

An Assessment of CES and Cobb-Douglas Production Functions ¹

Eric Miller

E-mail: eric.miller@cbo.gov

Congressional Budget Office

June 2008

2008-05

¹Working papers in this series are preliminary and are circulated to stimulate discussion and critical comment. These papers are not subject to CBO's formal review and editing processes. The analysis and conclusions expressed in them are those of the authors and should not be interpreted as those of the Congressional Budget Office. References in publications should be cleared with the author. Papers in this series can be obtained at www.cbo.gov (select Publications and then Working Papers). I would like to thank Bob Arnold, Juan Contreras, Bob Dennis, Matthew Goldberg, Angelo Mascaro, Delciana Winders, and Thomas Woodward for helpful comments.

Abstract

This paper surveys the empirical and theoretical literature on macroeconomic production functions and assesses whether the constant elasticity of substitution (CES) or the Cobb-Douglas specification is more appropriate for use in the CBO's macroeconomic forecasts. The Cobb-Douglas's major strengths are its ease of use and its seemingly good empirical fit across many data sets. Unfortunately, the Cobb-Douglas still fits the data well in cases where some of its fundamental assumptions are violated. This suggests that many empirical tests of the Cobb-Douglas are picking up a statistical artifact rather than an underlying production function. The CES has less restrictive assumptions about the interaction of capital and labor in production. However, econometric estimates of its elasticity parameter have produced inconsistent results. For the purpose of forecasting under current policies, there may not be a strong reason to prefer one form over the other; but for analysis of policies affecting factor returns, such as taxes on capital and labor income, the Cobb-Douglas specification may be too restrictive.

It takes something more than the usual “willing suspension of disbelief” to talk seriously of the aggregate production function.

(Robert Solow, 1957, p. 1)

1. Introduction

A macroeconomic production function is a mathematical expression that describes a systematic relationship between inputs and output in an economy, and the Cobb-Douglas and constant elasticity of substitution (CES) are two functions that have been used extensively. These functions play an important role in the economic forecasts and policy analysis of the CBO and others. For example, the CBO uses production functions to forecast potential output and the medium-term outlook for income shares. Additionally, several of the models that the CBO uses to analyze policy changes, such as the multisector Aiyagari (1994) model and the overlapping generations model, assume that aggregate output in the economy can be described by a production technology.¹

In all of these cases the CBO assumes that the economy’s underlying production technology takes the Cobb-Douglas form. This specification has long been popular among economists because it is easy to work with and can explain the stylized fact that income shares in the United States have been roughly constant during the postwar period. Economists have also been somewhat well disposed toward Cobb-Douglas because it gives simple closed-form solutions to many economic problems. However, empirical and theoretical work has often questioned the validity of the Cobb-Douglas as a model of the U.S. economy. Some economists believe that the more general CES may be a more

¹Although Cobb-Douglas does restrict the elasticity of substitution between the demand for labor and capital in production, it does not address labor-supply issues such as the substitution between labor and leisure, or the trade off between consumption spending and saving to increase the capital stock. These are issues which the CBO has examined extensively. For a further discussion see CBO (2008).

appropriate choice.

This paper will examine the major strengths and weaknesses of these two functional forms and suggest which form is better able to forecast income shares under given policies. This is important for budget estimation because one cannot predict the future path of the federal budget without first predicting how national income will be apportioned between capital and labor. These two sources of income are subject to different tax treatments and as a result their relative sizes are an important determinant of future tax revenues. While the focus of this paper is forecasting, many of the issues it raises are pertinent to policy analysis as well.

The paper begins with a review of some basic concepts in production theory. Readers with a solid understanding in this area should skip ahead to section 3, which discusses some general theoretical problems with the use of aggregate production functions. Section 4 reviews the empirical performance of the CES and Cobb-Douglas. The paper concludes with a recommendation to continue using the Cobb-Douglas in spite of theoretical concerns because it appears that additional costs and parameter uncertainties from the use of the CES are not outweighed by its benefits.

2. Production Functions

A production function is a heuristic device that describes the maximum output that can be produced from different combinations of inputs using a given technology. This can be expressed mathematically as a mapping $f : \mathbb{R}_+^N \rightarrow \mathbb{R}_+$ such that $Y = f(\mathbf{X})$, where \mathbf{X} is a vector of factor inputs $(X_1, X_1, \dots, X_n)'$ and $f(\mathbf{X})$ is the maximum output that can be produced for a given set of inputs $\mathbf{X}_i \in \mathbb{R}_+$.² This formulation is quite general and

²It is possible to have a production function that produces a vector of outputs, but in order to simplify the analysis this paper limits its scope to functions that map a vector of inputs to a single output.

can be applied at both microeconomic (i.e., individual firm) and macroeconomic (i.e., overall economy) levels. While production functions were originally designed with the individual firm in mind, macroeconomists came to realize that this methodology provides a useful tool for estimating certain parameters that cannot be directly measured from national accounts data. The most important of these is the elasticity of substitution between capital and labor.

Elasticity of substitution in production is a measure of how easy it is to shift between factor inputs, typically labor and capital. This measure is defined as the percentage change in factor proportions resulting from a one-unit change in the marginal rate of technical substitution (MRTS). MRTS is the rate at which labor can be substituted for capital while holding output constant along an isoquant; that is, it is the slope of the isoquant at a given point. Thus, for a two-input production function, $Y = f(K, L)$, the elasticity of substitution between capital and labor is given by

$$\sigma = \frac{\% \Delta(K/L)}{\% \Delta MRTS} = \left[\frac{d(K/L)}{dMRTS} \right] * \left[\frac{MRTS}{(K/L)} \right] = \frac{\partial \ln(K/L)}{\partial MRTS} \quad (1)$$

where σ can be thought of as an index that measures the rate at which diminishing marginal returns set in as one factor is increased relative to the other (Nelson 1964). When σ is low, changes in the MRTS lead to small changes in factor proportions. In the extreme case of fixed proportions or Leontief (1941) technology,

$$Y = \min(aK, bL) \text{ where } a, b > 0 \quad (2)$$

the resulting isoquants are L-shaped and $\sigma = 0$. This implies that changes in the MRTS will not cause any changes in factor proportions, so output is maximized by producing in fixed ratios. The other extreme is linear production technology,

