion around $\bar{Y}$, ignoring terms to the power of 2 and greater, as $\ln Y = \ln \bar{Y} + (1/\bar{Y})(Y - \bar{Y})$. Hence, $\ln(Y/\bar{Y}) = (Y - \bar{Y})/\bar{Y}$.

light on what will happen when we vary the proportion of labor to capital within one industry.' Moreover, by virtue of the above argument, this is equally true of Gunn & Douglas's approach. But Phelps Brown has not added anything new by his statistical argument.

This leads to Phelps Brown's next criticism. All that is being captured by Douglas's regressions is the accounting identity $P_i \equiv w_iL_i + r_iC_i$ and he was not estimating a technological relationship, in the form of a production function. at all. This point was first made by Marshak & Andrews (1944) who concluded that, on the whole, Douglas and his co-workers had been fitting a hybrid of a cost and production function and had confused it with the true production function. The implication is that there is an identification problem, similar to that exemplified by the familiar textbook example of supply and demand curves. An implication of this is that all is not lost, however, as it should be theoretically possible to find exogenous variables to identify the production function.

But Phelps Brown suggested that the problem is, in fact, insoluble. There is no identification problem; the estimates are unambiguously of the cost identity. The Phelps Brown critique was subsequently formalised by Simon & Levy (1963) and Cramer (1969), but the full import of the criticism went largely unnoticed. For example, Intriligator (1978, p. 270), while discussing Cramer's argument, only notes that it will lead to a bias in the estimates towards constant returns to scale and that factor shares will be approximately equal to the output elasticities. It is not mentioned that the problem removes entirely the possibility of interpreting the result of estimating a production function as a test of a technological relationship. To be fair, though, Cramer himself does not push his argument to its logical conclusion.

One reason for the relative neglect of this argument may be partly due to the fact that it was originally applied to inter-firm production functions and, as we have noted, these were already suspect on theoretical grounds.[10] But perhaps more importantly, it was (erroneously) not seen to be applicable to time-series estimations. We return to this important point below. In view of the significance of the Phelps Brown criticism, it is worth quoting him on this point:

> For on this assumption [that we can write $P \equiv wL + rC$ for any industry], the net products to which the Cobb–Douglas is fitted would be made 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 k by fitting the Cobb–Douglas function we are bound to arrive at the same value as when we reckon up total earnings and compare them with the total net product. In k we have a measure of the percentage change in net product that goes with a 1 per cent change in the intake of labor, when the intake of capital is 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 $P_i = wL_i + rC_i$, the difference between them will always approximate to the

[10] In the postwar period, a number of studies were undertaken for individual industries using US state data (for example, Hildebrand & Liu, 1965, and Moroney, 1972). This specification using cross-state data avoids the objection to the cross-industry studies that it was unlikely that all industries could be represented by the same production function. Nevertheless, the good statistical fits that were usually obtained using the regional data must be attributable to the accounting identity.

> compensation at the wage rate w of the difference in labor intake. The Cobb–Douglas k and the share of earnings, will be only two sides of the same penny. (Phelps Brown, 1957, p. 557)

The message of the cited passage is, in fact, simple; but no less devastating for that. Phelps Brown's argument implicitly starts with the definition of the output elasticity with respect to labour as $k = (\partial P/\partial L)/(L/P)$. Consequently, k is equal to the percentage change in output when the labour input increases by 1%, *ceteris paribus*. The accounting identity is given by $P_i \equiv wL_i + rC_i$, from which it may be seen that the difference in output between two firms that differ only in their labour input is $\Delta P = w\Delta L$, and

$$\frac{\Delta P}{\Delta L} \cdot \frac{L}{P} = \frac{wL}{P} = k \tag{13}$$

Consequently, k must, *by definition*, equal labour's share in output.

