

PROBLEMS WITH REGIONAL PRODUCTION FUNCTIONS AND ESTIMATES OF
AGGLOMERATION ECONOMIES: A CAVEAT EMPTOR FOR REGIONAL SCIENTISTS

Jesus Felipe and John McCombie

ABSTRACT Over the last twenty years or so, mainstream economists have become more interested in spatial economics and have introduced largely neoclassical economic concepts and tools to explain phenomena that were previously the preserve of economic geographers. One of these concepts is the aggregate production function, which is also central to much of regional growth theory. However, as Franklin Fisher, *inter alios*, has shown, the conditions to aggregate micro production functions into an aggregate production function are so stringent that in all probability the aggregate production function does not exist. It is shown that the good statistical fits that are commonly found empirically are solely due to the use of value data and an underlying accounting identity. The result is that the estimates obtained cannot be regarded as providing evidence of the underlying technological structure of the spatial economy, including the aggregate elasticity of substitution, the degree of returns to scale, or the rate of technical progress.

KEYWORDS: *Accounting identity, agglomeration economies, regional aggregate production functions*

JEL CLASSIFICATION: B50; O4; R11

CAMBRIDGE CENTRE FOR ECONOMIC AND PUBLIC POLICY

CCEPP WP04-12

DEPARTMENT OF LAND ECONOMY

UNIVERSITY OF CAMBRIDGE

May 2012



Jesus Felipe, Asian Development Bank, Manila, Philippines. Email: jfelipe@adb.org . John McCombie, UK. Email: jslm2@cam.ac.uk. The views in this paper represent solely those of the authors, and not those of the Asian Development Bank, its member countries, or those of its Executive Directors.

1. Introduction

The last twenty years or so have seen analytical tools of mainstream neoclassical economics being increasingly applied to analysing economic geography. There has been an ongoing lively debate between the “new” geographical economists and the economic geographers about the most appropriate methodology for modelling the space economy (Martin, 1999, Garretson and Martin, 2010). Economic geographers, such as Storper (2011), while criticising the assumption of general equilibrium inherent in the New Economic Geography, see the way forward as being a rapprochement between the economic geographers and the New Economic Geography. This is reflected, for example, in the 2006 symposium in the *Journal of Economic Geography*, reflecting “agglomeration and growth: a dialogue between economists and geographers” (Duranton and Storper, 2006). Nevertheless, the Regional Political Economy approach rejects the whole methodological stance of neoclassical economics from the perspective of Marx, Kalecki, and Sraffa (Sheppard and Barnes, 1990, Sheppard, 2000,).

The purpose of this paper is not to discuss the deep methodological issues that still remain, but to present a warning about the uncritical acceptance of a central concept in neoclassical economics that is being widely used in spatial economics, namely, the aggregate production function. It is being used for diverse purposes ranging from estimating the size of agglomeration economies and spatial economic spillovers to determining the rate of regional productivity convergence using Solow’s growth model. For example, Cohen and Morrison Paul (2009, p.106) in their survey of agglomeration, productivity, and regional growth argue that “recognizing and exploiting the potential of production theory models is thus an important step in the empirical literature on agglomeration economies”. See also the surveys on agglomeration economies by Rosenthal and Strange (2004, 2006) and the World Bank (2009, chapter 4). The rate of regional productivity convergence is considered by Armstrong (1995), Barro and Sala-i-Martin (1992, 2004, especially chapter 11), Eckey, *et al.*, (2007), Maier and Trippel (2009), and Tselios (2009). The aggregate production function is central to all these applied studies and used widely in theoretical spatial modelling.

The production function is essentially a microeconomic concept and should be theoretically specified in terms of output and capital measured in homogeneous physical units. There is a substantial technical literature on the “aggregation problem”, which shows that the aggregate production function cannot be derived from micro-production functions, except under the most implausible assumptions (Fisher 1992). Yet, paradoxically, statistical estimations of aggregate production functions give remarkably good fits with plausible estimates. The

answer to this conundrum is that aggregate production functions have to be estimated using data for output and capital measured in constant-price monetary units. This is not an innocuous procedure because of a definitional relationship between output, capital, and labour via an underlying accounting identity. It is this that is responsible for the surprisingly good statistical fits of aggregate production functions. A serious consequence is that it is not possible to interpret the estimates of putative aggregate production functions as technological parameters, such as the elasticity of substitution, the degree of returns to scale (including agglomeration economies) or the rate of exogenous technical progress.

In this paper, we draw together and extend this criticism, showing the importance of the critique for spatial modelling with the intention of drawing the attention of economic geographers and regional scientists to its limitations. For reasons of space, we only sketch the outline of the critique; the reader is referred to the literature for a discussion of the more complex associated issues, important though these may be.¹

We shall demonstrate a paradox. The best statistical fit given by estimating putative regional aggregate production functions must give estimates of “constant returns to scale” with the “output elasticities” equal to their factor shares. This result has been taken as confirming (or rather not refuting) the neoclassical marginal productivity theory of factor pricing (Douglas, 1976, *inter alios*). Regressions that find increasing returns to scale and any differences between the values of the output elasticities and the factor shares do so by virtue of being misspecified. This is irrespective of whether or not increasing returns to scale and agglomeration economies actually exist in reality. It also means that in many cases, with the help of some stylized facts such as a constant capital-output ratio and constant factor shares, which are not dependent upon the existence of an aggregate production function, it is possible to predict the regression estimates before running a single regression.

We reference some regional and urban papers that are subject to our critique. It is not possible, given space considerations, to undertake a comprehensive review of the many individual studies that potentially suffer from this problem. Following Solow (2000) in another context, we leave it to the reader to find Waldo² (or Wally) in other applied studies that use the (regional) aggregate production function. Waldo in this case is the role of the underlying accounting identity in determining the empirical results. Moreover, we do not argue that it is not possible to test for, and find evidence of, agglomeration economies (see, for example, Delgado, *et al.*, 2010a, 2010b, who do not rely on a neoclassical production function), but that any such results that are derived from the neoclassical aggregate production function are extremely problematic in the extreme.

2. Production Functions and External Economies to Scale: Some Theoretical Issues

One of the central tenets of spatial economics is the importance of, especially, agglomeration economies. These have been incorporated into the neoclassical production function in a number of ways. Most approaches usually commence with a firm's micro-production function of the form:³

$$Q_{ij} = A_{ij}F(K_{ij}, L_{ij}) \quad (1)$$

where i is the i th firm in industry I and which is located in city or region j . (We shall generally use the terms city and region interchangeably when discussing the production function.)

Q is the volume of output, K is the capital stock and L is employment. In terms of neoclassical production theory, Q and K are measured in homogeneous physical units.

The production function $F(\bullet)$ is assumed to exhibit constant returns to scale. For expositional ease, and because it generally gives a robust fit to the data, we shall assume that the production function is a Cobb-Douglas, but other more flexible forms such as the CES and the translog production function have also been used. (We discuss the translog below.) The use of other production functions does not affect the argument. Localisation economies in region j are sometimes assumed to be a function of total output Q_{Ij} of industry I in city j . Consequently, this variable is made an argument of the production function:

$$Q_{ij} = A_{ij}K_{ij}^{\alpha}L_{ij}^{(1-\alpha)}Q_{Ij}^{\gamma} \quad (2)$$

where $0 \leq \gamma < 1$.

As agglomeration economies are external to the firm, it remains possible to continue to assume that the individual firm's production function is subject to constant returns to scale and perfect competition. Consequently, wages and the rate of profit are often assumed to be determined by the neoclassical marginal productivity theory of factor pricing. The marginal productivity conditions are controversial because they assume that wages and profits are determined solely by the *technological* conditions of production and the utility maximising supply of labour function. While this does not mean that the distribution of income is necessarily optimal, it does follow that each factor is paid what it contributes to production,

with all the consequent normative implications (Clark, 1891). Consequently, in both theoretical and applied modelling this assumption removes the need to consider bargaining models or sociological explanations of wage determination, which are not normally amenable to mathematical modelling within this framework.

As the data for individual firms are often unavailable, it is erroneously assumed (see the next section) that because there are constant returns to scale (see Nakamura, 1985, p.122) equation (2) may be summed across firms. After rearranging equation (2), the following is obtained for the output of industry I in city or region j :

$$Q_{ij} = A_{ij} K_{ij}^{\alpha/(1-\gamma)} L_{ij}^{(1-\alpha)/(1-\gamma)} \quad (3)$$

where $1/(1-\gamma)$ is the degree of returns to scale. (Anticipating the argument below, the best statistical fit, however, will always give constant returns to scale.)

We could specify more complex relationships where a firm's productivity is affected by the output or productivity of firms in surrounding regions, subject to a distance decay function. These relationships can be picked by spatial econometric regression techniques.

Another approach is simply to assume that $A_{ij} = A_{0ij} \exp(\sum \mathbf{Z})$ where \mathbf{Z} is a vector of variables such as concentration/specialization ratios or an index of the diversity of the industry or total output in city j or its population (see Glaeser, 2000). The estimation often uses pooled data of the various industries. The absence of data for capital stocks at the urban level presents a problem. One way to overcome this is to simply assume the stylised fact that capital and output grow at roughly the same rate or that the output-capital ratio is constant. An alternative approach is to simply specify the production function in terms of labour only, which is the procedure that Glaeser *et al.*, (1992) follow. For a more recent study along these lines, see Bishop and Grippaios, (2010).