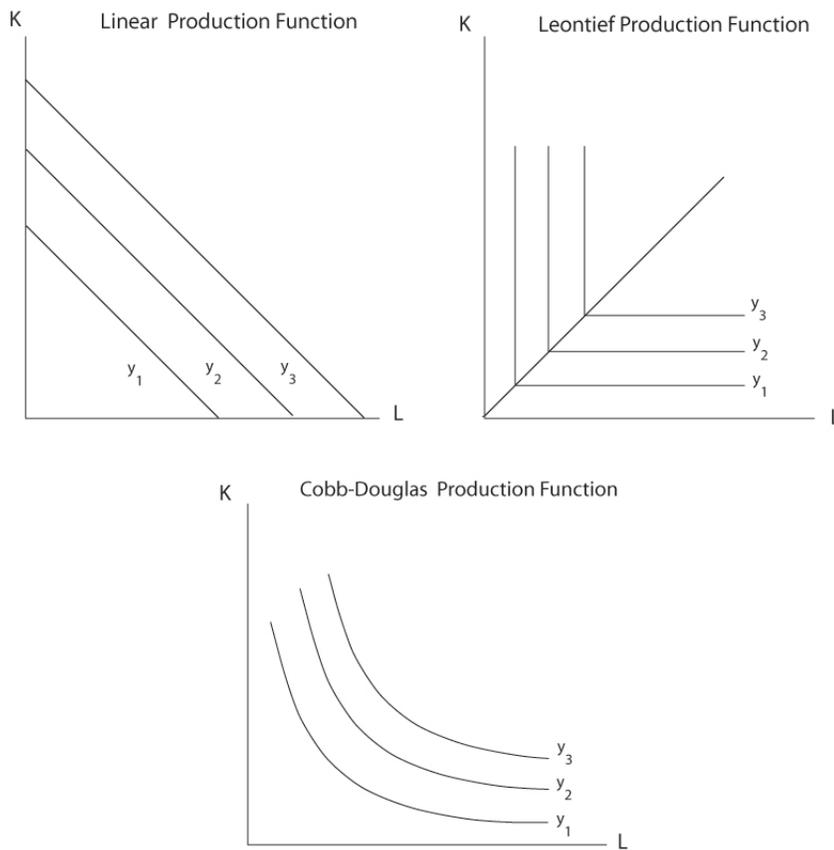
$$Y = aK + bL \tag{3}$$

where capital and labor are perfect substitutes. Here the MRTS is constant so the isoquants are straight lines and $\sigma = \infty$. The Cobb-Douglas form,

$$Y = AK^\alpha L^{1-\alpha} \tag{4}$$

lies between these two extremes, with $\sigma = 1$. This specification creates isoquants that are gently convex and easy to work with.

Figure 1. Isoquant Maps for Production Functions with Different Elasticities



Elasticities of substitution provide a powerful tool for answering analytical questions about the distribution of national income between capital and labor. If we assume that markets are competitive, then factors will be paid their marginal product. Hence, the wage rate will equal the marginal contribution from an additional worker and the return on capital will match the contribution in output that a marginal increment of capital provides. The elasticity of substitution can now be written as

$$\sigma = \frac{\% \Delta(K/L)}{\% \Delta(w/r)} = \frac{\partial \ln(K/L)}{\partial \ln(w/r)} \quad (5)$$

where w is the wage rate and r is the rental rate of capital. Differing values of σ have different implications for the distribution of income.

If $\sigma = 1$, any change in K/L will be matched by a proportional change in w/r and the relative income shares of capital and labor will stay constant. Any increase in the capital-labor ratio over time will be exactly matched by a percentage increase in the MRTS and an identical percentage increase in w/r . As a result, constant shares of output are allocated to capital and labor even though the capital-labor ratio may change over time. During the post-World War II period, the long-term trend in factor shares in the United States appears to have been roughly constant while the capital-labor ratio has been steadily increasing.³ This stylized fact initially led many economists to believe that $\sigma = 1$ in the United States.

If $\sigma > 1$, then a given percentage change in K/L will exceed the associated percentage change in w/r . For example, an increase in the capital stock would raise the ratio K/L but lower w/r by a smaller percentage, hence the share of capital in total income would rise as the capital-labor ratio increased. The opposite result occurs when $\sigma < 1$: an increase in the ratio K/L would tend to lower capital's share because the relative price

³While the overall labor share has been relatively constant during this period, there have been significant movements at the industry level (Solow (1958), Jones (2003), and Young (2006)).

of labor would rise in response to the increase in the amount of capital per worker.

The production functions that we have examined so far have been quite generic. In order to generate testable predictions, it is necessary to impose some further restrictions. Following Barro and Sala-i-Martin (2004), we define a production function of the form $Y = f(K, L, A)$, where K is capital, L is labor, and A is a measure of technology, as a neoclassical production function if the following three conditions are met:

1) *Constant returns to scale.* The function f exhibits constant returns to scale. If capital and labor are multiplied by a positive constant, λ , then the amount of output is also multiplied by λ :

$$f(\lambda K, \lambda L, A) = \lambda f(K, L, A) \text{ for all } \lambda > 0 \quad (6)$$

This definition includes only the two rival inputs, capital and labor; we did *not* define constant returns to scale as $f(\lambda K, \lambda L, \lambda A) = \lambda f(K, L, A)$.

2) *Positive and diminishing returns to private inputs.* For all $K > 0$ and $L > 0$, f exhibits positive and diminishing marginal products with respect to each input:

$$\begin{aligned} \frac{\partial f}{\partial K} > 0, \quad \frac{\partial^2 f}{\partial K^2} < 0 \\ \frac{\partial f}{\partial L} > 0, \quad \frac{\partial^2 f}{\partial L^2} < 0 \end{aligned} \quad (7)$$

If we hold constant the levels of technology and labor, then each additional unit of capital delivers additional output, but these additions decrease as the stock of capital rises. The same is true for labor.

