ARGUING A CASE FOR THE COBB-DOUGLAS PRODUCTION FUNCTION

K.V. Bhanumurthy*

Abstract

The paper tries to argue that Cobb-Douglas (CD) production function merits its use for analysing production process, not because it is looked as a simple tool which can be handled easily or is looked as a crude remedy for estimation ills, but because of the advantages it possesses. These advantages are due to the fact that it can handle multiple inputs in its generalised form. Even in the face of imperfections in the market it does not introduce distortions of its own. Unconstrained CD-function further increases its potentialities to handle different scales of production. Various econometric estimation problems, such as serial correlation, hetroscidasticity and multicolinearity can be handled adequately and easily. It is argued that most of its criticism is focussed on its inflexibility and admits that except for one obvious assumptions all other assumptions can be relaxed. It is further argued that it facilitates computations and has the properties of explicit representability, uniformity, parsimony and flexibility. Even the problem of simultaneity can be accounted for through the use of stochastic CD-production function. The paper argues that the technology can also be well represented by it.

INTRODUCTION

The Cobb-Douglas Production (CD) Function is often seen as an easy panacea for analysing production process or is dubbed as a simplistic tool based on restrictive conditions. The former attitude results in inappropriate use and the latter in its being relegated to an undeserved back seat. A critical analysis of its

* Reader, Department of Commerce, University of Delhi, Delhi School of Economics, Delhi.

merits and demerits is, hence, called for. The common criticisms about the restriction Cobb-Douglas production function forms the opening ground for such a review.

1. It cannot handle a large number of inputs.
2. The function is based on restrictive assumptions of perfect competition in the factor and product markets.
3. It assumes constant returns to scale (CRS).
4. Serial correlation and heteroscedasticity are common problems which beset this function too.
5. Labour and capital, are correlated and the estimates are bound to be biased.
6. Unitary elasticity of substitution is unrealistic.
7. It is inflexible in form.
8. Single equation estimates are bound to be inconsistent.
9. Other criticisms relate to the level of aggregation and nature of technology.
10. It cannot measure technical efficiency levels and growth very effectively.

The Cobb-Douglas production function is given as

$$Q = A L^{\alpha} K^{\beta}$$

where,

L = real value of labour input, K = real value of capital input, Q = real value added in output, α = output elasticity w.r.t. to labour, and β = output elasticity w.r.t. to capital

Log-linearizing and adding a stochastic term we have

$$\text{Log } Q = \text{Log } A + \alpha \text{ Log } L + \beta \text{ Log } K + u_{it}$$

In the event of pooling the time series data, belonging to two different sectors, for estimating a pooled Cobb-Douglas production function, the error term would consist of three components

$$u_{it} = \mu_i + \iota_t + \varepsilon_{it}$$

where,

U_{it} = total error term, μ_i = individual effect, and ε_{it} = random effect, ι_t = time effect.

1. Generalized Cobb-Douglas (CD) production function is very much capable of handling multiple inputs. It can be represented as

$$A = AF_1^{\alpha 1} F_2^{\alpha 2} F_3^{\alpha 3} ... F_n^{\alpha n}$$

where, F_1, F_2, F_3,.... F_n are n inputs,

α_1, α_2, α_3 ... α_n, are their elasticities of output with respect to inputs.

One aspect of the choice of a function on the basis of its ability to handle a number of inputs is the resultant measure of technical progress. Baily (1986) has shown that,

If, $Q = A (t) F[K, L, N]$

$G_a = G_Q - \alpha_l.G_l - \alpha_k.G_k - \alpha_n.G_n$

$G_b = G_{VA} - \alpha_l.G_1 - \alpha_k.G_k$

where

Q = Real Output, N = Material Input, L = Labour Input, K = Capital Input, G_Q = Growth in Output, G_{VA} = Growth in Value Added, G_1 = Growth in Labour, G_k = Growth in Capital, G_n = Growth in Material Inpout, α_1 = share of Labour in Output, α_k = Share of Capital in Output, α_l = Share of Labour V.A., and

if n = N/Q and d(n)/dt = 0

the two measures G_a and G_b are proportional, in time series estimates, and are actually equal in cross sectional estimates. This result holds, for all functional forms of Cobb-Douglas, CES and Translog production function. Thus, a two-factor CD function, under such assumptions, is equally efficient in measuring total factor productivity (TFP) growth.