A more formal demonstration of a similar proposition stems from the work of Simon & Levy (1963) and Cramer (1969). Since Douglas estimated the production functions using data expressed as indexes, let us express the accounting identity in an index form (this does not affect the generality of the argument):

$$\frac{P_i}{P_o} = \left(\frac{w_i L_o}{P_o}\right)\frac{L_i}{L_o} + \left(\frac{r_i C_o}{P_o}\right)\frac{C_i}{C_o} \tag{14}$$

or, if $w_i \simeq w_0$ and $r_i \simeq r_0$,

$$\tilde{P}_i = a\tilde{L}_i + (1 - a)\tilde{C}_i \tag{15}$$

where $a = w_0 L_0/P_0$ and $(1 - a) = r_0 C_0/P_0$. The subscript o denotes the base, or reference, industry and i denotes the ith industry. For convenience, the tilde explicitly denotes an index, where $\tilde{P}_0$ etc, equals unity.

As was mentioned above, $\ln \tilde{Y}_i \simeq \ln \tilde{Y}_0 + (\tilde{Y}_i - \tilde{Y}_0)/\tilde{Y}_0$. Consequently, the expression $\tilde{P}_i = \tilde{L}_i^k \tilde{C}_i^j$ may be expressed as:

$$\ln \tilde{P}_i = k \ln \tilde{L}_i + j \ln \tilde{C}_i \tag{16}$$

and as $\ln \tilde{P}_i = \tilde{P}_i - 1$,

$$\tilde{P}_i = k\tilde{L}_i + j\tilde{C}_i + (1 - k - j) \tag{17}$$

Comparing Equations (15) and (17), it is immediately apparent that the multiplicative Cobb–Douglas function provides a good approximation to the linear accounting identity and $k = a$ and $j = (1 - a)$. This assumes that there is not too much divergence of w_i and r_i from the base values. Consequently, j and k must sum to unity. This argument also holds even if the actual values of the variables are used rather than indexes.[11] Thus, if shares are constant, a

[11] This may be seen either by applying the above method using a Taylor series expansion (Simon & Levy, 1963) or, as here, by extending the method of Phelps Brown. Assuming a continuum of firms, totally differentiating the cost identity gives

$$\mathrm{d}P = (\mathrm{d}w)L + (\mathrm{d}r)C + w(\mathrm{d}L) + r(\mathrm{d}C) \tag{a}$$

and it follows that

Cobb–Douglas function will always give a good fit to the data even though there may be no well-defined underlying aggregate production function.

Simon (1979) provides a demonstration as to just how good an approximation the multiplicative function will give to the linear cost identity. There are two functional forms, namely $P_E = bL^kC^j$ and $P = wL + rC$, where P_E is the estimate of the accounting identity given by the power approximation. Simon calculates the ratio of P_E to P for a wide variation of capital–labour ratios, where the largest value is 25 times the smallest value. This is far larger than any ratios encountered empirically. Simon (1979, p. 466) finds that the ratio does not greatly diverge from unity and concludes that 'since in the data actually observed, most of the sample points lie relatively close to the mean value of [L/C], we can expect *average* estimating errors of less than 5 per cent'.

This point may be demonstrated even more dramatically. I constructed an artificial data set of 25 observations with L/C increasing from 0.20 to 5 (the range chosen by Simon) in increments of 0.20. The index of total costs (value added) $\tilde{TC}$ per unit of capital was constructed from the identity (dropping the i subscript for convenience):

$$\tilde{TC}/\tilde{C} \equiv \tilde{P}/\tilde{C} \equiv a(\tilde{L}/\tilde{C}) + (1 - a) \qquad (18)$$

The share of wages in value added was taken to be 0.75. The share was not assumed constant but was constructed so that it varied normally with a standard error of 0.02, which is plausible when compared with actual values of labour's share. (The value of a ranges from 0.80 to 0.72.) These data were used to estimate both the linear identity and the Cobb–Douglas specification. The results of estimating the relationships between indexes of output per worker and the capital–labour ratio are as follows:

(i) $\tilde{P}/\tilde{L} = \underset{(200.51)}{0.757} + \underset{(80.42)}{0.240}(\tilde{C}/\tilde{L}) \qquad \bar{R}^2 = 0.996 \qquad \text{SER} = 0.015$

(ii) $\ln(\tilde{P}/\tilde{L}) = \underset{(4.35)}{0.079} + \underset{(13.87)}{0.231}\ln(\tilde{C}/\tilde{L}) \qquad \bar{R}^2 = 0.888 \qquad \text{SER} = 0.068$

The identity, not surprisingly, gives an almost perfect fit. But what is interesting is the very good fit that the Cobb–Douglas function gives notwith-

---

$$\frac{dP}{P} = a\frac{dw}{w} + (1-a)\frac{dr}{r} + a\frac{dL}{L} + (1-a)\frac{dC}{C}$$

where $a = wL/P$ and $(1 - a) = rC/P$. If shares are constant, Equation (a) may be integrated to give

$$P = Aw^a r^{(1-a)} L^a C^{(1-a)} \qquad (c)$$

where A is the constant of integration and, if w and r show little variation between industries, Equation (c) may be written as

$$P = bL^a C^{(1-a)} \qquad (d)$$

It will be seen that so long as $a \ln w + (1 - a) \ln r$ does not vary greatly or is orthogonal to $a \ln L$ and $(1 - a) \ln C$, the accounting identity will give a very good approximation to the Cobb–Douglas, so long as factor shares are roughly constant.

standing the substantial variation in the capital–labour ratio. If the underlying data which generated this result were overlooked, the statistical results of Equation (ii) could be interpreted as not refuting the hypothesis that the manufacturing sector (or the whole economy) could be represented by an aggregate Cobb–Douglas production function. It would also be possible to infer, as did Douglas, that the equality of factor shares and the output elasticities support the marginal productivity theory. But, of course, there is no justification for either of these conclusions.

Finally, mention should be made of Houthakker's (1955) demonstration that if all individual production techniques exhibit fixed coefficients, but firms can combine two or more of these processes and the firm sizes followed a Pareto distribution, the aggregate production function would be a Cobb–Douglas. Like the other critiques, this also has been almost totally ignored (for an exception, see Blaug, 1974).

10. Anwar Shaikh's *Coup de Grâce*?

The outcome of the critique discussed in the previous section is that not only is it theoretically meaningless to estimate an unique production function using cross-industry data, but it would also be most surprising if the estimated coefficients did not closely correspond to the relative factor shares. This, of course, has nothing necessarily to do with the validity of the neoclassical marginal productivity theory. What are the implications of these criticisms for the use of time-series data for estimating a Cobb–Douglas function? Bronfenbrenner (1971), who had worked closely with Douglas in the early days, argued that there are none. This is because the specification of the production function should include a time-trend to capture technical progress, and this is absent from the cost identity. Leaving aside the problem that, even if this argument were valid, the shift of the production function would identify the cost equation and not the production function, Shaikh (1974) has shown that the time–series estimations are not immune from Phelps Brown's criticism.

The cost identity is given by (including time scripts for clarity):

$$P_t = w_t L_t + r_t C_t \tag{19}$$

Taking natural logarithms and differentiating equation (19) with respect to time gives

$$P_t = \{a_t \dot{w}_t + (1 - a_t)\dot{r}\} + a_t l_t + (1 - a_t)c_t \tag{20}$$

where the lower case denotes an exponential growth rate, except in the case of w and r where this is signified by $\dot{w}_t$ and $\dot{r}_t$. If shares are constant, so that $a_t = a$, equation (20) may be integrated to give

$$P = b \exp\left(\int (a\dot{w}_t + (1 - a)\dot{r}_t)\, \mathrm{dt}\right) L^a C^{(1-a)} \tag{21}$$