This methodology is subject to a number of serious aggregation problems, as the focus of attention moves from the plant or firm to the industry. Let us make the perhaps unrealistic assumption that there exist well-defined micro-production functions, with output and capital measured in physical terms. The question arises, under what conditions can these functions be aggregated to give a "well-behaved" aggregate production function? There is a large technical literature on this subject, but the conclusion is that the conditions are so stringent as to

preclude any meaningful representation of an aggregate production function (Felipe and Fisher, 2003).

There are essentially two problems. The first arises at the firm (plant) level from the necessity to sum over the individual factors of production to give aggregate measures of output, capital and labour. The second is the necessity of aggregating firms' (and industries') production functions. Fisher (2005, pp. 489-490) who has done more work than most on these problems within a quintessentially neoclassical approach (see his collection of essays in Fisher, 1992) came to the conclusion that "except under constant returns, aggregate production functions are unlikely to exist at all. ...Even under constant returns, the conditions for aggregation are so very stringent as to make the existence of aggregate production functions in real economies a non-event. ... One cannot escape the force of these results by arguing that aggregate production functions are only approximations."

Nearly all the work on aggregation has been done on aggregating across firms or industries. To the previous two problems, we must add a third, or spatial aggregation, problem. This results from summing the individual firms' output to give a total for the spatial unit (such as the state or city) under consideration (see McCombie and Roberts, 2007).

Given these theoretical difficulties, why is it that the aggregate production function continues to be used in both theoretical and applied work? The reason why the aggregation criticisms have been largely ignored by the economics profession is related to the instrumental methodology that pervades much of neoclassical economics. Ever since Friedman's (1953) famous "Essay on Positive Economics", the hallmark of an acceptable theory has been its predictive ability. The unrealism of the assumptions does not matter, as the hallmark of a good model is to "explain a lot from a little". Ever since Cobb and Douglas's (1928) first empirical study, estimates of the aggregate production function have generally given a close statistical fit with the sum of estimated output elasticities not significantly different from unity and the individual elasticities close to the factor shares. We shall next explain why this is the case.

3. The Empirical Estimation of Regional Aggregate Production Functions: Too Good to be True?

In this section, we discuss some approaches to the estimation of region and city production functions. In empirical work, physical measures of Q and K are almost never available (except in engineering production functions) and so constant-price monetary values of Q and K , the former henceforth denoted either by Y , (the value of gross output) or V (value added)

and the latter by J , must be used. The neoclassical approach assumes (erroneously) that all the theoretical results above follow through seamlessly to the aggregate level. This is illustrated by the often use of the term “volume of output” for Y and V . It is, of course, not a volume, but a constant-price value measure. In other words,

$$P_t Y_{(0)t} = P_t \sum_i p_{i0} Y_{it} \quad (4)$$

where $Y_{(0)t}$ is the monetary value of gross output at time t measured at base year prices, $t = 0$. p_{i0} are the individual prices of output i at time t , and Y_{it} is the physical quantity of i at time t . P_t is the price deflator, *not* the price, although the latter term is often misused in the literature. Value added is gross output minus the value of material inputs, i.e., their price (m_{i0}) multiplied by the quantity M_{it} , or:

$$P_t V_{(0)t} = P_t \sum_i (p_{i0} Y_{it} - m_{i0} M_{it}) \quad (5)$$

The picture is further complicated by the fact that for a good deal of the economy, most noticeably in the service sector, there is no measure of the output independent of the inputs. In this case, the constant price value of output is often some function of the deflated wage bill, with an arbitrary allowance made for productivity. J is usually calculated by the perpetual inventory method cumulating the constant-price depreciated value of investment. As we shall see, the fact that only constant-price value data and not physical data are available to the researcher becomes a crucial limitation of estimates of aggregate production functions. We define an aggregate production function as one where prices have been used to construct the measures of output and capital. In other words, it applies to not only the whole economy, but even, for example, to the level 4-digit SIC.

The remarkable goodness of fit that the Cobb-Douglas provides to regional data is demonstrated from the following regressions results. Table 1 reports the ordinary least squares estimation of the Cobb-Douglas production function for the 2-digit manufacturing industries for the US states for 1963 (as we are making a general point, the date is not important). Instrumental variable estimation did not make any significant difference to the results.

[Table 1 about here]

Two specifications of the Cobb-Douglas production function were estimated. The first used the aggregate industry values for the states. However, as it could be argued that these values

are determined by arbitrary (non-economic) administrative areas, the Cobb-Douglas was also estimated using state industry per establishment data. As may be seen from Table 1, there was not a great deal of difference between the results of these two specifications. The statistical fits were very close, with \bar{R}^2 s often over 0.950 (the lowest value is 0.854). The estimated output elasticities are close to their respective factor shares and constant returns to scale generally prevail. In fact, in only one industry (SIC 20, Food and Beverages) do both the total and the per establishment data show increasing returns to scale. Moreover, given the close statistical fit of the regressions it can be seen that there is very little variation in the intercept ($\ln A$) for agglomeration economies to explain.

Similar cross-sectional results have been found in a number of other estimations using regional data including Hildebrand and Liu (1965) and Moroney (1972) for the US and Griliches and Ringstad (1971) the latter using firm data for Norway. Moreover, Douglas's numerous estimations in the early 1930s using cross-industry data found equally close statistical fits with the sum of the estimated coefficients not significantly different from zero and the coefficients very close to their respective factor shares.

Constant returns to scale are also found when the production function is estimated for total manufacturing, which could be (erroneously as it will be shown) interpreted as implying that there are no external economies of scale at this higher level of aggregation. If increasing returns were found for total manufacturing, but not at the individual industry, this would be evidence of agglomeration economies. For example, Garofalo and Yamarik (2002) also found using panel-data estimation techniques for total US state manufacturing over the period 1947 to 1996 that the degree of returns to scale (1.008) was not significantly different from unity at the 95% confidence level. "The point estimates [of the output elasticities] furthermore, are close to the standard rule of thumb wherein labor is paid two-thirds of national income (p.320). Results such as this have been taken as implying that the marginal productivity theory of factor pricing holds at the aggregate level (Douglas, 1976).

Two observations are in order here. The first is that, notwithstanding the likely complex relationships between inputs and outputs at the plant level and the severe aggregation problems, it is usually argued by neoclassical economists that the estimation of aggregate production functions, such as those reported here, does not refute the hypothesis that there are "laws of production", which take a remarkably parsimonious form. Secondly, the results suggest the prevalence of constant returns to scale, which not all economists find implausible. (It is the standard assumption in much of neoclassical aspatial macroeconomics,

notwithstanding the development of endogenous growth models. (See, for example, Hoover, 2012, chapter 9, especially p. 330, Romer, 2006, p. 11)

4. On Identities and Behavioural Equations

Neoclassical production theory with the assumptions of homogeneous output and capital, both measured in physical units, constant returns to scale, perfectly competitive markets, no aggregation problems, and factors that are paid their marginal products will give the result that the total value of output will equal the sum of the remuneration of the factors of production. With the use of physical data, it is thus possible to estimate the parameters underlying technology of the production function and to test, for example, the marginal productivity theory of factor pricing. This will not be refuted if the estimated output elasticities equal the factor shares.

However, this all amounts to putting the cart before the horse, as we shall show that with the use of constant-price value data, which nearly all estimations of production functions use, this result must *always* hold, whether or not the above assumptions are satisfied. This is because of the existence of the underlying identity derivable from the product and income accounts. The following equation must always hold by virtue of being an identity:

$$V_i \equiv w_i L_i + r_i J_i \quad (6)$$

where w is the wage rate and r the ex post rate of profit (which may include economic rents).⁴ J is the constant price value of the capital stock. As equation (6) is an identity, an increase in any one of the four variables on the right-hand side, holding the other three constant, *must* by definition lead to an increase in V . There is no error term in equation (6). For example, partially differentiating equation (6) with respect to L gives $\partial V_i / \partial L_i \equiv w_i$, a result which led Phelps Brown (1957) to conclude that it was solely the accounting identity that generated the observed result that the marginal product of labour equals the wage rate. He also argued that this would occur, as equation (6) is an accounting identity, whether or not the usual neoclassical assumptions held. Equation (6) holds for all states of competition and even when no true aggregate production function holds.

The implications are far more serious than merely this, as it may be shown that all that the estimations of putative production functions are accomplishing is a regression of a mathematical transformation of equation (6) with no economic content. It will be shown that

the surprisingly good statistical results of estimating the Cobb-Douglas production function are not surprising at all, but inevitable. There are a number of ways of demonstrating this proposition, but perhaps the simplest is to totally differentiate equation (6) at any point in time to give:

$$\frac{dV_i}{V_i} \equiv a_i \frac{dr_i}{r_i} + (1 - a_i) \frac{dw_i}{w_i} + a_i \frac{dJ_i}{J_i} + (1 - a_i) \frac{dL_i}{L_i} \quad (7)$$

Integrating equation (7) gives the following result:

$$V_i \equiv a_i^{-a_i} (1 - a_i)^{-(1-a_i)} r_i^{a_i} w_i^{(1-a_i)} J_i^{a_i} L_i^{(1-a_i)} \quad (8)$$

or, equivalently,

$$V_i \equiv A_i J_i^{a_i} L_i^{(1-a_i)} \quad (9)$$

where $A_i \equiv a_i^{-a_i} (1 - a_i)^{-(1-a_i)} r_i^{a_i} w_i^{(1-a_i)}$

Equation (9) resembles a Cobb-Douglas form, but note that there is no error term as it is an identity. Equation (9) is exactly true, i.e., it is not an approximation of equation (6) as, given we are dealing with instantaneous changes, the factor shares are constant. Estimation of equation (9) using cross-regional data, provided factor shares are roughly the same for all regions, will show that factor elasticities are close to the factor shares. Therefore they will add up to unity, and the fit will be (almost) perfect. If, however, the factor shares vary, then the Cobb-Douglas production function will not necessarily give the best statistical fit. What is required is a more flexible functional form such as the CES (Felipe and McCombie, 2001) or the translog function (Felipe and McCombie, 2003). For example, the data, V , J , and L from the accounting identity could be used in estimating a Box-Cox transformation, which gives the linear relationship, the Cobb-Douglas or a functional form akin to the CES depending upon the variability of the factor shares. It is an error to assume that this critique only applies to the case where factor shares are constant and that, if they vary, somehow the estimates do refer to an actual aggregate production function.