2. Before going into the implications of the perfect competition assumption it is necessary to spell-out certain features of the function (see Hebden (1988)).

In general, the marginal productivity of the jth factor is

$$\delta Q_i / \delta f_j = \alpha_j (Q_i / F_j)$$

i.e. Marginal Productivity of factor F_j

= α_j times average productivity of factor j.

In a two-input case (K, L)

$$\delta L / \delta L = \alpha (Q / L) \delta Q / \delta K = \beta (Q / K)$$

For profit maximizing, optimum employment of factor inputs, the condition is that the ratio of the value of their marginal products be equal to the ratio of their marginal costs. If ‘P’ is the price per unit of the output (assuming perfect competition in the product market), w is the real wage rate and ‘r’ the return on capital (both marginal costs) (value of MP(L)/ value of MP(K)) = (MC of L/MC of K)

or $(P.\alpha. (Q/L) / P. \beta/ (Q/K)) = (w/r)$

so $(\alpha.K/\beta.L) = w/r$

$\alpha/\beta = w.L/r.K$

That is, the ratio of the exponents of the CD function represents the ratio of the factor payments, irrespective of the nature of competition in the factor markets.

Since $(\alpha/(\alpha + \beta)) = w.\,L/\,(w.L + r.K)) = w.L\,/\,P.Q$ and

$(\beta/(\alpha + \beta)) = (r.K/(w.L + r.K)) = r.K/P.Q.$

If perfect competition prevails, factors are paid the value of marginal product (VMP) i.e.

$P.\ \alpha.\ (Q/L) = w$; and $P.\beta.\ (Q/K) = r$

Thus, the exponents of the function equal the factor shares

$\alpha = w.L\,/\,P.Q;$ and $\beta = r.K\,/\,P.Q$

and as long as the factors are paid according to the ratio of their contribution, the assumption of perfect competition is not essential to the construction of the CD function. This can be generalized as a $\alpha_j/\Sigma\alpha_j$ for n inputs.

The question is, what happens if the product market has imperfections ? Once again, it does not introduce any distortion on its own, as far as the factor-price ratio is concerned. But if the individual wage rate is determined, it would contain an element of monopoly price because the value of marginal product would be greater than the marginal revenue product. This issue is being taken up later.

3. The marginal condition for efficient resource employment is converted into an exact condition for wage determination, if constant returns to scale obtain.

$(w.L\,/\,P.Q) = \alpha\,/(\alpha + \beta)$

$MR\ (L) = P.\ MP(L) = P.\ (\delta\ Q/\delta\ L) = P.\alpha.(Q/L)$ and since

MC(L) = w, MR(L)/MC(L) = P.α. Q/L. 1/w = α. (PQ/w) which equals

= α. (($\alpha + \beta$) / α) = $\alpha + \beta$

Similarly for capital we have,

MR(K)/MC(K) = $\alpha + \beta$

Therefore, for each factor,

MC/MR = $\alpha + \beta$ = returns to scale

But if returns to scale is constant, MC = MR (i.e. profit maximizes).

However, it is possible to estimate a CD function without imposing the CRS condition. Such a function is an unconstrained function. The validity of the constraint can be tested with an F-test. If a time variable is introduced in such a function it becomes a Tinbergen function, which is capable of measuring the rate of technical efficiency growth per annum (TFPG).

4. The problem of serial correlation is relevant to time series estimations. But, it is not an insolvable statistical problem. Firstly, the DW statistic is a fairly good measure of auto-correlation. Secondly, this issue pertains to any estimation involving time series data. Standard methods are available for appropriately treating such series – such ARIMA models.

As for heteroscedasticity, the problem arises usually in cross-sectional estimates, especially, if the data is ordered according to classes. In the case of annual time series it is rare. There is a simple solution of prior testing. If the residuals are plotted against Log Q and if it has any significant trend, it is indicative of heteroscedasticity. Moreover, it does not affect the bias of the estimators, but affects only their efficiency.