3) *Inada conditions.* The marginal product of capital (labor) approaches infinity as

capital (labor) goes to zero and approaches zero as capital (labor) goes to infinity:

$$\lim_{K \rightarrow 0} \frac{\partial f}{\partial K} = \lim_{L \rightarrow 0} \frac{\partial f}{\partial L} = \infty$$
$$\lim_{K \rightarrow \infty} \frac{\partial f}{\partial K} = \lim_{L \rightarrow \infty} \frac{\partial f}{\partial L} = 0 \tag{8}$$

The two most popular neoclassical production functions are the Cobb-Douglas and the CES. The Cobb-Douglas is a simple production function that is thought to provide a reasonable description of actual economies. It was created by labor economist Paul H. Douglas and mathematician Charles W. Cobb in an effort to fit Douglas's empirical results for production, employment, and capital stock in U.S. manufacturing into a simple function (Cobb and Douglas 1928). This functional form has been extremely popular among economists because of its ease of use and its extreme flexibility. It can be written as

$$Y = AK^\alpha L^{1-\alpha} \tag{9}$$

where Y is output, K is capital, L is labor, α is a constant that takes values between zero and 1, and A is the level of technology ($A > 0$). It can be easily demonstrated that this function exhibits constant returns to scale and $\sigma = 1$. If we assume that markets are competitive and that factors are paid their marginal product, then α and $1 - \alpha$ are equal to capital and labor's share of output respectively. As mentioned earlier, an elasticity of substitution equal to unity implies that these factor shares will remain constant for any capital-labor ratio because any changes in factor proportions will be exactly offset by changes in the marginal productivities of the factor inputs. Thus, the observed income shares will be constant through time.

The Cobb-Douglas is a special case in a more general class of production functions

with constant elasticity of substitution.⁴ Following Klump and Preissler (2000), we can rewrite equation (1) as

$$\sigma = \left[\frac{d(K/L)}{dMRTS} \right] * \left[\frac{MRTS}{(K/L)} \right] = \frac{\tilde{f}'(k)[\tilde{f}(k) - k\tilde{f}'(k)]}{-k\tilde{f}''(k)\tilde{f}(k)} \quad (10)$$

This definition can then be transformed into a second-order partial differential equation in k with the solution

$$y = \frac{Y}{L} = \tilde{f}(k) = \gamma_1[k^\rho + \gamma_2]^{1/\rho} \quad (11)$$

or

$$Y = f(K, L) = \gamma_1[K^\rho + \gamma_2L^\rho]^{1/\rho} \quad (12)$$

where γ_1 and γ_2 are constants of integration and $\rho = (\sigma - 1)/\sigma$. If we set $\alpha = 1/(1 + \gamma_2)$ and $C = \gamma_1(1 + \gamma_2)^{1/\rho}$ then we arrive at the standard CES form introduced by Arrow, Chenery, Minhas, and Solow (1961) (hereinafter referred to as ACMS):

$$Y = C[\alpha K^\rho + (1 - \alpha)L^\rho]^{1/\rho} \quad (13)$$

where C is a measure of technical progress and the coefficients α and $1 - \alpha$ are distribution parameters between zero and 1 and can be used to determine factor shares. The substitution parameter ρ can be used to derive the elasticity of substitution σ .

This formulation of the CES function has been criticized as being unduly restrictive because it assumes that technological progress has no effect on the marginal productivities of input factors.⁵ If we set $A_K = \gamma_1$ and $A_L = \gamma_2\gamma_1^\rho$ we obtain the functional form

⁴The CES is a special case of an even more general class of production functions. For a further discussion see Christensen, Jorgenson and Lau (1973).

⁵Barro and Sala-i-Martin (2004) have proposed a CES variant $C[\alpha(bK)^\rho + (1 - \alpha)((1 - b)L)^\rho]^{1/\rho}$ that generalizes the ACMS function by allowing for biased technological change. Klump and Preissler (2000) show that this specification does not follow directly from the definition of σ ; instead, the authors suggest starting with equation (14) and then applying a suitable normalization procedure.

proposed by David and van de Klundert (1965):

$$Y = [(A_K K)^\rho + (A_L L)^\rho]^{1/\rho} \quad (14)$$

where A_K represents capital-augmenting technological change and A_L represents labor-augmenting technological change. From equation (14) it follows that

$$\text{Marginal Product of Capital} = \frac{\partial f}{\partial K} = A_k^{(\sigma-1/\sigma)} \left(\frac{Y}{K} \right)^{1/\sigma} \quad (15)$$

$$\text{Marginal Product of Labor} = \frac{\partial f}{\partial L} = A_L^{(\sigma-1/\sigma)} \left(\frac{Y}{L} \right)^{1/\sigma} \quad (16)$$

$$MRTS = \sigma > 0 \quad (17)$$

In the extreme case where $(\sigma - 1)/\sigma \rightarrow -\infty$, the elasticity of substitution tends toward zero (no substitution) and when $(\sigma - 1)/\sigma \rightarrow 1$ the elasticity of substitution tends toward infinity (perfect substitutes).

For any variant of the CES production function, the value of σ depends on three given baseline values: a given capital intensity ($k_0 = K_0/L_0$), a given marginal rate of substitution $[f_L/f_K]_0 = u_0$, and a given level of per capita production $y_0 = Y_0/L_0$. A set of CES functions are considered to be in the same “family” if they share the same baseline values but differ only in their values of σ and one point of tangency that is characterized by the given baseline values. Using Klump and de La Grandville’s (2000) normalization procedure, one can benchmark current data to some initial period and then assess how σ varies relative to different benchmark choices.⁶

⁶Klump and de La Grandville (2000) advocate the use of normalized production functions in empirical work because it makes for more consistent cross-study estimates of σ . For a discussion of how this normalization procedure can be carried out in empirical work, see Klump, McAdam, and Willman (2008).

Finally, we examine the role that production functions play in estimating labor's share of income. If we assume that markets are competitive and returns to scale are constant, then labor's share S_L can be modeled solely as a function of the capital-output ratio k^* . The relationship between S_L and k^* is unaltered by changes in factor prices, factor quantities, or the presence of labor-augmenting technical progress. When our production function is CES ($\sigma \neq 1$), then the relationship between S_L and k^* is monotonic. If labor and capital are substitutes ($\sigma > 1$), a lower capital intensity will increase the labor share; if they are complements ($\sigma < 1$), then the converse holds. In models that use more general production functions, the relationship between S_L and k^* need not be monotonic: the labor share can go up and then down as some variable driving changes in k^* (such as real wages or interest rates) varies.

3. Production Functions: Theoretical Problems

Reducing the U.S. economy to a simple mathematical function is bound to mask many of the complex interactions that characterize a real economy. However, if this stripped-down model can produce reasonably accurate predictions, then it can be a useful forecasting tool regardless of how well it describes reality.