The supposed aggregate production function (which is erroneously assumed to exist) is given in value terms by $V_i = f(A_i, J_i, L_i)$. Again, assuming a continuum of firms and differentiating, we obtain

$$\frac{dV_i}{V_i} = \frac{dA_i}{A_i} + \alpha_i \frac{dJ_i}{J_i} + \beta_i \frac{dL_i}{L_i} \quad (10)$$

where α and β are the output elasticities.

Comparing this with equation (7) which is definitionally true, the following relationships must *always* hold.

$$\frac{dA_i}{A_i} \equiv a_i \frac{dr_i}{r_i} + (1 - a_i) \frac{dw_i}{w_i} \quad (11.1)$$

$$\alpha_i \equiv a_i \quad (11.2)$$

$$\beta_i \equiv (1 - a_i) \quad (11.3)$$

Consequently, from the identity, it follows that

$$\frac{dV_i}{V_i} \equiv \frac{dA_i}{A_i} + a_i \frac{dJ_i}{J_i} + (1 - a_i) \frac{dL_i}{L_i} \quad (12)$$

which is the same result that would be obtained if all the usual neoclassical assumptions held, but, of course, we have not invoked them in deriving equation (12). It can be simply shown that if there is a neoclassical aggregate production function and all the neoclassical assumptions are met, then from the “dual” of the production function, equation (11.1) will hold. Of course, what we have shown is that it must hold *regardless* of whether or not a well-behaved aggregate production function, or its dual, exists.

If we were able to use physical measures where Q , L and K are not definitionally related, then we have a behavioural relationship. In this case, it is theoretically possible for changes in, say, K to have no significantly statistical effect on Q , holding L constant, the effect being merely captured by the error term. Alternatively, it is theoretically possible to find the coefficients of both K and L statistically insignificant. Consequently, under these circumstances, the existence of a production function is capable of being refuted.

However, as we have seen above, this is *not* the case if we use value data. If we regard equation (9) as relating to cross-section data for, say, regions or cities, then it will only hold exactly if the wage rate, the rate of profit and the factor shares are identical. If they are not,

the goodness of fit obtained by estimating the Cobb-Douglas will now depend upon the degree of the variation of the variables a , $(1-a)$, w , and r across the firms or regions. For reasons that are discussed below, a and $(1-a)$ tend to be relative constant between firms in the same industry. Moreover, w and r in a regional context are unlikely to vary markedly over space, at least compared with the variation in J and L . Thus, we are likely to get a very good fit to the Cobb-Douglas relationship. Simon (1979b) and McCombie (1998) show that even considerable variations in factor shares will still give a good fit to the data, with t -ratios large in comparison with many other cross-sectional regressions of behavioural relationships. Thus, in the light of this, the estimates of the regional production functions estimated discussed above are hardly surprising.

Ideally, in the estimation of the Cobb-Douglas above, spatial econometric techniques should be used. Nevertheless, as all we are estimating are identities, we know what the estimated values of the coefficients should be, namely the values of the factor shares. There are no economic spillover effects between states as the model consists merely of an identity for each state. There is, however, an omitted variable resulting from the assumption in the model that $A_i (= r_i^a w_i^{(1-a)})$ is spatially constant and this misspecification may lead to spatial autocorrelation. Consequently, any spatial autocorrelation will be due, especially, to those states with above average wages being located next to other states with above average wages, and *vice versa*. It seems generally that the weighted logarithm of wage rate and the rate of profit is orthogonal to the other regressors, as the estimates of the output elasticities tend to be close to their factor shares. Consequently, the spatial error model could be used to estimate the Cobb-Douglas relationship. (See Florax and Folmer, 1992 and, for a general discussion, Corrado and Fingleton, 2011.) But this econometric problem is very much of a secondary issue.

The accounting identity critique also applies to the use of time-series data. The accounting identity can be expressed in growth rate form as:

$$\hat{V}_t \equiv a_t \hat{r}_t + (1 - a_t) \hat{w}_t + a_t \hat{J}_t + (1 - a_t) \hat{L}_t \quad (13)$$

Compare this with any neoclassical production function expressed in growth rate form using value data:

$$\hat{V}_t \equiv \hat{A}_t + \alpha_t \hat{J}_t + \beta_t \hat{L}_t \quad (14)$$

A comparison of equations (13) and (14) leads to the conclusion that $\hat{A}_t \equiv a_t \hat{r}_t + (1 - a_t) \hat{w}_t$, $\alpha_t \equiv a_t$ and $\beta_t \equiv (1 - a_t)$.

If there is a “true” aggregate production function with an elasticity of substitution of unity (a Cobb-Douglas) and the usual neoclassical assumptions hold then this will always give constant factor shares. But, as we have seen, it is extremely unlikely that the aggregate production function does exist. However, if for some other reason factor shares are constant, e.g., firms employ a constant mark-up pricing policy, this will ensure that the data give a good statistical fit to the Cobb Douglas (see Fisher, 1971).

To illustrate how the critique applies to more general production functions than the Cobb-Douglas, consider the translog production function:

$$\ln V_t = c + \lambda t + \gamma_J \ln J_t + \gamma_L \ln L_t + 0.5\gamma_{JL} \ln J_t \ln L_t + 0.5\gamma_{JJ} (\ln J_t)^2 + 0.5\gamma_{LL} (\ln L_t)^2 \quad (15)$$

The expressions for the output elasticities may be obtained by differentiating $\ln V$ with respect to $\ln J$ and $\ln L$ respectively. If the assumptions of profit maximization and perfect competition are justified, the output elasticities of capital and labour will equal their respective shares.

$$\frac{\partial \ln V}{\partial \ln J} = \alpha(t) = \gamma_J + 0.5\gamma_{JL} \ln L_t + \gamma_{JJ} \ln J_t = a_t \quad (16)$$

$$\frac{\partial \ln V}{\partial \ln L} = \beta(t) = \gamma_L + 0.5\gamma_{JL} \ln J_t + \gamma_{LL} \ln L_t = (1 - a_t) \quad (17)$$

Furthermore, differentiating equation (15) with respect to time gives:

$$\hat{V}_t = \lambda + \alpha_t \hat{J}_t + (1 - \alpha_t) \hat{L}_t \quad (18)$$

which, using the marginal productivity conditions, can be expressed as:

$$\hat{V}_t = \lambda + a_t \hat{J}_t + (1 - a_t) \hat{L}_t \quad (19)$$

which is identical to equation (14). Consequently, given the observed path of factor shares, the translog “production function” will give a good approximation to the identity.⁵

The heart of the problem is the fact that the production function is a technological relationship, but is estimated using constant price monetary or value data. Estimations of aggregate regional costs functions (which in neoclassical production theory are derived as the “dual” of the aggregate production function) and labour demand functions are also subject to this critique (see Felipe and McCombie, 2009, who also consider the related important critique by Anyadike-Danes and Godley, 1989).

Estimates of “Regional Production Functions” and “Elasticities of Substitution”

Nevertheless, there have been numerous estimations of aggregate regional production functions, most notably the CES, that have attempted to provide an estimate the regional elasticity of substitution. The relevance of this is that these estimates have an important implication for the efficacy of regional and capital and/or labour subsidies (Armstrong and Taylor, 2001, pp.241-246).⁶ Much ink has been spilt in refining the estimation techniques using different specifications. Early estimates, such as those of Tooze (1968), Buck and Atkins (1976), and O’Donnell and Swales (1979) used the marginal productivity of labour side relation in the absence of the capital stock, whereas later studies used explicit measures of the capital stock (Harris, 1991). The problems that aggregation poses for the CES production have been examined by the simulations of Fisher *et al.*, (1977). (Ironically, one of the authors was Solow.) They postulated true underlying micro CES production functions but where aggregation problems meant that no true aggregate production function existed. The simulations generated estimates of the elasticity of substitution that were statistically significant. Nevertheless, as Fisher *et al.* point out, “the elasticity of substitution in these production functions is an ‘estimate’ of nothing; there is no ‘true’ aggregate parameter to which it corresponds” (Fisher *et al.*, 1977, p.312). Indeed, there was a case where the estimated aggregate elasticity of substitution lay outside of the range of the individual micro elasticity of substitutions.