5. The problem that is often either exaggerated, or is uncritically glossed over, is that of multicollinearity. As is known, it is a problem only when it is extreme in degree. Firstly, the logic that K and L are substitutes and are bound to be correlated in a Cobb-Douglas production function is erroneous. The inherent converse relationship is true only if output remains constant, that is, along the same production isoquant. This is not possible in time series data. The stronger argument is that if industries with the same K/L are grouped for an estimation, it introduces intercorrelation. For instance, if the defined technology is K/L = 3, it implies K = 3L which indicates a high degree of collinearity. The alternative is to take dissimilar industry such that K/L varies. This would ensure better estimates, but would be fitted across a heterogeneous group of industries. The three aspects of this problem are prior assessment, implications and solutions.

The method for prior assessment is to apply the Klien's rule. In a regression equation of y on x_1, x_2 and x_3 and correlation of R^2 $x_1.$ x_2x_3 must be less than the overall R^2 $y.x_1x_2x_3$ of the estimated equation. In a CD function it simply means checking the correlation between L and K and comparing it with that of the R^2 of the estimated function. The implication of a high degree of collinearity is that the estimated βs will have high variances and so would result in unreliable estimates of β s. The easiest check is to ensure that the estimates have t-values that are highly significant, at least at the 1 percent level. It must be added that the use of panel regression is in itself a solution because really it is the cross-sectional estimate that is beset with such a problem.

6. The elasticity of the CD function is given by

$$\sigma = [d\,(K/L) / (K/L)] / [d\,(\delta L/\delta K) / (\delta L / \delta K)]$$

and is assumed to be constant and equal to unity throughout. Even this need not be assumed a priori. There exists a method whereby it is possible to make a prior assessment about the nature of the elasticity. Here, the following relationship can be estimated before attempting a production function estimation.

Log Q/L = $\alpha + \sigma$ Log w; where w = real wage rate, Q/L = labour productivity and σ = elasticity of substitution.

Now, if the coefficient of Log w is subjected to either a t-test for the hypothesis that $\sigma = 1$, or if a restriction is placed and an F-test is conducted it would either confirm or deny the existence of $\sigma = 1$. Secondly, as it will be shown later even this test can be improved upon by assigning arbitrary values, and by doing so, arriving at a true measure of the elasticity. Finally, Klien's viewpoint is that since all production function estimations are approximations, if elasticity is near about unity, a CD production function can be estimated. This important aspect is taken up again in the next part.

7. The most serious criticism is that the Cobb-Douglas production function is inflexible in its functional form.

Lau (1986) has described,

> "Flexibility of functional form is desirable because it allows the data the opportunity to provide information about critical parameters. An inflexible functional form often prescribes the values, or at least the range of values, of critical parameters (which should ideally be) free to attain any set of theoretically consistent values." (pp.1540)

He has specified a set of criteria which relates to the functional form. These are enlisted below. The relative merit of the CD function can be verified from them.

The conditions imposed by a restricted CD function are :

(i) $\alpha > 0, \beta > 0$, (ii) $\sigma = 1$,

(iii) $\alpha + \beta = 1$ and (iv) $P_k > 0, P_L > 0$.

Except (iv), which is essential for obvious reasons the rest of the restrictions need not be imposed a priori.

(a) Computational Facility : Here, the requirement is that the parameters must be linear. If necessary a non-linear form can be log-linearized. But the parameters must not be non-linear in the estimated final form even if the original form be non-linear. The CD function satisfies this requirement.

(b) Explicit Representability : The function itself and any function derived from it, must be represented in explicit closed form. It is not enough that the production function alone is explicitly representable, even the input demand function must be so and must also preferably be linear in parameters. In this case a profit or cost functions scores over it, since in their case it is automatically ensured.

(c) Uniformity : If the function belongs to a complete system, different functions of the same system must have the same algebraic form, but different parameters. The same procedure and programme can be used for all different functions. The CD function satisfies this condition.