But empirically the weighted growth of wages and the rate of profit can be approximated by a time-trend (McCombie & Dixon, 1991) so that $a\dot{w} + (1 - a)\dot{r} \simeq \lambda$ (a constant). (In practice for the advanced countries, the rate of profit shows little secular variation and the trend is primarily capturing the growth of wages.) Consequently, Equation (21) can be written as

$$P = be^{\lambda t}L^{a}C^{(1-a)} \tag{22}$$

which is formally identical to the Cobb–Douglas 'production function'. As Shaikh put it, the Cobb–Douglas relationship is due not to the 'laws of production' but rather 'the laws of algebra'.[12]

Simon (1979) has a similar, if less general, argument. He assumes that the rate of profit r does not change over time and C is a constant proportion of P, i.e. $C = sP$. (This implies that the factor shares will be constant.) Commencing with the accounting identity, $P_t \equiv w_tL_t + r_tC_t$, using the aforementioned assumptions and assuming that wages grow at the exponential rate $\dot{w}$, it may be shown that

$$P_t = (P_0/L_0)e^{(\dot{w}t)}L_t \tag{23}$$

Turning to the Cobb–Douglas, namely $P_t = be^{\lambda t}L_t^{k}C_t^{j}$, and using the assumption $C_t = sP_t$, it is a straightforward matter to show that

$$P_t = [P_0/L_0^{k/(1-j)}]e^{\lambda t/(1-j)}L_t^{k/(1-j)} \tag{24}$$

Comparing Equations (23) and (24), it can be seen that $k = (1 - j)$ and λ, the rate of technical progress, will be definitionally equal to the growth of real wages multiplied by its factor share. The fitted Cobb–Douglas will also be homogeneous of degree one.

Moreover, Simon (1979) and McCombie & Dixon (1991) show, not surprisingly, that generalisations of the Cobb–Douglas, such as the CES and the translog production function, are not immune from this criticism.

Why then did the early time-series analysis produce such poor results? The answer is that even though the weighted sum of the wage rate and the rate of profit demonstrate a strong trend, the variation in $\ln L$ and $\ln C$ is relatively so small (with often the latter especially showing very little fluctuation) that the estimates of the coefficients are not well determined. One solution, following Shaikh (1980), is to abandon the assumption that technical progress is a smooth function of time. If we were to fit non-linear (cyclical) time-trends, then the one that proxied the exact path of $w_t^{a}r_t^{(1-a)}$, or, in a dynamic context $a\dot{w}_t + (1 - a)\dot{r}_t$, would give the regression an almost perfect fit, with the estimated output elasticities being almost identical to the factor shares. On the other hand, we could persist with a linear time-trend, and adjust the factor inputs, particularly

[12] Shaikh's (1974) critique has been largely ignored in the literature, even more so than that of Phelps Brown and it is ironic that even Simon (1979) does not cite it. The only places where the author has seen Shaikh referenced is Heathfield & Wibe (1987), Lavoie (1992) and Harcourt (1982). Part of the reason may be that Shaikh's argument was very much bound up with a critique of Solow's (1957) method of quantifying the contribution of technical change to growth as well as the difficulties posed by the accounting identity, although the two are logically distinct. The fact that Solow's rejoinder was totally, if erroneously, dismissive—'Mr Shaikh's article is based on misconception pure and simple'—may have been largely responsible for the neglect of his criticism. (This seems to be the reason Heathfield & Wibe discount the argument.) The Shaikh/Solow interchange is dealt with at greater length in McCombie (1997b).

capital, for variations in capacity utilisation. To the extent that capacity utilisation is highly correlated with the deviations of the log of the weighted wage and profit rates from trend, we will find a good fit for the Cobb–Douglas. This is precisely how the time-series estimations of the Cobb–Douglas were rescued in the postwar period (see, for example, Lucas, 1970; Tatom, 1980 and Shapiro, 1993). Often the proxy for capacity utilisation was derived from deviations of output or the output–capital ratio from their trend values and so it is not surprising that it proves to be a reasonably good proxy for the deviations of the weighted wage and profit rates from their trend values. Moreover, it is often found that $a\dot{w}_t + (1-a)\dot{r}_t$ varies procyclically, tending to fall as the economy moves into recession and output growth falls. Thus, it is not surprising that even surveys of capacity utilisation also provide a good proxy for the deviation of the weighted average of the wage and the rate of profit (McCombie, 1997b).