Another representative example is Graham (2000) who used an aggregate production function to estimate the factors that affected the productivity levels of the UK counties. This usefully further illustrates the generic problems posed by the accounting identity. One specification he estimated was the Cobb-Douglas in the form

$$\ln(V_{it} / L_{it}) = \ln A_t + \alpha \ln(J_{it} / L_{it}) + (\alpha + \beta - 1) \ln L_{it} + \ln h(s_{it}) \quad (20)$$

where if the coefficient of $\ln L_i > 0$, there are increasing returns to scale. $\ln h(s_i)$ is a vector of variables supposedly capturing spatial and other externalities. From the identity, given by equation (13), we know that if factor shares are constant the following specification will give a perfect fit to the data:

$$\ln(V_{it} / L_{it}) \equiv c + (1 - a) \ln w_{it} + a \ln r_{it} + a \ln(J_{it} / L_{it}) + (a + (1 - a) - 1) \ln L_{it} \quad (21)$$

The coefficient of $\ln L$ will equal zero. Of course, the term $((1-a)\ln w + a \ln r)$, which is assumed to be spatially invariant in equation (20), is likely to show some variation at the county level. However, provided the term is orthogonal with the other regressors, it is not likely to bias the values of the other coefficients.

For manufacturing the share of capital is about 0.4 and that of labour around 0.6 (for the whole economy, the figure is generally between 0.25 and 0.3). Estimating equation (21) for each of the two years 1984 and 1991 gives, not surprisingly, a very close statistical fit with the two R^2 s of over 90%. The coefficients of $\ln(J/L)$ are 0.426 (t -value of 17) and 0.489 (t -value of 21) for 1984 and 1991 respectively, and those of $\ln L_i$ were only statistically significant for 1984, but, even so, the magnitude was small (-0.075, t -value of -2.6).

In equation (20), $\ln h(s)$ included variables as a skills proxy, the proportion of part-time workers and a dummy for London. All of which were statistically significant. How is this to be explained? If we were estimating the identity, these would be merely irrelevant regressors. However, the fact that some of them are statistically significant is because, especially, $\ln w$ shows some variation across counties. Consequently, it is not surprising that these “externality” variables are significant in the regression as they are likely to be correlated with the logarithm of the wage rate. However, they cannot be interpreted as externality effects within a production function framework.

Graham (2000) also repeated the exercise using the linear approximation to the CES production function (which he calls the translog). Not surprisingly, the statistical fit is better (the R^2 s are 0.96 and 0.95) as this functional form is more flexible than the Cobb-Douglas in that it allows for inter-county variations in factor shares. However, not surprisingly the elasticities of capital at the mean are 0.413 and 0.453, again merely reflecting the value of capital’s share.

The Regional Growth Accounting Approach

In a simulation exercise, Felipe and McCombie (2006) also found that the hypothetical “true” rate of growth of technical change or the growth of total factor productivity differed from the rate calculated using value data. In fact, the growth of total factor productivity calculated using value data is, by definition, the weighted growth of the real wage rate and the rate of profit (see equations (13) and (14)) (Denison, 1967). This has obvious implications for the regional estimates of total factor productivity growth and the regional growth accounting approach.

The growth accounting approach is based on the neoclassical production function and the neoclassical theory of factor pricing. This allows the output elasticities to be measured by the factor shares and the growth of total factor productivity is given by

$$TFP_t \equiv \hat{V}_t - \alpha_t \hat{J}_t - (1 - \alpha_t) \hat{L}_t \equiv \hat{V}_t - a_t \hat{J}_t - (1 - a_t) \hat{L}_t \quad (22)$$

Hulten and Schwab, (1987) used the growth accounting approach to calculate the growth of total factor productivity for the US regions. They found that that output grew twice as fast in what they termed the Sun Belt states compared with the older Snow Belt regions. This could not be explained by differences in the growth of total factor productivity. But all this tells us is that the growth of the real wage rate and the rate of profit, each weighted by its factor share did not greatly differ between the regions, a result hardly surprising given the relative homogeneity of states compared with nations. Using their data, the growth of total factor productivity (which, it is recalled is sometimes misleadingly interpreted as technical progress) accounts for 75 per cent of productivity growth in the Snow Belt and 68 per cent in the Sun Belt, leaving relatively little to be explained by the growth in the capital-labour ratio. This is similar to the magnitudes found by Solow (1957) for the US private sector from 1909 to 1948.

However, if the rate of profit is roughly constant over time, then the *measured* growth of total factor productivity will definitionally be equal to $(1 - a) \hat{w}$ and, if factor shares are roughly constant, this will equal $(1 - a)(\hat{V} - \hat{L})$. Consequently, total factor productivity growth will equal 60 to 75 percent of productivity growth, depending upon labour’s share – a result Solow (1988) for some reason, surprisingly found “startling”.

Other examples include Moomaw and Williams (1991), who calculated total factor productivity for the 48 US states for 1954 -1976. Harris and Trainor (1991) estimated total

factor productivity for regional manufacturing for the UK regions over the period 1968-91. A twist in their procedure is to use Hall's (1988) method to estimate the mark-up (price-cost margin or μ) due to imperfect product markets and to use $\mu(1-a)$ as labour's true output elasticity where the estimates of μ are generally greater than unity. However, without going into details, Hall's model is also a misspecification of the accounting identity, where the misspecification biases the estimate of μ upwards from the "true" value of unity (Felipe and McCombie, 2002). The underlying problems caused by the identity remain. Calculations of total factor productivity growth and its use in regression analysis continues a pace. See Barro and Sala-i-Martin (2004, Chapter 10, "Growth Accounting), and Ascari and Di Cosmo (2004).

5. Regional Convergence and the Aggregate Production Function

As an example of the misleading interpretation that the use of the aggregate production function can entail, we next consider the Solow growth model and its use in estimating the speed of absolute and conditional "beta" regional convergence (see Mankiw *et al.*, 1992). The popularity of this model with some regional economists is surprising given the underlying assumptions, which include perfectly competitive markets, constant returns to scale, a common technology, and a rate of exogenous technological progress that is the same for all regions (i.e., technological progress is a pure public good). These would seem to be most implausible both in a regional and an international context. There are now numerous studies that have estimated absolute and conditional beta convergence for the regions with the more recent ones mainly concentrating on conditional beta convergence. See, for example, Barro and Sala-i-Martin (1992, 2004), Armstrong (1995), Evans and Karras (1996), Rey and Montouri (1997), Garofalo and Yamarik (2002), Badinger *et al.*, (2004), Egger *et al.*, (2006), Benos and Karagiannis (2008), Esposti and Bussoletti (2008), and Tselios (2009). See the survey of Martin and Sunley (1998).

The steady-state productivity growth rate in the Solow growth model is equal to the rate of technical change. The reason why some regions have faster growth than others is that their capital-labour ratios are below the steady-state level. There is thus a transitional disequilibrium component to their growth rate as these regions catch up and converge to their steady-state growth rates. Conversely, regions where the capital-labour is above their steady-state value will have temporarily slower growth than the steady-state rate. (A major problem with this analysis is that the Solow model predicts convergence, but for this to have occurred there needs to have been *divergent* growth in the first place. Indeed, world growth since the

industrial revolution has been as Prichett (1997) puts it “divergence, big time”. Apart from resorting to an explanation in terms of substantial *ad hoc* shocks, the theory is silent on this point.)

Testing this is done by log-linearizing the aggregate production function around its steady-state growth path to give:

$$\frac{1}{\tau} \ln \left(\frac{y_{it}}{y_{i,0}} \right) = c - \left(\frac{1 - e^{-\eta\tau}}{\tau} \right) \ln y_{i,0} \quad (23)$$

where $y_{i,t}$ is the level of labour productivity in region i at time t , and η is the speed of convergence. τ is the length of period over which the growth rate is calculated i.e., $\tau = t - 0$ where 0 denotes the initial year. For many regional data sets, the estimate of η is statistically significant and, consequently, the rate of convergence takes a value of around 2 per cent per annum (Barro and Sala-i-Martin, 2004).

The problem with this specification, remaining within the neoclassical framework, is that it ignores regional variations in the investment rate, which, while not affecting the common steady-state rate of productivity growth, will affect the steady-state level of productivity. In a seminal paper Mankiw *et al.*, (1992) overcame this shortcoming. Starting with a conventional Cobb-Douglas production function of the form $Q = (A(t)L)^{(1-\alpha)} K^\alpha$, Mankiw, Romer and Weil (MRW) showed that the *steady-state* Solow growth model can be expressed in level form as:

$$\ln y_{it} = \ln A_0 + \lambda t + \frac{\alpha}{(1-\alpha)} \ln s_{Kit} - \frac{\alpha}{(1-\alpha)} (n_{it} + \delta + \lambda) \quad (24)$$

A_0 is again the level of total factor productivity, which is assumed to be the same for all countries (and $\ln A_t = \ln A_0 + \lambda t$), s_K is the ratio of investment in physical (as opposed to human) capital to total output, n is the growth of employment (which is also sometimes assumed to be constant across regions), δ is the rate of depreciation and λ is the common rate of technical progress. The parameters α and $(1-\alpha)$ are the output elasticities of physical capital and labour. It is assumed that $\ln A_t = \ln A_0 + \lambda t$ can be proxied by a constant when

cross- country (or cross-regional) data are used. In other words, both the level and rate of growth of technology are assumed to be the same for all the regions.

This equation was estimated by MRW using the Penn-World Tables and cross-country value data for advanced and developing countries. While the results were reasonable, except for the OECD countries, they found that the estimate of α took a value of about two-thirds, higher than its factor share of approximately one-third. This led them to augment the Solow growth model by introducing the logarithm of the share of human capital in total output as an explanatory variable. The estimate of the coefficient of physical capital now became about one-third, roughly equal to its factor share.