(d) Parsimony : This requires that the number of parameters be minimized. Here, the CD function scores heavily since CD has only three parameters to be estimated. The main contender – the translog (TL) cost function has certain problems. Firstly, TL requires around 25 to 30 data points. A CD function can be fitted with very few data points because it has few parameters, and yet gives good results. Secondly, this long time period involves two problems. One, the estimates of TFP growth can vary greatly if such a long period is chosen. Two, it may mean that during the period some structural change has taken place and if it is so the estimate of TFP growth loses relevance.

(e) Flexibility : Has been defined above. The CD function is inflexible, insofar as it specifies a unitary elasticity condition. Here, Lau comments, "of course if it is known a priori whether the elasticity of substitution is closer to 1 or zero the choice is easier to make" (p.1539). The opinion of Klien (1974) is important in this respect.

"In my own research on empirical estimates of production functions would lead me to favour an elasticity of substitution somewhat less than unity in many lines of economic activity : nevertheless the CD form may be a tolerably good approximation (p.44)".

8. The simultaneous context is often ignored in the estimation of a CD function. The most obvious way out is to resort to simultaneous estimation. However, if the interest lies in obtaining consistent estimators for only the production function, this can be done by ensuring certain prior conditions.

Kmenta, Zellner and Dreze (1966) have proposed a method for a stochastic CD production function in a complete system, which gives consistent results.

Let $X = AL^{\alpha_1} K^{\alpha_2} e^{u_{it}}$

$\pi = p.X - w.L - r.K$ be the profit function

The profit maximizing condition $\delta \pi / \delta L = \delta \pi / \delta K = 0 \Rightarrow$ $X/L = (w/\alpha_1.p)^{-\frac{1}{2}\sigma u_{1i}}$ and $X/K = (r/\alpha_2.p)^{-\frac{1}{2}\sigma u_{1i}}$. Another source of randomness occurs because of the differences in between the realized and actual prices. This results in disturbances u_{2i} and u_{3i}. The assumption is that the prices are either known with certainty or are independent of the production function disturbance.

Let $y_{1i} = \log X_i$, $y_{2i} = \log L_i$, $y_{31} = \log K_i$

$y_{1i} - \alpha_1.y_{2i} - \alpha_2.y_{3i} = \iota_1 + u_{1i}$

$y_{1i} - y_{2i} = \iota_2 + u_{2i} + u_{1i}$

$y_{1i} - y_{3i} = \iota_3 + u_{3i} + u_{1i}$

where ι_1, ι_2 and ι_3 stand for log A, log $(w/\alpha_1.p)$ and log $(r/\alpha_2.p)$. And u_{1i}, u_{2i}, u_{3i} are error terms.

$y_{1i} = k_1 + [1/1-\alpha_1 - \alpha_2]. [(1-\alpha_1 - \alpha_2) u_{1i} - \alpha_1.u_{2i} - \alpha_2.u_{3i}]$

$y_{2i} = k_2 + [1/1-\alpha_1 - \alpha_2]. [(\alpha_2 - 1) u_{2i} - \alpha_2.u_{3i} - \alpha_2.u_{3i}]$

$y_{3i} = k_3 + [1/1-\alpha_1 - \alpha_2]. [(-\alpha_1.u_{2i}) + (\alpha_1 - 1).u_{3i}]$

Here, the Cov $(y_{2i}, u_{1i}) = 0$ and Cov $(y_{3i}, u_{1i}) = 0$, because y_{2i} and y_{3i} do not depend on u_{1i} but on u_{2i} and u_{3i}.

Now, if it is assumed that u_{1i} is uncorrelated with u_{2i} and u_{3i} since the former depends on acts of nature like weather, and the latter depends on human error, then the estimates of α_1, α_2 and 'A', belonging to the production function will be consistent and unbiased. They would also be asymptotically efficient.

In the earlier Cobb-Douglas estimates used by Marschak and Andrews (1944), the estimates of ι_2 and ι_3 included systematic errors, R_1 and R_2. The random disturbance u_{2i} and u_{3i} were related to the profit maximizing behavioural equations. On the other hand, u_{1i*} is a random, neutral, disembodied, productivity differential when α_1 and α_2 are common to all individuals being considered. The shift parameter A however varies as $A_i = A_0.e^{u1i*}$, A_0 being a common parameter and the e^{u1i} being the random element. Along with the purely random element, the production function would have two error components which cannot be estimated from a single set of data. Firstly, this calls for pooling of estimates and secondly, it is likely to yield inconsistent estimates of the parameters. In Zellner, Kmenta and Dreze's (1966) method, $R_1 = R_2 = 1$.