However, once we move beyond simple forecasting and attempt to draw inferences from model parameters, internal consistency becomes much more important. All of the estimated parameters in the CES and Cobb-Douglas models have a behavioral interpretation. This means that if the model is not internally consistent, then these parameters are not describing a meaningful economic relationship. However, even if our model is misspecified and the parameters are in fact statistical artifacts, they may still be useful for forecasting purposes. The theoretical literature on aggregation shows that a macroeconomic production function has economic content only if a very stringent set of

conditions is met. Since these conditions aren't satisfied in real economies, it is likely that the good fit observed in empirical studies of aggregate production functions is the result of a statistical artifact.

Aggregate production functions rely heavily on the use of marginal products and factor elasticities, both of which are microeconomic concepts that macroeconomists have found very useful for simplifying their models. While it is common practice to estimate these parameters for capital and labor in the whole economy, it is not entirely clear that these measurements capture an economically meaningful relationship. It is worth remembering that a production function does not express a relation between inputs and *actual output*, but between combinations of inputs and *maximum potential output*. Thus, macroeconomic parameter estimates have an economic interpretation when the production possibility frontiers (PPF) for the individual firms can be combined to form an aggregate PPF for the entire economy. As we shall see, a very stringent set of conditions is required for meaningful aggregation across firms.

The first problem that arises is how to develop a sensible measure of aggregate capital. Following Jorgenson et al. (2000), capital enters our production function exogenously as a flow of services K , which is assumed to be proportional to the physical stock of capital.⁷ Many parameterized production function models describe the capital stock in dollar amounts rather than in physical quantities. This is problematic, because according to neoclassical theory, the rate of interest is determined by the overall amount of physical capital in the economy. Since the monetary value of a capital asset is influenced by the rate of return it generates over its productive lifetime, the model will suffer from endogeneity problems.⁸ One can get around this difficulty (to some extent) by using the BEA's quantity index, which measures changes in the quality-adjusted stock of physical

⁷The constant of proportionality accounts for changes in quality and treats new capital as an equivalent number of efficiency units of older capital.

⁸Joan Robinson popularized this line of argumentation during the Cambridge Capital Controversies. For a further discussion of this issue see Cohen and Harcourt (2003).

capital. However, it is important to keep in mind just how diverse U.S. capital stock really is. Everything from steel mills to palm pilots comes under this heading, and the Bureau of Economic Analysis is faced with the monumental task of boiling down this heterogeneity to a single index number.⁹ Additionally, there is some arbitrariness in the way that income is partitioned between capital and labor in the national accounts. If an individual is employed by a firm, her compensation comes in the form of a wage and counts as labor income. On the other hand, if this individual owns her own business where she performs the exact same work, her compensation is considered proprietors' income, a separate category in the national income and product accounts, and researchers are left to determine what part of proprietors' income to treat as labor income and what part capital income.¹⁰

Assuming that it is possible to create a meaningful measure of capital services in the economy, it remains to be seen whether aggregation of firms' PPF leads to a coherent PPF for the economy as a whole. The following argument draws heavily on Felipe and Fisher's (2003) survey of the theoretical literature on aggregation. The first major result presented is Leontief's theorem, which provides the necessary and sufficient conditions for aggregation of any twice differentiable production function. The theorem states that aggregation is possible if and only if the MRTS of the variables in the aggregate production function are independent of the variables that are not included. In an economy where each firm has a production function $f(K, L, A)$ and aggregation occurs across K and L , we can construct an aggregate production function F if and only if

$$\frac{\partial(f_K/f_L)}{\partial f_A} = 0 \tag{18}$$

⁹Similar problems (though not as severe) arise in the measurement of services from labor. While labor hours provide a convenient numeraire, it is clear that not all labor hours are created equal. For a discussion of how to adjust labor hours for differences in human capital see Ho and Jorgenson (1999).

¹⁰The CBO convention (CBO (2006)) has been to use the same average fraction as is found in the corporate sector. However, tax considerations may induce people to report their income in ways that shift the balance between capital and labor.

That is, aggregation is possible if and only if the marginal rate of substitution between K and L is independent of A . In the current example both capital and labor enter as arguments, so technology must be neutral with respect to both. If we construct the production function in per capita terms, then we need to aggregate only across capital. In this case, we can have a coherent production function when technological change is labor augmenting. Let us return for a moment to the problem of heterogeneous capital. If we have a production function $f = f(k_1, \dots, k_n, L)$, where each k_i represents a unique type of capital, then $f = F(K, L)$, where K is the aggregate stock of capital if and only

$$\frac{\partial}{\partial L} \left(\frac{\partial Q / \partial k_i}{\partial Q / \partial k_j} \right) = 0 \text{ for all } i \neq j \quad (19)$$

Thus, Leontief's theorem requires that labor has no effect on the substitution possibilities between the capital inputs. This makes intuitive sense because if labor has an effect on substitution possibilities, then it becomes impossible to collapse the capital stock to a single dimension. Clearly this condition is not satisfied in the real world, where the choice of capital is influenced by the quantity and quality of labor available.

Franklin Fisher expanded on this work and found that if one assumes that labor is efficiently allocated to firms, then an aggregate production function can be found to exist if and only if all of the micro units are identical except for their capital efficiency coefficient. When firms choose the amount of labor to hire, a labor aggregate exists only if there is no specialization in employment and aggregate output exists only if there is no specialization in production, i.e., firms all produce the same market basket of goods at different scales. What does all this mean for our aggregate production function? We give Felipe and Fisher (2008) the last word on this matter:

The conditions under which aggregate production functions exist are so stringent that real economies surely do not satisfy them. The aggregation results

pose insurmountable problems for theoretical and applied work in fields such as growth, labour or trade. They imply that intuitions based on micro variables and micro production functions will often be false when applied to aggregates.

Proponents of the neoclassical production function argue that despite these theoretical shortcomings, aggregate production functions can be defended on instrumentalist grounds if they provide a reasonably good description of the data. To test this proposition, Fisher (1971) simulated numerous fictitious simple economies where he knew the conditions for successful aggregation would be violated. He fit the Cobb-Douglas to the generated data and found that without exception it yielded an almost perfect fit. Fisher et al. (1977) ran a similar experiment and found that as the constancy of the labor share decreased, the Cobb-Douglas's fit decreased substantially. When the experiment was repeated using the CES, the authors were unable to predict the model's fit based on changes in the labor share.