However, if the stylized facts that factor shares and the capital-output ratio are constant hold, then, *solely* from the accounting identity, we can derive the following equation (see Felipe and McCombie 2005b)⁷:

$$\ln y_{it} \equiv c + 1.0 \ln w_{it} + \frac{a}{(1-a)} \ln r_{it} - \frac{a}{(1-a)} \ln \left(n_{it} + \delta + \frac{(1-a)\hat{w}_{it} + a\hat{r}_{it}}{(1-a)} \right) \quad (25)$$

It can be seen that this is identical to equation (24) where from the two equations (25) and (24):

$$\ln A_i = \ln A_o + \lambda t = \ln w_{i0} + \frac{a}{(1-a)} \ln r_{i0} + \hat{w}_i t + \frac{a}{(1-a)} \hat{r}_i t \quad (26)$$

if the wage rate and the rate of profit grow at a constant rate over time. Note that empirically the initial level of the (log) of the wage rate varies considerably across countries, whereas $\ln r_{i0}$ shows little systematic variation. We can see from equation (25) that assuming that $\ln A_i = \ln A_o + \lambda t$ is constant across countries/regions will cause omitted variable bias if one estimates equation (26). This is because as we have noted $\ln w_i$ varies significantly across countries (λ and the growth rates of the wage rate and the rate of profit will also differ between countries). It is solely for this reason that MRW failed to find a perfect statistical fit to the underlying accounting idea and were compelled to introduce human capital.

Support for our argument is provided by Easterly and Levine (2001) who generalised the MRW model by using dummy variables to allow $\ln A$ to differ between the major regions of the world.⁸ They found that “South Asia and Sub-Saharan Africa have significantly lower

productivity than the other regions (income differences that are not explained by the MRW term). The OECD has higher productivity than the rest of the world by a factor of 3 Once the productivity level is allowed to vary, the coefficient on MRW implies a capital share of .31 – which is line with most estimates from national income accounting” (Easterly and Levine (2001, p.190). This is without including the human capital variable. Including the human capital variable in the equation brings the estimate of α closer to a because it is closely correlated with lnw and hence brings the estimating equation closer to the identity.

But we can see immediately why they get this result. Because of the underlying accounting identity, the result is inevitable as the dummies are closely proxying the variation of the lnw_{0i} term in the identity (or lnw_{it} as the two will be highly correlated) bringing the estimated regression coefficients closer to those of the identity. If we were also to allow the growth rates of “total factor productivity” (λ , or $\hat{w} + a/(1-a)\hat{r}$) to vary between countries (as the neoclassical growth accounting approach shows should be the case) rather than imposing a constant growth rate, we would end up merely estimating the identity. (MRW’s assumptions are inconsistent with the standard neoclassical growth accounting results on this point, and *vice versa*.)

The rate of conditional beta convergence is estimated by subtracting the logarithm of initial level of productivity from both the right hand and the left hand side of equation (25). It can be seen that if equation (26) is used (which under a neoclassical interpretation implies that countries have different levels and growth rates of total factor productivity) then the coefficient on lny_0 will equal minus unity. This implies an infinitely fast rate of convergence. The fact that lnw is highly correlated with lny means that excluding the former, will bias the absolute value of the coefficient of lny downwards, giving the estimates generally found empirically. Nevertheless, the MRW model has been used at the regional level.

It is interesting that when Lee *et al.*, (1998) allow λ to vary the estimate of η increases from 0.04 to 0.29. As Islam (1998, p. 326) concedes “when, in addition, heterogeneity in growth rates is allowed convergence becomes, in essence, an empty concept”. If Lee *et al.*, (1998) had allowed lnA to vary as well, then the estimated speed of convergence would have tended towards infinity, as we have noted above. Convergence, within MRW’s framework of the augmented Solow growth model is indeed an empty concept.

The fact that there is an inverse correlation between the growth of regional productivity and the initial level of productivity does show regional per capita GDP levels are converging. But

there could be many reasons for this. It may be because the regions with lower levels of productivity are technologically more backward and are benefiting from the diffusion of innovations from the more advanced regions. There are strong political pressures to reduce income per capita disparities, especially in the advanced countries. This is accomplished not only through fiscal transfers, but also through the spatial pattern of government expenditure and regional subsidies, either as the result of an explicit or implicit regional policy. In the enlarged EU many of the low productivity regions are those that depended heavily on agriculture, and with the relative decline of this sector and the transfer of labour to the higher productivity industrial sector, regional productivity can be expected to rise in the less developed regions. Thus, the estimation of either absolute or conditional beta convergence is not a test of the (augmented) Solow model.

The endogenous growth models, to the extent that they use the aggregate production function also suffer from the problem of the accounting identity. Take, for example, the “linear-in-capital” or $Q = AK$ model (expressed here in physical terms), where the exponent of capital is unity and captures the effect of both the direct contribution of capital and the effect of induced technical change. A is now a constant. Let us assume that the Kaldorian stylized facts of constant factor shares, no trend in the rate of profit and no discernable growth in the output capital ratio. See Simon (1986, Appendix A, pp. 172-183) for an explanation of why value data will give a constant capital output ratio this irrespective of the ratio between *physical* output and capital. Under these circumstances the accounting identity, equation (13) may be written as:

$$\hat{V}_t \equiv (1-a)\hat{w}_t + (1-a)\hat{L}_t + a\hat{J}_t \equiv (1-a)(\hat{V}_t - \hat{L}_t) + (1-a)\hat{L}_t + a\hat{J}_t \equiv \hat{J}_t \quad (27)$$

Consequently, integrating this equation gives $V_t = AJ_t$ where A is the constant of integration and does not vary over time.

A second generation of endogenous growth models allows A to change over and which is interpreted as the “stock of knowledge” or “ideas”. Consequently, A is produced by a separate R&D production function of the form:

$$\Delta A_t = \delta L_{R\&D}^\sigma A_t^\phi \quad (28)$$

where $L_{R\&D}$ is the number of scientists and $0 < \sigma < 1$ and $\phi < 1$ are parameters of the R&D production function. (See Martin and Sunley, 1998, Armstrong and Taylor, 2000, and

Roberts and Setterfield, 2010, for discussions of endogenous growth in a regional context.) But, as Steedman (2003) points out, this requires A to be cardinal measure and no explanation has been put forward to show that it is. Consequently, the “analysis of the [New Growth Theory] equations can having no meaning cannot yield convincing conclusions” (p. 133).

Izushi's (2008) attempt to test various specifications of equation (28) using data for the NUTS1 regions of the European Union clearly demonstrates this problem. \hat{A} was calculated as the growth of total factor productivity, which, as we have seen, is equal to $a\hat{w}$, given the assumption that there is no rate of change in the rate of profit.⁹ The only statistically significant relationship found by Izushi, and which “shows strong support for the ... Lucasian framework” (Izushi, 2008, 955), is the one where regressing the growth of total factor productivity on that private sector R&D workers gives a slope coefficient of 0.25. This value is most implausible. It implies increasing the growth rate of R&D workers by, say, ten percentage points per annum will increase the growth of real wages in the EU regions by over 2.5 percentage points per annum. Such an increase in the growth rate of R&D workers should not be difficult, given the supposed economic benefits, as they only consist of about 1% of total EU employment. (The results also suggest that, from the estimated coefficient of the spatial-lag term, a one percentage point increase in the growth of the wage rate of the surrounding regions will implausibly increase the region's growth rate by 0.6 percentage points. Increasing the share of R&D personnel in total employment also had a further positive impact on the growth of the real wage.) Clearly, to test the endogenous growth model satisfactorily requires an independent measure of the stock of knowledge.

6. A Way Forward?

The above discussion is somewhat nihilistic and is what Lawson (2004) claims an exercise in under-labouring; or as Locke (1690) put it: “removing some of the rubbish that lies in the way of knowledge” (cited by Lawson, 2004, p. 317). But as Kuhn (1970) pointed out, one paradigm, no matter how (logically) flawed, is only abandoned if it is replaced by another paradigm. It is beyond the scope of this paper to discuss alternative paradigms and we merely make a few suggestions as to possible ways forward.

At the aggregate level, a more fruitful way of analysing regional growth process is the post Keynesian demand-oriented export-led growth approach which emphasizes not only increasing returns and agglomeration economies but the role of the structure of production

(captured through differences in income elasticities of demand for a regions exports). It is flexible enough to capture the importance of monetary factors, interregional capital flows and sovereign debt problems that are currently threatening the existence of the euro zone and on which the neoclassical approach is noticeably silent (Rowthorn, 2010, McCombie, 2012). One central tenet of this approach is the Verdoorn law (see McCombie *et al.*, 2002), a linear specification of Kaldor's technical progress function. It has similarities to the production function although Kaldor eschews the marginal productivity theory of factor pricing and the distinction between movements along and shifts of the production function. What then are the implications of the critique for the Verdoorn law given the need to use value data here as well? There are a number of specifications of the Verdoorn law, one of which explicitly includes the growth of capital and takes the form

$$TFP_t = \lambda + v\hat{V}_t \quad (29)$$

where TFP is the growth of total factor productivity, λ the exogenous growth of total factor productivity, \hat{V} the growth of output and v is the Verdoorn coefficient ($1 > v > 0$ implies encompassing increasing returns to scale). More elaborate specifications include other variables such as the density of production and the model is now routinely estimated using spatial econometric methods. What are the implications of the accounting identity critique for the Verdoorn law? The Verdoorn law does have behavioural implications, notwithstanding the accounting identity critique. From the accounting identity in growth rate form (equation (13)), equation (29) may be alternatively expressed as:

$$TFP_t \equiv a\hat{r}_t + (1-a)\hat{w}_t = \lambda + v\hat{V}_t \quad (30)$$

or

$$(a\hat{r}_t + (1-a)\hat{w}_t) = \frac{\lambda}{v} + \frac{v}{(1-v)}(a\hat{J}_t + (1-a)\hat{L}_t) \quad (31)$$

where λ is a constant.