11. What Remains of the Aggregate Production Function?

The implications of this critique are serious for the very notion of the aggregate production function. Suppose that it is assumed that there is a well-behaved Cobb–Douglas production function $P = be^{\lambda t}L^kC^j$. We have noted that under conditions of perfect competition, $\partial P/\partial L = w = kP/L$ and $\partial P/\partial C = r = jP/C$. The elasticity of substitution is given by $d\ln(C/L)/d\ln(w/r)$, which equals unity, and the degree of returns of scale is given by $k + j$. In spite of the severe aggregation problems underlying the concept of the aggregate production function (not to mention the implications of the Cambridge Capital Theory Controversies), the concept is still widely used in macroeconomics. These serious *theoretical* difficulties are often dismissed along the lines of Solow's (1966) comment:

> 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 better comes along.

Moreover, Solow further points out that aggregation is necessary for most of macoeconomics and the concept of the aggregate production function is only marginally less justifiable than, say, the aggregate consumption function (Solow, 1957). Using the criterion of 'good empirical results' as an indication of the usefulness of a theory is reminiscent of Friedman's (1953) instrumental position in his 'Methodology of Positive Economics'.

Judging by the extensive work of both a theoretical and empirical nature on the aggregate production function, it would seem that many neoclassical economists consider that it does give good empirical results and that this legitimises the concept. Recent studies which have used the aggregate production function have included those that have estimated the degree of returns to scale (Hall, 1990), external economies of scale (Caballero & Lyons, 1992) and procyclical productivity growth (Basu, 1996). The aggregate elasticity of substitution has been used to determine the effect of investment grants and labour subsidies on the level of employment (Harris, 1991).

But the argument outlined above with respect to the Cobb–Douglas production function (which may be generalised to other production functions) suggests that because of the underlying cost identity, it will always be possible to find an 'aggregate production function' that gives good statistical results, and the supposed output elasticities must necessarily equal their respective factor shares. (The estimated output elasticities will also show that the production function exhibits constant returns to scale.) In other words, if, for whatever reason, factor shares are constant, the data will always seem to confirm the expected relations of the Cobb–Douglas production function. But this conclusion may well be erroneous. Suppose, for example, that the true aggregate production function is a CES with an elasticity of substitution of 0.4, but trade union bargaining results in factor shares being constant. It follows that the data will give a good fit to the Cobb–Douglas function, even though the 'true' elasticity of substitution is less than unity. Consequently, the good statistical fit provides no guarantee that the estimate parameters are necessarily capturing the true technological relationship of an aggregate production function, even if it exists. One can always proceed on the assumption that the aggregate production function does exist, in spite of the impossibility of providing an empirical justification for this assumption. But then one is faced with explaining away the problems of aggregation, etc. It is necessary to ignore the conclusion of Fisher (1987, p. 55) that 'such results [of the consideration of the aggregation problem] show that the analysis of such aggregates as 'capital', 'output', 'labour' and 'investment' as though the production side of the economy could be treated as a single firm is without sound foundation'. Nor can the neoclassical economist draw any comfort from the analogy with the aggregate consumption function. It is true that there is an underlying identity in this case as well, namely, income equals consumption plus saving. However, taking this relationship into account does not mean that the parameters of interest, such as the marginal propensity to consume, must take predetermined values, as in the case of the aggregate production function.