Fisher's findings have been confirmed by subsequent experiments in which the requisite conditions for the existence of a production function were clearly violated. Shaikh (1974) showed that *any* production series Y , K would fit the Cobb-Douglas well, provided that factor shares are constant and capital and labor are uncorrelated with the rate of technological growth. To illustrate this point, he created a data set that spelled out the word HUMBUG when values of Y/L were plotted against values of K/L . Additionally, Shaikh chose his data so that it would produce the same capital share as Solow's (1957) famous paper on economic growth. Using Solow's (1957) procedure, Shaikh fit the Cobb-Douglas to his data and found an extremely high R^2 value.¹¹

Nelson and Winter (1982) used simulation analysis to generate a data set from their evolutionary model of the economy. The model makes the following assumptions: firms do not maximize profits, technology is available only in fixed coefficients, and firms

¹¹In a short note, Solow (1974) claims to have refuted Shaikh's study. However, Solow's reply completely ignores the problem of identification.

operate in conditions of disequilibrium. Surprisingly, the fitted Cobb-Douglas provided an excellent description of this data and generated elasticity of substitution parameters that were very close to the factors' share of revenue.

In each of these experiments, the researchers created a model economy in which neoclassical theory would provide a poor causal explanation for the observed data. Yet in all cases the Cobb-Douglas still produced a remarkably good fit. These findings show that the Cobb-Douglas functional form is flexible enough that it can fit the data well even when it does not have a meaningful economic interpretation.

By the principle of Occam's razor, if many theories can explain a set of empirical results, we should accept the explanation that requires the fewest assumptions. Indeed, the simplest explanation for the Cobb-Douglas' excellent empirical fit in many real world studies is the presence of an accounting identity that exists in any real economy, regardless of the underlying production technology. Observe the following equation:

$$Y(t) = W(t) + P(t) = wL(t) + rK(t) \tag{20}$$

where Y is real value added at time t , W is the wage bill, P is total payments to owners of capital, w is the real average wage rate, r is the real rate of return, L is the total labor employed, and K is the total capital stock. This accounting identity shows how output is broken down between wages and profit and does not rely on the standard neoclassical assumptions. From this definition we can use elementary calculus to construct a mathematical function that looks very much like the Cobb-Douglas. We follow Felipe and McCombie (2005a) and begin by differentiating equation (20) with respect to time and dividing through by total output to get

$$\frac{\dot{Y}}{Y} = \frac{wL}{Y} \left(\frac{\dot{L}}{L} \right) + \frac{wL}{Y} \left(\frac{\dot{w}}{w} \right) + \frac{rK}{Y} \left(\frac{\dot{K}}{K} \right) + \frac{rK}{Y} \left(\frac{\dot{r}}{r} \right) \tag{21}$$

For convenience, let $\alpha = wL/Y$, labor's share of output and let $(1 - \alpha) = rK/Y$, capital's share of output. Now, with the help of several stylized facts we can derive the Cobb-Douglas from equation (20). Assume that factor shares are constant and integrate equation (21). This gives us the function

$$Y = Bw^\alpha r^{1-\alpha} L^\alpha K^{1-\alpha} \quad (22)$$

If we further assume that the wage rate grows at a nearly constant rate and profit rate does not exhibit a secular trend ($\dot{r} \approx 0$), then $a\dot{w} + (1 - \alpha)\dot{r} \approx \alpha\dot{w} = \lambda$, which is a constant. Thus, we can rewrite equation (22) as the Cobb-Douglas production function with exogenous technical change:

$$Y(t) = Ae^{\lambda t} L^\alpha K^{1-\alpha} \quad (23)$$

This equivalence between the Cobb-Douglas and the accounting identity holds whenever factor shares are roughly constant and the correlation between the capital labor ratio is weakly correlated with changes in the technology parameter.¹² The above derivation was carried out without assuming that factors are paid their marginal product, that returns to scale are constant, or that production can be described by a particular technology. Felipe and Holz (2001) used Monte Carlo simulation to empirically test the conditions under which this link will hold. The pair found that the Cobb-Douglas is well described by the accounting identity even when there are large variations in the factor shares. However, the link between the Cobb-Douglas and the accounting identity begins to break down when there are variations in the growth rates of wages and the return to capital.

Felipe and McCombie (2001) examined ACMS's CES production function and found

¹²This derivation applies to time-series data. For a similar transformation of cross-sectional data see Felipe and McCombie (2005a).

that regression results that are purportedly well explained by a CES production function can be equally well explained by an underlying accounting identity. The original motivation behind the CES was the observation that in a given industry, value added per unit of labor varied across countries with the wage rate (and hence the elasticity of substitution is not equal to unity). To explain this fact, ACMS worked backward and hypothesized the following relationship:

$$\ln\left(\frac{Y}{L}\right) = \ln[(C^{1-\sigma})(\alpha^{-\sigma})] + \sigma \ln w \quad (24)$$

From equation (24) they were able to derive the CES production function. When equation (24) was fitted to the data it gave a very high R^2 and a value of σ below unity. However, it is highly plausible that the good fit of the model has nothing to do with an underlying production function and instead reflects the following identity, which comes directly from the definition of the labor share S_L , where

$$\frac{Y}{L} = \frac{w}{S_L} \quad (25)$$

If labor's share remains constant across units of observation then the model fit will be very good and estimates of σ will be close to unity. If labor's share and wages are negatively correlated then estimates of σ will be greater than unity, while positive correlation between wages and labor's share will yield an estimate of σ that is less than unity. While all of these outcomes can be explained by an underlying CES production function, they can be more easily explained by a simple accounting identity.¹³

¹³The above argument applied to cross-sectional data, but Felipe and McCombie (2001) show that a similar argument applies to time-series data.

4. CES and Cobb-Douglas: An Empirical Assessment

If income shares are forecast using a neoclassical production function, the elasticity of substitution parameter σ is of prime importance. When $\sigma > 1$, an increasing share of national income goes to capital as the capital-labor ratio increases. If $\sigma < 1$, then capital's share declines as this ratio increases. When $\sigma = 1$, income shares are unaffected by changes in the capital-labor ratio. Cobb-Douglas models assume that σ equals unity, but there is no theoretical reason why this must be the case. The correct choice of σ is a purely empirical question. Ever since the pathbreaking work of Charles Cobb and Paul Douglas in 1928, economists have been well disposed toward the Cobb-Douglas function because it can explain the stylized fact that income shares in the United States have been roughly constant over time while the capital-labor ratio has been rising steadily.