9. In the case of Cobb-Douglas production function pooling can be resorted to for countering the problems stated above. However, firstly, Cov $(X_{jt}, U_{jt}) = 0$ can be tested for eliminating the simultaneous equation bias. Secondly, aggregation itself introduces two kinds of problems. One, that of the likely aggregation bias due to cross correlation amongst individuals, that might occur on account of the presence of a time error. The other, due to the bias because of the violation of the structure of the error term of the classical ordinary least square (OLS) model. The first problem can be handled by testing for Cov $(U_{j1}, U_{j2}) = 0$. That is, by testing for cross correlations amongst the residual of the production equations of different individuals. The second problem has been taken care of by arriving at the Generalized Least Square (GLS) estimates.

Another issue related to aggregation is that of the level of aggregation and the problem of homogeneity of the underlying technology. The alternative is to choose a level of aggregation that is low enough to permit a commonness of the K/L ratio. Since the analysis is based on time series data it does not involve the problem of collinearity. But on the other hand, it does not result in a meaningful analysis at the overall level in terms of over all policy. Similarly, cross sectional analysis cannot reveal much about long-term policy changes and is likely to be biased by immediate influences, though it has the advantage of being in long-term equilibrium. Therefore, the choice falls upon a level of aggregation that is not beset with statistical problems, reflects something about overall policy and is based on average technology 'types' of different sectors. Finally, as Klien (1974) has stated in the work quoted above,

> "Good evidence has been accumulated over the years to suggests that technology on a macroscopic level can pretty well (be) described by the Cobb-Douglas production function ... we recognised that technology could be plausibly explained by this function and that it fitted well with .. a whole system". (pp.44)

Thus, for the aggregate level, the clear candidate is the Cobb-Douglas production function.

10. The assumption in production function estimation is that long-run equilibrium prevails. However, recent literature such as the exposition by Hulten (1986) points out that temporary equilibrium is possible. The effect of temporary equilibrium for the measurement of Technical progress assuming u-shaped cost curves is discussed hereafter.

Given the production function in the form

$$Q(t) = A(t).\ F[K(t), L(t)]$$

where,

A, F, K and L have usual meanings and t is the time point at which the variable is being measured.

Apart from other things that are known, A(t) denotes the efficiency parameter that allows for Hicks-neutral technical progress. The following relationship gives the function for growth in output, with (.) indicating the growth dimension

$$(\dot{Q}/Q) = \dot{A}/A + E_k(\dot{K}/K) + E_L(\dot{L}/L) \qquad \text{(i)}$$

and when the elasticities of capital and labour are given as

$$E_K = (\delta Q/\delta K).(K/Q) \qquad E_L = (\delta Q/\delta K).(L/Q)$$

$$W_K = (P_k.K/P.Q), \qquad W_L = (P_L.L/P.Q),$$

such that for a CD function $E_K = W_K = \beta$ and $E_L = W_L = \alpha$.

Under Constant Returns to Scale (CRS) $W_K + W_L = 1$. But often independent estimates of P_K are used in which the CRS assumption is not required. This is true of a Translog function where P_K is as follows :

$$P_K = (R_t + D_t)/P$$

It is the sum of the real rate of interest (R_t) (or the opportunity cost of using capital, i.e., the return on long-term government capital bonds) and the rate of depreciation (D_t).

As pointed out, usually the assumption of long-run equilibrium ensures that $P_K = Z_K$, where

$$Z_K = P.(\delta Q/\delta K) \neq P_K$$

This discrepancy arises on account of the quasi-fixed factor, capital, earning a rental because of which Z_K exceeds P_K, when optimal capacity is surpassed i.e.,

$U = Q_a/Q_o > 1$ where Q_a is actual production and Q_o is optimal capacity respectively.