12. Conclusions

This paper has traced the development of the Cobb–Douglas production function since its inception. It has been partly an exercise in the history of economic thought, but the paper has gone beyond this. It has been shown, by using the original data set, that the initial remarkably good fit using time-series data and closeness of the estimated coefficients to their factor shares is largely illusory, being the result of one or two extreme observations and the absence of a time-trend. The irony is that if Cobb and Douglas had attempted, in the modern terminology, a predictive failure test, the Cobb–Douglas production function might never have seen the light of day. Moreover, when a time-trend is included, the results using Douglas's own data collapse, a finding in common with other studies of the period using different data sets. The emphasis of Douglas's work subsequently changed to the cross-industry studies where the results also seemed to confirm the neoclassical theory of distribution and were more robust. But even those commentators sympathetic to Douglas's aims questioned the meaning of fitting data from a number of industries to a single production function. Moreover, Phelps Brown (1957) later demonstrated that the reason why there was such a good fit and such a close correspondence between the estimated 'output elasticities' and factor shares was that the regression was simply picking up the underlying accounting identity. The subsequent studies, each of which found that k approximately equalled 0.75 and j equalled 0.25, provided no new evidence in support of the marginal productivity theory, since the data could not refute the theory. Shaikh (1974), Simon (1979), McCombie (1987) and McCombie & Dixon (1991) have further elaborated on this line of reasoning, showing that it applies to estimating the Cobb–Douglas and other production functions using time-series data.

The Cobb–Douglas production function has enjoyed a renaissance since Solow's (1957) paper, but the fact that it has achieved much better fits when adjustments for capacity utilisation have been introduced merely serves to strengthen my argument. What these criticisms have established is that the statistical estimation of production functions cannot provide an independent test of the existence or otherwise of an aggregate production function.

References

Basu, S. (1996) Procyclical productivity: increasing returns or cyclical utilisation? *Quarterly Journal of Economics*, 100, pp. 719–751.

Blaug, M. (1974) *The Cambridge Revolution. Success or Failure?* (London, Institute of Economic Affairs).

Blaug, M. (1978) *Economic Theory in Retrospect*, 3rd edn (Cambridge, Cambridge University Press).

Bronfenbrenner, M. (1944) Production functions: Cobb–Douglas, interfirm, intrafirm, *Econometrica*, 12, pp. 35–44.

Bronfenbrenner, M. (1971) *Income Distribution Theory* (London, Macmillan).

Bronfenbrenner, M. & Douglas, P.H. (1939) Cross-section studies in the Cobb–Douglas function, *Journal of Political Economy*, 47, pp. 761–783.

Brown, M. (1966) *On The Theory and Measurement of Technological Change* (Cambridge, Cambridge University Press).

Caballero, R.J. & Lyons, R.K. (1992) External effects in U.S. procyclical productivity, *Journal of Monetary Economics*, 29, pp. 209–226.

Clark, J.B. (1899) *The Distribution of Wealth* (New York, Macmillan).

Clark, J.M. (1928) Inductive evidence on marginal productivity, *American Economic Review*, 18, pp. 451–467.

Cobb, C.W. & Douglas, P.H. (1928) A theory of production, *American Economic Review*, 18 (suppl.), pp. 139–165.

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

Daly, P. & Douglas, P.H. (1943) Production function for Canadian manufacturers, *Journal of the American Statistical Association*, 38, pp. 178–186.

Daly, P., Olsen, E. & Douglas, P.H. (1943) The production function for manufacturing in the United States, *Journal of Political Economy*, 51, pp. 61–65.

de Marchi, N. & Gilbert, C. (1989) Introduction, *Oxford Economic Papers*, 41, pp. 1–11.

Dobb, M. (1973) *Theories of Value and Distribution since Adam Smith. Ideology and economic theory* (Cambridge, Cambridge University Press).

Douglas, P.H. (1934) *The Theory of Wages* (New York, Macmillan).

Douglas, P.H. (1948) Are there laws of production? *American Economic Review*, 38, pp. 1–41.

Douglas, P.H. (1967) Comments on the Cobb–Douglas production function, in: M. Brown (Ed.) *The Theory and Empirical Analysis of Production: studies in income, and wealth*, 31 (National Bureau of Economic Research).