In their original 1928 study, Cobb and Douglas used time-series data to examine the relationship between capital, labor, and manufacturing output in the United States. The authors found a tight correspondence between their theoretical production function and the observed results. Over the next few decades, a number of studies confirmed Cobb and Douglas's initial findings and lent strong empirical support to their model.¹⁴ However, modern economists have questioned the methodological soundness of these early studies.

Fraser (2002) reexamined five of these original time series studies that claimed the Cobb-Douglas provided an excellent fit to the data. Collinearity diagnostics revealed that once a time trend was accounted for, the data in all five studies was highly collinear, which can lead to imprecise point estimates. Fraser also examined the time-series properties of the five data sets and found a mix of stationary and nonstationary variables.

¹⁴For an overview of these early studies see Douglas (1948).

In the presence of nonstationary variables standard Ordinary Least Squares estimates are dubious because they may lead to spurious regressions.¹⁵

After the seminal work of ACMS (1961), estimating elasticities using a fitted Cobb-Douglas largely fell out of favor.¹⁶ Early CES studies that used cross-sectional data tended to find elasticity values close to 1, while time series analyses tended to produce much lower estimates (Berndt 1976). More recent studies tend to eschew flat cross-sectional data which contain inherent biases (Lucas 1969) and instead focus on either panel or time-series data. Most of these studies find values of σ that are consistently below unity, but a great deal of variation in the results persists. Pereira (2003) surveyed major papers in the field from the past 40 years and found that, in general, elasticity values were below unity. Hamermesh (1993) surveyed a number of studies that used the CES specification and labor-hours data to estimate factor elasticities in the United States and found results ranging from 0.32 to 1.16. Again, in most cases the results were significantly below unity. A recent survey by Chirinko (2008) looked at modern studies of the elasticity parameter and found considerable variation in cross-study results. However, the weight of the evidence suggested a range of σ that is between 0.4 and 0.6, with the assumption of Cobb-Douglas being strongly rejected.

One problem with interpreting these cross-study results is that the various analyses are not all measuring the same thing. The CES production function has a number of different variants that can be tested using either cross-sectional or time-series data. As mentioned earlier, the value of σ is dependent on a given capital intensity, a given marginal rate of substitution, and a given level of per capita production. Klump and de La Grandville (2000) argue that cross-study results become far more meaningful when

¹⁵Felipe and Holz (2001) find that spuriousness makes only a minor contribution to the high R^2 in regressions that use a fitted Cobb-Douglas.

¹⁶The notable exception is Mankiw, Romer, and Weil (1992). The authors find that when international GDP growth is estimated using a Cobb-Douglas model with human capital, the fit is quite good. Felipe and McCombie (2005b) have criticized this result and argue that the good fit is due to an accounting identity rather than an underlying production function.

they are within a “family” of CES functions that differ only in their σ values.

Another reason why the CES results are not robust across data sets is because time-series estimates of σ are not well measured by least squares regression techniques. Kumar and Gapinski (1974) used Monte Carlo simulation to study the small-sample properties of CES estimators and found that the simulated parameters generally performed well except for estimates of σ , where estimates of the true parameters exhibited enormous variance and, in one case, bias on the order of 10^{109} . The authors concluded that these wild results are due to the flatness of the sum of squared error function near the true parameter values. These results were confirmed by Thursby (1980), who showed that parameter estimates of σ may not have an expected value and even when the first and second moments do exist, the variance of this estimator tends to be large.

Building on this work, Pereira (2003) used Monte Carlo simulation to test the reliability of econometric estimates of the CES substitution parameter ρ . The data were generated from Cobb-Douglas and CES production functions with various specifications (true parameters). When the underlying data was generated by a Cobb-Douglas production function, the simulated CES estimates consistently did a poor job of identifying the true parameter. When the true parameter was generated by a CES production function, the reliability of the simulated estimators varied greatly depending on the chosen model specification. The experiment was rerun but this time a Box-Cox transformation of the CES estimator was used to identify the true parameters.¹⁷ In nearly all cases, the transformed estimator produced results with less bias and lower variance.

Additionally, there is the issue of econometric identification in production function estimates. The fundamental idea is nicely captured in an older paper by Marschak and Andrews (1944, p. 144):

¹⁷A Box-Cox transformation can be used to normalize error terms and reduce heteroskedasticity. For a further discussion see Zarembka (1974).

Can the economist measure the effect of changing amounts of labor and capital on the firm's output-the "production function"- in the same way in which the agricultural research worker measures the effect of changing amounts of fertilizers on the plot's yield? He cannot because the man power and capital used by each firm is determined by the firm, not by the economist. This determination is expressed by a system of functional relationships; the production function, in which the economist happens to be interested, is but one of them.

This criticism can be restated in econometrics jargon as a violation of strict exogeneity assumptions. For unbiased coefficient estimates, OLS requires that the regressors be strictly exogenous: $(E[\epsilon_i|x_1, \dots, x_n]) = 0$, where ϵ_i is an error term and the x_i terms are regressors. However, many of the variables that influence firms' choice of capital and labor are unknown to the econometrician, but nevertheless exhibit a systematic pattern. Studies that simply regress output on the factors of production are subject to Marschak and Andrews' critique and are likely to violate this assumption and produce biased estimates. This problem may be overcome using two-stage least squares or a fixed-effects estimator, but such techniques were uncommon in early studies when these methods were still in their infancy.

Another source of inconsistency between studies may come from misspecifications of technical progress. The vast majority of production function studies make the simplifying assumption that technological progress is Hicks-neutral, i.e., it does not change the marginal products of capital or labor for a given ratio of inputs. Given this assumption, one can show that Cobb-Douglas is the only functional form that is capable of explaining the U.S. experience of constant factor shares and a rising capital-labor ratio (Antras 2004). This is because Cobb-Douglas is the only functional form in which Hicks-neutral technical change can be equivalently expressed as labor-augmenting technical change and only the latter can deliver the theoretical result of a constant labor share that has approximately been seen in the data. In contrast, empirical studies such as Greenwood, Hercowitz, and Krusell (1997), Krusell et al. (2000), and Dupuy (2006)) strongly sup-

port the notion that capital-augmenting technology has a significant effect on the U.S. economy. However, as Acemoglu (2003) has shown, such technology can change labor income's long-run share from one level to another, provided that it does not dominate labor-augmenting technical progress in the long run. Thus, functional forms other than Cobb-Douglas may be able to explain the presence of constant factor shares alongside a rising capital-labor ratio.