Consequently, the Verdoorn law does, in fact, demonstrate an empirical relationship, even allowing for the accounting identity. That is to say, a faster growth of the region's weighted factor inputs will increase the weighted growth rate of the factor prices.¹⁰

The importance of agglomeration economies and clusters is being extensively studied by Michal Porter's Cluster mapping project which uses statistical analysis to determine the extent of clusters. While the concept of the cluster used by Porter is not without its critics (Martin and Sunley, 2003), there is no doubt that it is providing useful insights into regional growth. The approach, with its emphasis on clusters determining the export performance of the region (Porter, 2003, p. 572, fn 13), complements the post Keynesian approach noted above (McCombie, 2012).¹¹ Porter (2003) presents some statistical estimates of the importance of clusters without using a production function. A more recent study by Delgado *et al.*, (2011) also confirms the importance of clusters. It is found that the growth of employment of 4-digit SIC clusters is determined by the degree of regional specialization in the industry, the relative strength of the cluster environment surrounding that industry and a measure of cluster strength of other industries used in the cluster. Delgado *et al.*, (2010) also demonstrate statistically the importance of clusters in the start-up of new firms which reflects the degree of entrepreneurship.

There is also a considerable literature explaining regional growth disparities through regional innovation systems, industrial districts and regional inter-linkages, which in many cases use case studies and does not rely on regional production functions. There is not space to discuss these approaches here but surveys can be found in Clark, *et al.*, (2000 Part IV, "the Geography of Learning". See also, *inter alia*, the articles in Dunning (2000) and Breschi *et al.*, (2005).

6. Conclusions

The aggregate production function has had a chequered history ever since it was first introduced by Cobb and Douglas (1928). In particular, it is now well established theoretically that micro Cobb-Douglas production functions cannot be summed to give an aggregate production function, except under most implausible assumptions. These reservations (together with those of the capital controversies) were discussed in most textbooks on economic growth prior to around 1975 and then were conveniently forgotten – there is no mention of them in growth textbooks, such as Barro and Sala-i-Martin (2004). (See the discussion in McCombie, 2011.) The standard instrumentalist defence that aggregate production functions "work", in that they empirically give close statistical fits with plausible estimates is unsound. These estimations of the aggregate production function have to use constant price monetary data for output and capital and underlying accounting is responsible for the goodness of fit. The only case where this problem does not occur is when physical or engineering data are used and such studies, which almost certainly must be at the plant level, are far and few between.

The aggregate production function has now become widely used in spatial economics, for example, to estimate the degree of agglomeration economies, to calculate the rate of regional technical progress, to model regional economic growth and the rate of convergence or divergence, and to estimate the elasticity of substitution. It also has important policy implications. For example, the elasticity of substitution of the regional production function has been used to estimate the effect of regional capital and/or labour subsidies. But such calculations are fatally flawed, as the aggregate elasticity of substitution does not exist.

This paper serves as a warning to the continued uncritical use of the aggregate production function in economic geography (or, indeed, in macroeconomics, *per se*).

Notes

[Here]

References

- Anyadike-Danes, M. & Godley, W. (1989), Real wages and employment: a sceptical view of some recent empirical work, *The Manchester School of Economic & Social Studies*, 57(2), 172-87.
- Armstrong, H.W. (1995) Convergence among regions of the European Union, 1950-90, *Papers in Regional Science*, 74(2), 143-52.
- Armstrong, H.W. & Taylor, J. (2000) *Regional Economics and Policy*, Oxford: Basil Blackwell.
- Ascari, G. & Di Cosmo, V. (2004) Determinants of Total Factor Productivity in the Italian Regions, (available at <http://economia.unipv.it/new/q47.pdf>).
- Badinger, H., Müller, W., & Tondl, G. (2004) Regional convergence in the European Union, 1985-1999: a spatial dynamic panel analysis, *Regional Studies*, 38(3) 241-253.
- Barro, R. & Sala-i-Martin, X. (1991) Convergence across states and regions, *Brookings Papers*, 1, 107-82.
- Barro R. & Sala-i-Martin, X. (2004) *Economic Growth*, Cambridge Mass: MIT Press (first edition, 1993).
- Benos, N. & Karagiannis, S. (2008) Convergence and economic performance in Greece: evidence at regional and prefecture level, *Journal of the Applied Regional Science Conference*, 20(1), 52-69.
- Bishop, P. & Gripaos, P. (2010) Spatial externalities, relatedness and sector employment growth in Great Britain, *Regional Studies*, 44(4), 443-454.

- Breschi, S. & Malerba, F. (eds) (2005) *Clusters, Networks, and Innovation*, Oxford: Oxford University Press.
- Buck, T.W. & Atkins, M.H. (1976) Capital subsidies and unemployed labour, a regional production function approach, *Regional Studies*, 10(2), 215-222.
- Clark, G.L., Feldman, M. P. & Gerlter M.S. (2000) *The Oxford Handbook of Economic Geography*, Oxford: Oxford University Press.
- Clark, J. B., (1891) Distribution as determined by the law of rent, *Quarterly Journal of Economics*, 5(3) 289-318.
- Cobb, C. W. & Douglas, P.H. (1928) A theory of production, *American Economic Review Supplement, Papers and Proceedings*, 18(1) 139-165,.
- Cohen, J.P. & Morrison Paul, C.J. (2009) Agglomeration, productivity and regional growth: production theory approaches, in Capello, R., & Nijkamp, P. (eds) *Handbook of Regional Growth and Development Theories*, Cheltenham: Edward Elgar.
- Corrado, L and Fingleton B (2011), Where is the economics in spatial econometrics? *Journal of Regional Science*, 52(2), 210-239.
- Delgado, M., Porter, M.E, & Stern, S. (2010) Clusters and entrepreneurship, *Journal of Economic Geography*, 10(4), 495 -518.
- Delgado, M., Porter, M.E, & Stern, S. (2011) Clusters, convergence and economic performance. (available at <http://www.isc.hbs.edu/econ-clusters>)
- Denison, E.F. (1967) *Why Growth Rates Differ. Post-war Experience in Nine Western Economies*, Washington, D.C.: The Brookings Institute.
- Douglas, P.H. (1976). The Cobb-Douglas production function once again: its history, its testing, and some new empirical values, *Journal of Political Economy*, 84(5), 903-15.
- Dunning, J.H. (ed.) (2000) *Regions, Globalization and the Knowledge-based Economy*, Oxford: Oxford University Press.
- Duranton G. & Storper, M. (2006) Agglomeration and growth: a dialogue between economists and geographers, *Journal of Economic Geography*, 6(1), 1-7.
- Easterly, W. & Levine, R. (2001) It's not factor accumulation: stylized facts and growth models, *World Bank Economic Review*, 15(2), 177-219.
- Eckey H-F, Kosfeld, R. & Turck, M. (2007) Regional convergence in Germany: a geographically weighted regression approach, *Spatial Economic Analysis*, 2(1), 45-64.
- Egger, P.M. & Pfaffermay K.R. (2006) Spatial convergence, *Annals of Regional Science*, 85(2), 199-215.
- Esposti, R. & Bussoletti, S. (2008) Impact of Objective 1 Funds on regional growth convergence in the European Union: a panel-data approach, *Regional Studies*, 42(2), 159-173.
- Evans, P. & Karras, G. (1996) Do economies converge? evidence from a panel of U.S. states, *Review of Economics and Statistics*, 78(3), 384-388.