Douglas, P.H. (1972) *In the Fullness of Time: the memoirs of Paul H. Douglas* (New York, Harcourt, Brace, Jovanovich).

Douglas, P.H. (1976) The Cobb–Douglas production function once again: its history, its testing, and some empirical values, *Journal of Political Economy*, 84, pp. 903–115.

Durand, D. (1937) Some thoughts on marginal productivity, with special reference to Professor Douglas's analysis, *Journal of Political Economy*, 45, pp. 740–758.

Fisher, F.M. (1969) The existence of aggregate production functions, *Econometrica*, 37, pp. 553–577.

Fisher, F.M. (1987) Aggregation problem, in: J.L. Eatwell, M. Milgate & P. Newrnan, *The New Palgrave. A dictionary of economics* (Basingstoke, Macmillan).

Fisher, F.M. (1992) Aggregation. Aggregate production functions and related topics, J. Monz (Ed.) (London, Harvester Wheatsheaf).

Flux, A.W. (1894) Review of Wicksteed's *Essay on the Co-ordination of the Laws of Distribution*, reprinted in: W.J. Baumol & S.M. Goldfeld (1968) *Precursors in Mathematical Economics* (London, LSE).

Friedman, M. (1953) The methodology of positive economics, in his *Essays in Positive Economics* (Chicago, Chicago University Press).

Gunn, G.T. & Douglas, P.H. (1941) The production function for American manufacturing for 1919, *American Economic Review*, 31, pp. 67–80.

Gunn, G.T. & Douglas, P.H. (1942) The production function for American manufacturing for 1914, *Journal of Poiitical Economy*, 50, pp. 595–602.

Haavelmo, T. (1944) The probability approach in econometrics, *Econometrica*, 12 (suppl), pp. 1–11.

Hall, R.E. (1990) Invariance properties of Solow's productivity residual, in: P. Diamond (Ed.) *Growth, Productivity, Employment* (Cambridge, MA, MIT Press).

Handsaker, M.J. & Douglas, P.H. (1937) The theory of marginal productivity tested by data for manufacturing in Victoria, *Quarterly Journal of Economics*, 52, pp. 1–36.

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

Harcourt, G.C. (1982) *The Social Science Imperialists*, P. Kerr (Ed.) (London, Routledge & Kegan Paul).

Harris, R.D. (1991) The employment creation effects of factor subsidies, some estimates for Northern Ireland manufacturing industry, 1955–1983, *Journal of Regional Science*, 31, pp. 49–64.

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

Hildebrand, G. & Liu, T.C. (1965) *Manufacturing Production Functions in the United States, 1957* (Ithaca, NY, Cornell University Press).

Houthakker, H.A. (1955) The Pareto distribution and the Cobb–Douglas production function in activity analysis, *Review of Economic Studies*, 23, pp. 27–31.

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

Keynes, J.M. (1939) Professor Tinbergen's method, *Economic Journal*, 49, pp. 558–568.

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

Leser, C.V. (1954) Production functions for the British industrial economy, *Applied Statistics*, III, pp. 174–183.

Lucas, R.E. (1970) Capacity, overtime, and empirical production functions, *American Economic Review, Papers and Proceedings*, 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, pp. 1121–1136.

McCombie, J.S.L. (1997a) Rhetoric, paradigms and the relevance of the aggregate production function, in: P. Arestis & 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. (1997b) *Solow's 'Technical Change and the Aggregate Production Function' Revisited.* Downing College, Cambridge (mimeo).

McCombie, J.S.L. & Dixon, R. (1991) Estimating technical change in aggregate production functions: a critique, *International Review of Applied Economics*, 5, pp. 24–46.

Marshak, J. & Andrews, W.H. (1944) Random simultaneous equations and the theory of production, *Econometrica*, 12, pp. 143–205.

Mendershausen, H. (1938) On the significance of Professor Douglas' production function. *Econometrica*, 6, pp. 143–153. [A correction, *ibid.* 7, p. 1938.]