Choosing the correct functional form in the presence of biased technological change is challenging because it is impossible to determine whether the evolution of factor shares and factor ratios over time is due to elasticity effects or technological bias unless a priori assumptions about the structure of technological change are imposed (Diamond, McFadden, and Rodriguez (1978)). Thus, in order to generate tractable results, economists are forced to either make very strong assumptions about how technological change affects the overall economy or assume that technological change is Hicks-neutral. If these assumptions turn out to be unwarranted, this will result in model misspecification and the resulting bias may be large. Antras (2004) found that in econometric models where technological change is assumed to be Hicks-neutral, estimates of the elasticity of substitution between capital and labor are biased toward unity. If the economy's underlying technology turns out to be labor or capital augmenting, then the values of the technological parameters A_K and A_L cannot be estimated and the resulting omitted variable bias pushes the estimated elasticity toward unity. If one assumes that A_K and A_L both grow at constant rates, then estimates of σ are significantly below unity.¹⁸

Finally, there is the issue of short-run versus long-run analysis. A recent paper by Charles Jones (2003) presents readers with four stylized facts:

1. Growth rates in U.S. per capita GDP have not shown a major trend for the last

¹⁸Given the difficulty of directly measuring the rate of technological progress in the economy, this assumption may be equally unwarranted. Additionally, this specification should produce an upward or downward trend in labor's share.

125 years.

2. The capital share shows a significant trend in many countries and in many U.S. industries over time.
3. Elasticity of substitution estimates are often below unity.
4. The price of capital goods in the “equipment” category e.g. computers, machine tools, has been falling relative to the price of nondurable consumption goods, where the falling price is taken to indicate that technological progress is being embodied in capital goods at a faster rate than in consumption goods.

Taken together, these results are somewhat puzzling because a well-known theorem in growth theory states that if a neoclassical growth model possesses a steady state with positive growth and a growing capital share, either technological change must be labor-augmenting or the production function must be Cobb-Douglas. Jones (2003) believes this result can be reconciled with the above facts if production functions are Cobb-Douglas in the long run, but take some other form in the short run. While this hypothesis seems plausible, it is undermined by a recent survey by Chirinko (2008), who finds that in most studies where σ is estimated by techniques that explicitly focus on the long-run position of the economy, the value of σ is still significantly below unity.

5. Conclusion

After surveying the literature on macroeconomic production functions, there does not appear to be overwhelming evidence that would lead one to choose the CES over the Cobb-Douglas for forecasting GDP and income shares. When empirical estimates are restricted to the Cobb-Douglas form, the fit tends to be quite good. However, there are strong theoretical reasons why this fit is likely due to an accounting identity rather

than an underlying production function. To some extent, the CES specification gets around this problem because it allows for more degrees of freedom. However, it still suffers from serious identification problems. The generality of the CES is also a liability for forecasting purposes because it is difficult to get an estimate of σ that is entirely consistent across studies.

Aggregation issues pose a serious theoretical challenge to both functional forms and one should be aware of these problems in order to avoid making strong inferences that cannot be justified. In spite of this, the Cobb-Douglas has in the past been able to produce reasonably accurate long-term economic forecasts. Thus, despite the strong theoretical case against the continued use of Cobb-Douglas, it does not appear that using a CES framework would significantly improve the CBO's ability to forecast GDP and income shares. For analysis of policies affecting factor returns, such as taxes on capital and labor income, however, the Cobb-Douglas specification is too restrictive.

References

- Acemoglu, D. 2003. "Labor- and Capital-Augmenting Technical Change." Journal of the European Economic Association, Vol. 1, pp. 1-37.
- Aiyagari, R. 1994. "Uninsured Idiosyncratic Risk and Aggregate Saving." Quarterly Journal of Economics, Vol. 109, pp. 659-684.
- Antras, P. 2004. "Is the U.S. Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution." B.E. Journal of Macroeconomics, Vol. 4,1, Article 4.
- Arrow, K; Chenery, H; Minhas, B; and Solow, R. 1961. "Capital-Labor Substitution and Economic Efficiency." Review of Economics and Statistics, Vol. 63, pp. 225-250.
- Barro, R. and Sala-i-Martin, X. 2004. Economic Growth. Cambridge: MIT Press.
- Basu, S. and Fernald, J. 1997. "Returns to Scale in U.S. Production: Estimates and Implications." Journal of Political Economy, Vol. 105, pp. 249-283.
- Bentolila, S. and Saint-Paul, G. 2003. "Explaining Movements in the Labor Share." B.E. Journal of Macroeconomics, Vol. 3,1, Article 9.
- Berndt, E. 1976. "Reconciling Alternative Estimates of the Elasticity of Substitution." Review of Economics and Statistics, Vol. 58, pp. 59-68.
- Box, G. and Cox, D. 1964. "An Analysis of Transformations." Journal of the Royal Statistical Society, Vol. 26, pp. 211-252.
- Chirinko, R. 2008. " σ : The Long and Short of It." Journal of Macroeconomics, Vol. 30, pp. 671-686.
- Christensen, L. Jorgenson, D. and Lau, L. 1973. "Transcendental Logarithmic Production Frontiers." Review of Economics and Statistics, Vol. 55, pp. 28-45.

- Cobb, C. and Douglas, P. 1928. "A Theory of Production." American Economic Review, Vol. 18, pp. 139-250.
- Cohen, A. and Harcourt G. 2003. "Whatever Happened to the Cambridge Capital Controversies?" Journal of Economic Perspectives, Vol. 17, pp. 199-214.
- Congressional Budget Office. 2001. "CBO's Method for Estimating Potential Output: An Update."
- . 2006. "How CBO Forecasts Income."
- . 2008. "An Analysis of the President's Budgetary Proposals for Fiscal Year 2009."
- David, P. and van de Klundert, T. 1965. "Biased Efficiency Growth and Capital-Labor Substitution in the U.S., 1899-1960." American Economic Review, Vol. 55, pp. 357-394.
- Diamond, P. McFadden, D. and Rodriguez, M. 1978. "Measurement of the Elasticity of Factor Substitution and Bias of Technical Change." In Production Economics: A Dual Approach to Theory and Applications. Fuss, M. and McFadden, D. (ed.). Amsterdam:North-Holland.
- Douglas, P. 1948. "Are There Laws of Production?" American Economic Review, Vol. 38, pp. 1-41.
- Dupuy, A. 2006. "Hicks Neutral Technical Change Revisited: CES Production Function and Information of General Order." B.E. Journal of Macroeconomics, Vol. 6, 2, Article 2.
- Felipe, J. and Fisher, F. 2003. "Aggregation in Production Functions: What Applied Economists Should Know." Metroeconomica, Vol. 54, pp. 208-262.