If U < 1, it falls short of the optimum $Z_K < P_K$. While it equals P_K when U = 1.

Now if the growth equation is re-estimated with V_K and V_L and with capital stock replacing value of capital services,

$$(\dot{Q}/Q) = (\dot{A}^a/A^a) + V_K(\dot{K}/K) + V_L(\dot{L}/L) \quad \text{(ii)}$$

$\dot{A}^a/A^a$ = true TFP measure.

where

$$V_K = (Z_K.K/P.Q), \quad V_L = (P_L.L/P.Q)$$

if, $S = P_K.K + P_L$

where, S = unit cost at Qo

If instead of V_K and V_L, W_K and W_L are used, then

$W_K + W_L \neq 1$

This can b justified on grounds of non-CRS homogeneity of degree P.Q./S. The other alternnative is to use as weights

$W_K = (P_K.K/S)$, $W_L = (P_L.L/S)$

whose sum will be equal to one, but will yield another false index of TFP growth. It would be noticed that $V_L \neq W'_L$ unless P.Q = S.

$$(\dot{Q}/Q) = \dot{A}'A' + W_K(\dot{K}/K) + W_L(\dot{L}/L) \quad \text{(iii)}$$

Subtracting (iii) from (ii) and rearranging we have

$$\dot{A}'/A' = \dot{A}^a/A^a + (V_K - W'_K)(\dot{K}/K) + (V_L - W'_L)(\dot{L}/L) \quad \text{(iv)}$$

Therefore, the only true TFP index is that which is based on V_L and V_K. When U > 1 then $V_K > W_K$ and the false residual obtained from (i), over-estimates the TFP growth. Similar is the case with the false index obtained from (iii). Conversely, if $U < 1$ the false measures underestimate the TFP growth.

Thus, if an unconstrained Tinbergen function is used it may apparently appear simplistic, but it has the merit of measuring the TFP while making appropriate correction for capacity-utilisation and the capital rentals, since the measure of capital is the product of Z_K and K_s, which weights capital stock by the correct capital price.

Even the Translog cost function which has the most favourable functional properties, and is unrestrictive in its assumption, would yield similar results if the weight of capital input is W_K, which is usually the case. Moreover, estimating a translog cost function for such a purpose would be futile, because it can be shown that the residual derived from a cost function (which is only the dual of the production function) is precisely the reverse of that estimated from the product function. The other choice, of abandoning a production function in favour of cost function, is equally futile because, given the fact that elasticity of substitution is unity, the Translog Cost function also collapses to a Cobb-Douglas cost function (Diwert, (1974)) which is nothing but the dual of the CD production function.

REFERENCES

Baily, M.N. (1986). "Productivity Growth of Material Use." *Quarterly Journal of Economics*. VCI, Issue 1, pp.185.

Diwert, W.E. (1974). "Duality Between Unit Cost and Production Functions.' In Intriligator, M.D. and Kendrick, D.A. eds., *Frontiers of Quantitative Economics*. V.2. North Holland, New York, pp.115.

Hulten, C.R. (1986). "Productivity Change, Capacity Utilization and the Source of Efficiency Growth." *Journal of Econometrics*. V.33, n.1-5, pp.31-49.

Klein, L.R. (1974). "Issues in Econometric Studies of Investment Behaviour." *Journal of Economic Literature*. Pp.43-50.

Krishna, K.L. (1987). "Industrial growth and productivity in India." In Panchmukhi, V.R. and Brahmananda, P.R. eds., *The Development Process of the Indian Economy*. Himalaya, New Delhi.

Lau, L. (1986). "Functional Forms of Econometric Model Building." In Griliches, Z. And Intriligator, M.D. eds., *Handbook of Econometrics*, V.3, pp.1513-1566.

Marschak, J. and Andrews, W.J. (1944). "Random Simultaneous Equations and the Theory of Production." *Econometrica*. V. 12, pp.143-204.

Zellner, A., Kmenta, J. and Dreze, J. (1966). "Specification and Estimation of Cobb-Douglas Production Function Models." *Econometrica*. October, pp.784-795.

Manuscript Received April, 2002

Accepted June, 2002