- Felipe, J. & Fisher, F.M. (2003) Aggregation in production functions: what applied economists should know, *Metroeconomica*, 54(2-3), 208-262.
- Felipe, J. & McCombie, J.S.L. (2001) The CES production function, the accounting identity and Occam's razor, *Applied Economics*, 33(10), 1221-1232.
- Felipe, J. & McCombie, J.S.L. (2002) A problem with some recent estimations and interpretations of the mark-up in manufacturing industry. *International Review of Applied Economics* 16 (2), 187–215.
- Felipe, J., and McCombie, J.S.L. (2003), Methodological Problems with Neoclassical Analyses of the East Asian Miracle, *Cambridge Journal of Economics*, 27(5), 695-721.
- Felipe, J & McCombie, J.S.L. (2005a) How sound are the foundations of the aggregate production function? *Eastern Economic Journal*, 31(3), 467-488. Updated version in Harcourt, G.C. & Kriesler, P (eds) *The Oxford Handbook of Post-Keynesian Economics*, Oxford: Oxford University Press (forthcoming, 2013).
- Felipe, J. & McCombie, J.S.L. (2005b) Why are some countries richer than others? A skeptical view of Mankiw-Romer-Weil's test of the neoclassical growth model, *Metroeconomica*, 56(3), 360-392.
- Felipe J. & McCombie, J.S.L. (2006) The tyranny of the identity. Growth accounting revisited, *International Review of Applied Economics*, 20(3), 283-299.
- Felipe, J. & McCombie, J.S.L. (2007) On the rental price of capital and the profit rate. The perils and pitfalls of total factor productivity growth, *Review of Political Economy*, 19(3), 317-345.
- Felipe, J. & McCombie, J.S.L. (2009) Are estimates of labour demand functions mere statistical artefacts?, *International Review of Applied Economics*, 23(2), 147-168.
- Felipe, J and McCombie, J.S.L. (2012a) The aggregate production function: Not even wrong, *Cambridge Centre for Economic and Public Policy*, Department of Land Economy, University of Cambridge, WP 03-12.
- Felipe, J and McCombie, J.S.L. (2012b) The aggregate production function and the accounting identity critique: Further reflections on Temple's criticisms and misunderstandings, *Levy Institute of Economics*, Bard College, Working Paper No. 718, May.
- Fisher, F.M. (1971) Aggregate production functions and the explanation of wages: a simulation experiment, *Review of Economics and Statistics*, 53(4), 305-325.
- Fisher, F.M. (1992) *Aggregation. Aggregate Production Functions and Related Topics*, Monz, J. (ed.) London: Harvester Wheatsheaf.
- Fisher, F.M. (2005) Aggregate production functions – a pervasive, but unpersuasive, fairytale, *Eastern Economic Journal*, 31(3), 489-491.
- Fisher, F. M., Solow R. M. & Kearn J. M. (1977) Aggregate production functions: some CES experiments, *Review of Economic Studies*, 44(2), 305-20.

- Florax, R.J.G.M. & Folmer H. (1992) Specification and estimation of spatial linear regression models. Monte Carlo evaluation of pre-test estimators, *Regional Science and Urban Economics*, 22(3), 405-432.
- Friedman, M. (1953) The methodology of positive economics, in Friedman. M (ed.) *Essays in Positive Economics*, Chicago: Chicago University Press.
- Garofalo G.A. & Yamarik, S. (2002) Regional convergence: evidence from a new state-by-state capital stock series," *Review of Economics and Statistics*, 84(2), 316-323.
- Garretson, H. & Martin R. L (2010) Rethinking (new) economic geography models: taking geography and history more seriously, *Spatial Economic Analysis*, 5(2), 127-160
- Glaeser, E.L. (2000) The new economics of urban and regional growth in Clark, G.L., Feldman, M.P., & Gertler, M.S. (eds) *The Oxford Handbook of Economic Geography*, Oxford: Oxford University Press.
- Glaeser, E.L., Kallal, H.D., Scheinkman, J.A. & Shleifer, A. (1992) Growth in cities, *Journal of Political Economy*, 100(6), 1126- 1152.
- Graham, D.J. (2000), Spatial variation in labour productivity in British manufacturing , *International Review of Applied Economics*, 14(3), 323- 340.
- Griliches, Z. & Ringstad, V. (1971) *Economies of Scale and the Form of the Production Function*, Amsterdam: North Holland.
- Hall, R. E. (1988) Market structure and macroeconomic fluctuations", *Brookings Papers on Economic Activity*, 2, 285-322.
- Harris, R.I.D. (1991) The employment creation effects of factor subsidies: some estimates for Northern Ireland manufacturing. 1955-83, *Journal of Regional Science*, 31(1), 49-64
- Harris, R.I.D. & Trainor, M. (1997) Productivity growth in the UK regions, 1968-1991, *Oxford Bulletin of Economics and Statistics*, 59(4), 485-509.
- Hausmann, R., Hwang, J. and Rodrik, D. (2006), "What You Export Matters", *Journal of Economic Growth*, 12(1), 1-25,
- Hildebrand, G. & Liu, T.C. (1965) *Manufacturing Production Functions in the United States, 1957*, Ithica, New York: Cornell University Press.
- Hoover, K.D. (2012) *Applied Intermediate Macroeconomics*, Cambridge: Cambridge University Press.
- Hulten, C. R. & Schwab, R.M. (1984) Regional productivity growth in U.S. manufacturing: 1951-78, *American Economic Review*, 74(1), 152-162.
- Islam, N. (1998) Growth empirics: a panel data approach - a reply *Quarterly Journal of Economics*, 113(1), 325-329.
- Izushi, H. (2008), What does endogenous growth theory tell us about regional economies? Empirics of R&D and worker-based productivity growth, *Regional Studies*, 42(7), 947- 960.

- Kuhn, T. (1970) *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press (second edition).
- Lawson, T. (2004) Philosophical under-labouring in the context of modern economics: aiming at truth and usefulness in the meanest of ways, in Davis, J.B., Marciano, A. & Runde, J. (eds) *The Elgar Companion to Economics and Philosophy*, Cheltenham: Edward Elgar.
- Lee, K., Pesaran, M.H. & Smith, R. (1998) Growth empirics: a panel data approach - a comment, *Quarterly Journal of Economics*, 113(1), 319-323.
- Leichenko, R.M. (2000) Exports, Employment, and Production: A Causal Assessment of U.S. States and Regions, *Economic Geography*, 76(4), 303-325.
- Locke J. (1690) *An Essay Concerning Human Understanding*, Menston: Scolar Press, 1970.
- McCombie, J.S.L. (1998). Are there laws of production?: an assessment of the early criticisms of the Cobb-Douglas production function," *Review of Political Economy*, 10(2), 141-173.
- McCombie, J.S.L. (2011) 'Cantabrigian Economics' and the aggregate production function, *Intervention*, 8(1), 165-182.
- McCombie, J.S.L. (2012) Explaining regional growth rate disparities; An assessment of the demand-oriented approach. *Cambridge Centre for Economic and Public Policy, Department of Land Economy, University of Cambridge* (mimeo).
- McCombie, J. S. L. & Roberts, M. (2007) Returns to scale and regional growth: the static-dynamic Verdoorn law paradox revisited, *Journal of Regional Science*, 47(2), 179-208,
- McCombie, J.S.L., Pugno, M., & Soro, B. (2002), *Productivity Growth and Economic Performance. Essays on Verdoorn's Law*. Basingstoke: Macmillan.
- Maier, G. & Trippel, M. (2009) Location/allocation of regional growth in Capello, R., & Nijkamp, P. (eds) *Handbook of Regional Growth and Development Theories*, Cheltenham: Edward Elgar.
- Mankiw, N.G., Romer, D., & Weil, D.N. (1992) A Contribution to the Empirics of Economic Growth, *Quarterly Journal of Economics*, 107(2), 407-437.
- Martin R. (1999) Critical survey. The new 'geographical turn' in economics: some critical reflections, *Cambridge Journal of Economics*, 23(1), 65-91.
- Martin, R. L. and Sunley, P. (1998), Slow convergence? The new endogenous growth theory and regional development, *Economic Geography*, 74(3), 201-227,
- Martin R.L. and Sunley P. (2003) Deconstructing clusters: chaotic concept or policy panacea? *Journal of Economic Geography*, 3(1), 5-35.
- Minondo, A. (2010), Exports' productivity and growth across Spanish regions, *Regional Studies*, 44 (5), 568-577.
- Moomaw, R. & Williams, M. (1991). Total factor productivity growth in manufacturing: further evidence from the States, *Journal of Regional Science*, 31(1), 17-34.

- Moroney, J.R. (1972) *The Structure of Production in American Industry*, Chapel Hill: University of North Carolina Press.
- Nakamura, R (1985) Agglomeration economies in urban manufacturing industries: a case of Japanese cities, *Journal of Urban Economics*, 17(1), 108-124.
- O'Donnell, A. T. and Swales, J.K. (1979) Factor substitution, the C.E.S. production function and U.K. regional economics, *Oxford Economic Papers*, 31(3), 460-476.
- Phelps Brown, E.H. (1957) The meaning of the fitted Cobb-Douglas function, *Quarterly Journal of Economics*, 71(4), 546-560.
- Porter, M.E. (2003), The economic performance of regions, *Regional Studies*, 37(6-7), 549-578.
- Pritchett, L., (1997), Divergence, big time, *Journal of Economic Perspectives*, 11(3), 3-17.
- Rey, S J. & Montouri, B.D. (1997) US regional income convergence: a spatial econometric perspective, *Regional Studies*, 33(2), 143-156.
- Roberts, M. & Setterfield, M. (2010) Endogenous regional growth: a critical survey in Setterfield, M. (ed.) *Handbook of Alternative Theories of Economic Growth*, Cheltenham: Edward Elgar.
- Romer, D. (2006) *Advanced Macroeconomics*, New York, McGraw-Hill Irwin.
- Rosenthal, S. S. & Strange, W. C., (2004) Evidence on the nature and sources of agglomeration economies, in Henderson, J.V. & Thisse, J.F. (eds) *Handbook of Regional and Urban Economics, Volume 4*, chapter 49.
- Rosenthal, S. S. & Strange W.C. (2006) The micro-empirics of agglomeration economies in Arnott, R.J. and McMillen, D.P. (eds) *A Companion to Urban Economics*, Oxford: Basil Blackwell, 7-23.
- Rowthorn, R.E. (2010) Combined and uneven development: Reflections on the North-South divide, *Spatial Economic Analysis*, 5(4), 363-388.
- Shaikh, A. (1974) Laws of production and laws of algebra: the Humbug production function, *Review of Economics and Statistics*, 56(1), 115-120.
- Shaikh, A. (1980) Laws of production and laws of algebra: Humbug II, in Nell, E.J. (ed.) *Growth, Profits and Property, Essays in the Revival of Political Economy*, Cambridge: Cambridge University Press, 80-95.
- Sheppard, E. (2000) Geography or economics? Conceptions of space, time, interdependence, and agency in Clark, G.L., Feldman, M.P., Feldman & Gertler, M.S. (eds) *The Oxford Handbook of Economic Geography*, Oxford: Oxford University Press.
- Sheppard, E. & Barnes, T.J. (1990) *The Capitalist Space Economy: Geographical Analysis After Ricardo, Marx and Sraffa*, London: Unwin Hyman.
- Simon, H. A. (1979a) Rational decision making in business organizations, *American Economic Review*, 69(4), 493-513.