- . 2008. “Aggregation (Production).” In The New Palgrave: A Dictionary of Economics, Durlauf, S. and Bloom, L. (eds.). London: Macmillan Reference Ltd.
- Felipe, J. and Holz, C. 2001. “Why do Aggregate Production Functions Work? Fisher’s Simulations, Shaikh’s Identity and Some New Results.” International Review of Applied Economics, Vol. 15, 3.
- Felipe, J. and McCombie, J. 2001. “The CES Production Function, the Accounting Identity and Occam’s Razor.” Applied Economics, Vol. 33, pp. 1221-1232.
- . 2005a. “How Sound Are the Foundations of the Aggregate Production Function?” Eastern Economic Journal, Vol. 31, No. 3.
- . 2005b. “Why are Some Countries Richer than Others? A Skeptical View of Mankiw-Romer-Weil’s Test of the Neoclassical Growth Model.” Metroeconomica, Vol. 65, pp. 360-392.
- Fisher, F. 1971. “Aggregate Production Functions and the Explanation of Wages: A Simulation Experiment.” Review of Economics and Statistics, Vol. 53, pp. 305-325.
- Fisher, F. Solow, R. and Kearn, J. 1977. “Aggregate Production Functions: Some CES Experiments.” Review of Economic Studies, Vol. 44, pp. 305-320.
- Fraser, I. 2002. “The Cobb-Douglas Production Function: An Antipodean Defence?” Economic Issues, Vol. 7, Part 1.
- Greenwood, J. Hercowitz, Z. and Krusell, P. 1997. “Long-Run Implications of Investment-Specific Technological Change.” American Economic Review, Vol. 87, pp. 342-362.
- Griliches, Z. and Mairesse, J. 1995. “Production Functions: The Search for Identification.” In Econometrics and Economic Theory in the 20th Century: The

- Ragnar Frisch Centennial Symposium, Strom, S. (ed.). Cambridge: Cambridge University Press.
- Hamermesh, D. 1993. Labor Demand. Princeton: Princeton University Press.
- Hayashi, F. 2000. Econometrics. Princeton: Princeton University Press.
- Ho, M. and Jorgenson, D. 1999. "The Quality of the U.S. Work Force, 1948-95." Harvard University, mimeo.
- Jones, C. 2003. "Growth, Capital Shares and a New Perspective on Production Functions." University of California, Berkeley, mimeo.
- Jorgenson, D. Stiroh, K. Gordon, R. and Sichel, D. 2000. "Raising the Speed Limit: U.S. Economic Growth in the Information Age." Brookings Papers on Economic Activity, Vol. 2000, pp. 135-235.
- Klump, R. and de La Grandville, O. 2000. "Economic Growth and the Elasticity of Substitution: Two Theorems and Some Suggestions." American Economic Review, Vol. 90, pp. 282-291.
- Klump, R; McAdam, P and Willman, A. 2008. "Unwrapping Some Euro Area Growth Puzzles: Factor Substitution, Productivity and Unemployment." Journal of Macroeconomics, Vol. 30, pp. 645-666.
- Klump, R. and Preissler, H. 2000. "CES Production Functions and Economic Growth." Scandinavian Journal of Economics, Vol. 102, pp. 41-56.
- Krusell, P. Ohanian, L. Rios-Rull, V. and Violante, G. 2000. "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis." Econometrica, Vol. 68, pp. 1029-1053.

- Kumar, K. and Gapinski, J. 1974. "Nonlinear Estimation of the CES Production Parameters: A Monte Carlo Study." Review of Economics and Statistics, Vol. 56, pp. 563-567.
- Leontief, W. 1941. The Structure of the American Economy, 1919-1939. New York: M.E. Sharpe Inc.
- Lucas, R. 1969. "Labor-Capital Substitution in U.S. Manufacturing." In The Taxation of Income from Capital, Harberger, A. and Bailey, M. (ed.). Washington D.C.: The Brookings Institution, pp. 223-274.
- Mankiw, N; Romer, D and Weil, D. 1992. "A Contribution to the Empirics of Economic Growth." Quarterly Journal of Economics, Vol. 107, pp. 407-437.
- Marschak, J. and Andrews, W. 1944. "Random Simultaneous Equations and the Theory of Production." Econometrica, Vol. 12, pp. 143-205.
- Nelson, R. 1964. "Aggregate Production Functions and Medium Range Growth Projections." American Economic Review, Vol. 54, pp. 575-606.
- Nelson, R and Winter, S. 1982. An Evolutionary Theory of Economic Change. Cambridge: Belknap Press.
- Nicholson, W. 2002. Microeconomic Theory. New York: Thomson Learning.
- Pereira, C. 2003. "Empirical Essays on the Elasticity of Substitution, Technical Change, and Economic Growth." Ph.D. dissertation, North Carolina State University.
- Shaikh, A. 1974. "Laws of Production and Laws of Algebra: The Humbug Production Function." Review of Economics and Statistics, Vol. 56, pp. 115-120.
- Simon, H. 1979. "On Parsimonious Explanations of Production Functions." Scandinavian Journal of Economics, Vol. 81, pp. 459-474.

- Solow, R. 1957. "Technical Change and the Aggregate Production Function." Review of Economics and Statistics, Vol. 39, pp. 312-320.
- . 1958. "A Skeptical Note on the Constancy of Relative Shares." American Economic Review, Vol. 48, pp. 618-631.
- . 1974. "Laws of Production and Laws of Algebra: The Humbug Production Function: A Comment." Review of Economics and Statistics, Vol. 56, p. 121.
- Stiroh, K. 1998. "Long-Run Growth Projections and the Aggregate Production Function: A Survey of Models Used by the U.S. Government." Contemporary Economic Policy, Vol. 16, pp. 467-479.
- Thursby, J. 1980. "Alternative CES Estimation Techniques." Review of Economics and Statistics, Vol. 62, pp. 292-295.
- Young, A. 2006. "One of the Things We Know That Ain't So: Why U.S. Labor's Share Is Not Relatively Stable." University of Mississippi, mimeo.
- Zarembka, P. 1974. "Transformation on Variables In Econometrics." in Frontiers in Econometrics, P. Zarembka (ed.). New York: Academic Press, pp. 81-104.