Menger, C. (1871) *Grundsätze der Volkswirthschaftslehre* (Wien, Wilhelm Braumiller).

Moroney, J.R. (1972) *The Structure of Production in American Industry* (Chapel Hill, University of North Carolina Press).

Morgan, M.S. (1990) *The History of Econometric Ideas* (Cambridge, Cambridge University Press).

Muth, J.F. (1961) Rational expectations and the theory of price movements, *Econometrica*, 29, pp. 1–23.

Pesaran, M.H. & Pesaran B. (1991) *Microfit 3.0* (Oxford, Oxford University Press).

Phelps Brown, E.H. (1957) The meaning of the fitted Cobb–Douglas function, *Quarterly Journal of Economics*, 71, pp. 546–560.

Pigou, A.C. (1933) *Theory of Unemployment* (London, Macmillan).

Robinson, J.V. (1933) *The Economics of Imperfect Competition* (London, Macmillan).

Robinson, J.V. (1970) Capital theory up to date, *Canadian Journal of Economics*, 3, pp. 309–317.

Samuelson, P.A. (1979) Paul Douglas's measurement of production functions and marginal productivities, *Journal of Political Economy*, 87, pp. 923–939.

Sandelin, B. (1976) On the origin of the Cobb–Douglas function, *Economy and History*, 19, pp. 117–125.

Shapiro, M.D. (1993) Cyclical productivity and the workweek of capital, *American Economic Review, Papers and Proceedings*, 83, pp. 229–233.

Shaikh, A. (1974) Laws of production and laws of algebra, the Humbug production function, *Review of Economics and Statistics*, 56, pp. 115–20.

Shaikh, A. (1980) Laws of production and laws of algebra: Humbug II, in: E.J. Nell (Ed.) *Growth, Profits and Property* (Cambridge, Cambridge University Press).

Simon, H.A. (1979) On parsimonious explanations of production relations, *Scandinavian Journal of Economics*, 89, pp. 459–474.

Simon, H.A. & Levy, F.K. (1963) A note on the Cobb–Douglas function, *Review of Economic Studies*, 30, pp. 93–94.

Solow, R.M. (1957) Technical change and the aggregate production function, *Review of Economics and Statistics*, 39, pp. 312–320.

Solow, R.M. (1966) Review of Hicks's Capital and Growth, *American Economic Review*, 56, pp. 1257–1260.

Tatom, J.A. (1980) The 'problem' of procyclical real wages and productivity, *Journal of Political Economy*, 88, pp. 385–394.

Thünen, J.H. von (1826) *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie* (1966 reprint, Stuttgart, Gustav Fischer).

Tinbergen, J. (1939) *Statistical Testing of Business-Cycle Theories, Vol I: a method and its application to business cycle theories* (Geneva, League of Nations).

Tinbergen, J. (1942) On the theory of trend movement, [in German] *Weltwirtschaftliches Archiv*. Reprinted in Tinbergen, J. (1959) *Selected Essays*, L.H. Klassen, L.M. Koyck & H.J. Witteveen (Eds) (Amsterdam, North-Holland).

Wall, B. (1948) Cobb–Douglas function for U.S. manufacturing and mining, 1920–40, *Econometrica*, 16, pp. 211–213.

Wallis, K.F. (1973) *Topics in Applied Econometrics, Lectures in Economics*, 5 (London, Gray-Mills).

Walras, L. (1874) *Éléments d'Economic Pure: Ou Theorie de la Richesse Sociale* (Lausanne, R. Rouge).

Wicksell, K. (1895) Zur Lehre von der Steuerincidenz, Doctoral dissertation, Uppsala University.

Wicksell, K. (1900) Marginal productivity as the basis for distribution in economics, *Economisk Tidskrift*. English translation in: E. Lindhal (Ed.) (1958) K. Wicksell, *Selected Papers on Economic Theory* (London, George Allen and Unwin).

Wicksteed, P. (1894) *An Essay on the Co-ordination of the Laws of Distribution* (London, Macmillan).