- Simon, H.A. (1979b) On parsimonious explanation of production relations, *Scandinavian Journal of Economics*, 81(4), 459-474.
- Simon, J. L. (1986) *Theory of Population and Economic Growth*, Oxford: Basil Blackwell.
- Steedman, I. (2003) On 'measuring' knowledge in new (endogenous) growth theory in Salavadori, N. (ed.) *Old and New Growth Theories. An assessment*, Cheltenham, Edward Elgar 127-133.
- Simon, H. A. & Levy, F.K. (1963) A note on the Cobb-Douglas function, *Review of Economic Studies*, 30(2), 93-96.
- Solow, R.M. (1957) Technical change and the aggregate production function, *Review of Economics and Statistics*, 39(3) 312-320.
- Solow, R.M. (1988) Growth theory and after, *American Economic Review*, 78(3), 307-317.
- Solow, R.M. (2000) *Growth Theory: An Exposition*, Oxford: Oxford University Press (first edition, 1970).
- Storper, M. (2011) From retro to avant-garde: a commentary on Paul Krugman's 'The new economic geography, now middle-aged'. *Regional Studies*, 45(1), 9-15.
- Tooze, M.J. (1968) Regional elasticities of substitution in the United Kingdom in 1968, *Urban Studies*, 13(1), 35-44.
- Tselios, V. (2009) Growth and convergence in income per capita and income inequality in the regions of the EU, *Spatial Economic Analysis*, 4(3), 343-370.
- World Bank (2009), *Reshaping Economic Geography*, World Development Report, Washington, DC: World Bank,.

Table 1 Estimates of a Cobb-Douglas Production Function, US Manufacturing, State Data, 1963

$$\ln V_i = c + \alpha \ln J_i + \beta \ln L_i + \varepsilon_i$$

<i>Ind. Code</i>		α	t	β	t	$v = \alpha + \beta$	t^*	\bar{R}^2
20	(i)	0.600	6.32	0.450	4.53	1.050	3.11	0.990
	(ii)	0.661	7.30	0.537	5.08	1.198	3.55	0.921
22	(i)	0.038	0.15	0.923	3.68	0.961	1.32	0.983
	(ii)	0.010	0.05	0.875	3.85	0.885	2.38	0.946
23	(i)	0.372	4.96	0.625	7.24	0.997	0.11	0.982
	(ii)	0.312	3.60	0.619	8.66	0.931	1.17	0.918
24	(i)	0.576	5.91	0.345	2.87	0.921	2.03	0.969
	(ii)	0.444	4.60	0.683	4.24	1.127	1.27	0.878
25	(i)	0.195	2.59	0.842	10.65	1.037	1.36	0.981
	(ii)	0.189	2.12	0.808	10.25	0.997	0.05	0.921
26	(i)	0.499	9.97	0.535	10.18	1.034	1.82	0.990
	(ii)	0.492	7.72	0.512	5.19	1.004	0.08	0.970
27	(i)	0.511	3.56	0.523	3.36	1.034	1.77	0.993
	(ii)	0.579	4.42	0.512	3.07	1.091	1.45	0.936
28	(i)	0.223	2.70	0.794	9.14	1.017	0.93	0.974
	(ii)	0.217	2.52	0.806	8.40	1.023	0.63	0.962
29	(i)	0.626	5.66	0.282	2.05	0.908	2.20	0.985
	(ii)	0.619	4.93	0.256	1.41	0.875	1.65	0.961
30	(i)	0.620	4.64	0.425	3.23	1.045	1.26	0.974
	(ii)	0.590	4.45	0.474	3.27	1.064	1.06	0.937
31	(i)	0.224	1.43	0.774	4.55	0.998	0.06	0.989
	(ii)	0.225	1.82	0.667	4.12	0.892	1.32	0.960
32	(i)	0.304	3.52	0.703	8.78	1.007	0.28	0.983
	(ii)	0.260	2.86	0.697	8.74	0.957	0.73	0.909
33	(i)	0.408	6.79	0.576	8.67	0.984	0.94	0.993
	(ii)	0.409	5.39	0.578	5.49	0.987	0.33	0.980

Table 1 (continued)

<i>Ind. Code</i>		α	t	β	t	$\nu = \alpha + \beta$	t^*	\bar{R}^2
34	(i)	0.363	3.60	0.637	6.24	1.000	0.07	0.993
	(ii)	0.356	3.47	0.628	5.99	0.984	0.32	0.910
35	(i)	0.200	1.69	0.828	6.66	1.028	1.74	0.992
	(ii)	0.263	2.19	0.731	5.36	0.994	0.13	0.925
36	(i)	0.337	2.93	0.708	6.00	1.045	2.02	0.986
	(ii)	0.334	2.70	0.704	4.76	1.038	0.62	0.908
37	(i)	0.225	1.49	0.831	4.48	1.056	1.06	0.952
	(ii)	0.242	1.88	0.892	5.31	1.134	1.78	0.899
38	(i)	0.015	0.07	1.091	4.14	1.106	2.03	0.977
	(ii)	0.192	0.79	0.913	2.80	1.105	0.76	0.854

Notes: SIC: 20 Food & Beverages; 22 Textiles; 23 Apparel; 24 Lumber; 25 Furniture; 26 Pulp & Paper; 27 Printing; 28 Chemicals; 29 Petroleum & Coal; 30 Rubber & Plastics; 31 Leather; 32 Stone Clay & Glass; 33 Primary Metals; 34 Fabricated Metals; 35 Non-Electrical Machinery; 36 Electrical Machinery; 37 Transportation Equipment; 38 Instruments.

- (i) State total data
- (ii) State per establishment data

t are the t -values, t^* is based on the null hypothesis that the coefficient is unity.

Source: US Annual Survey of Manufactures, Area Statistics.

¹ See Felipe and McCombie (2005a, 2012a). Simon and Levy (1962) first applied the critique to cross sectional data. Shaikh (1974, 1980) generalised the criticism to time-series estimations of supposed production functions and Simon (1979a) thought the problem so important he mentioned it in his Nobel prize speech. (See also Simon 1979b.) However, the problems have generally been ignored by economists who use the production function with the exception of Temple who fundamentally misunderstands the issues (see Felipe and McCombie (2012b) and the references cited therein).

² “Finding Waldo” is a children’s game where the character Waldo has to be found where he is hidden in a coloured picture with many other individuals. Solow (2000) uses this as an analogy for the *ad hoc* linear restriction necessary in many endogenous growth models to ensure a steady-state solution, but which is often merely implicit rather than explicitly stated.

³ Ideally, the production function should include land as a factor production. However, this is often excluded because of data availability and we shall abstract from this factor of production here.

⁴ It is sometimes argued that the critique does not hold because neoclassical production theory assumes perfect competition (i.e., there are no economic rents). This is erroneous for the reasons set out in Felipe and McCombie (2007) and, for reasons of space, will not be discussed here.

⁵ Times-series estimation of production functions sometimes give a poor fit to the data with the estimates of the output elasticities taking often implausible values. This is because proxying the weighted logarithm of the wage rate and the profit rate by a *linear* time trend (or their weighted growth rate by a constant) is misspecified as the wage rate and, especially, the rate of profit has a pronounced cyclical fluctuation. It is always possible to find a more flexible proxy that make the estimating equation equivalent to the accounting identity.

⁶ Although, as O’Donnell and Swales (1979) and Armstrong and Taylor (2000) point out, other factors such as the supply elasticities of capital and labour are also important.

⁷ Note that, as we have seen above, the accounting identity critique does not depend upon these stylized facts. If they do not hold, then we know immediately from the accounting identity that MRW's specification will give a poor statistical fit.

⁸ These are OECD, East Asia, South Asia, Sub-Saharan Africa, Western Hemisphere, Middle East and North Africa, and Europe.

⁹ The growth of total factor productivity was adjusted for the rate of change in human capital. Generally, this does not make a great deal of difference as the latter is proxied by a function of the growth of the change in *ratio* of the wage rates of the skilled and unskilled workers, which is generally relatively small, and could even be negative (Solow, 1988, p.313).

¹⁰ Estimating equation (33) directly generally finds the implicit estimate of ν not significantly different from zero, while estimating equation (32) gives ν approximately equal to one half. This is due to econometric reasons and will not be discussed here. Suffice it to say, equation (32) is generally seen as the preferred specification. It should be noted that the use of the factor shares as weights is, in the light of the accounting critique, somewhat of an *ad hoc* assumption. The neoclassical interpretation of them as reflecting output elasticities is invalid.

¹¹ See Hausmann *et al* (2007) for a theoretical and empirical analysis importance of exports in determining national growth rates and Minondo (2009) and Leichenko (2000) for regional examples.