*Agricultural Economics Research Review*
Vol. 21 July-December 2008 pp 289-292

*Research Note*

# Fitting of Cobb-Douglas Production Functions: Revisited

**Prajneshu***
Division of Biometrics, Indian Agricultural Statistics Research Institute (ICAR), New Delhi - 110 012

**Abstract**

The set of Cobb-Douglas production functions is usually fitted by first linearizing the models through logarithmic transformation and then applying method of least squares. However, this procedure is valid only when the underlying assumption of multiplicative error-terms is justified. Unfortunately, this assumption is rarely satisfied in practice and accordingly, the results obtained are of doubtful nature. Further, nonlinear estimation procedures generally yield parameter estimates exhibiting extremely high correlations, implying thereby that the parameters are not estimated independently. In this paper, use of *expected-value parameters* has been highlighted and the advantages of their use have also been discussed. Finally, the developed methodology has been illustrated by applying it to the wheat yield time-series data of Punjab.

## Introduction

A large number of research papers (see e.g. Anupama *et al.*, 2005; Mandal *et al.*, 2005; Mruthyunjaya *et al.*, 2005; Pouchepparadjou *et al.*, 2005; Shaheen and Shiyani, 2005; Srinivas and Ramanathan, 2005) dealing with Cobb-Douglas production functions published in the area of agricultural economics is a testimony to the important role played by these models. The model, in its simplest form, when there is only one explanatory variable ($U$) and one response variable ($Y$), is given by Equation (1):

$$Y = aU^b \qquad \ldots(1)$$

where, $a$ is scale parameter and $b$ is a measure of curvature. To estimate the parameters, the usual procedure is to assume a multiplicative error exp($\varepsilon$) in Equation (1) so that the model may be linearized by means of logarithmic transformation, giving Equation (2):

$$\ln(Y) = \ln(a) + b\ln(U) + \varepsilon \qquad \ldots(2)$$

This equation is then fitted to data using "method of least squares" and goodness of fit is assessed by computing coefficient of determination $R^2$.

Main drawback in this procedure is that a proper justification of assumption of multiplicative error is hardly ever provided and this assumption is usually made only for mathematical convenience. As pointed out by Ratkowsky (1990), the assumption tends to be valid only when variability of response variable $Y$ increases with increasing values of explanatory variable $U$, which happens very rarely. Further, one frequent mistake occurs when goodness of fit of even the original nonlinear model given by Equation (1) is assessed by reporting the same value of $R^2$ as has been obtained for the linearized model given by Equation (2). In fact, as discussed in detail by Prajneshu and Chandran (2005), it is not at all possible to get $R^2$ for Equation (1) from that for Equation (2).

Accordingly, in this paper, procedure that should be followed for fitting Cobb-Douglas production functions has been discussed. An illustration for wheat yield time-series data of Punjab state has also been presented.

*Email: prajneshu@yahoo.co.in
The author is grateful to Dr P. Kumar, formerly Professor, Division of Agricultural Economics, I.A.R.I., New Delhi-110 012 for providing the data. Thanks are also due to the referee for his valuable comments.

## Suggested Procedure

In order to apply Equation (1) to data, an additive error-term, assumed to be independently and identically distributed, is added on the right hand side of this equation, thereby yielding corresponding "statistical model". It may be noted that parameters ($a$, $b$) appear in a nonlinear manner. Thus, nonlinear estimation procedures, like 'Levenberg–Marquardt algorithm', or 'Does not Use Derivatives (DUD)' procedure are required to be employed for fitting the models to data. A good description of these procedures is given in Draper and Smith (1998). Fortunately, most of the software packages, like SPSS, SAS, SPLUS and GENSTAT contain computer programs to accomplish the task. Subsequently, residual analysis may be carried out by employing "Run test" (Gujarati, 2004) to examine validity of assumption of independence of errors. Finally, goodness of fit of fitted models may be examined by computing mean square error (MSE).

However, quite often, when above procedure is followed, it is found that the parameter estimates have very large correlations, implying thereby that the parameters have not been estimated independently. As the scale parameter $a$ is related to shape of the curve determined by the curvature parameter $b$, any change in the latter requires a corresponding change in the former. Therefore, the likelihood contours are bound to be extremely elongated ellipsoids. Hence, it is very important to reduce correlations between parameter estimates, as it is these that make the algorithms for "nonlinear estimation" inaccurate or fail to converge.

One way out for the above problem is through 're-parameterization' by using 'expected-value parameters' (Ratkowsky, 1990). These parameters correspond to the fitted (predicted) values of the response variable $Y$. The only restriction on expected-value parameters is that they should fall within the observed range of data.

### Finding Expected-Value Parameters

To find the expected-value parameters corresponding to $k$ parameters, $k$ values of explanatory variable $U$ are chosen. The new parameters are the expected values, denoted by $y_1$, $y_2, \ldots, y_k$, after the original parameters are eliminated. For the two-parameter model given by Equation (1), the first step is to choose values $U_1$ and $U_2$ of the explanatory variable $U$; they may respectively be minimum and maximum observed values of $U$ in the data set. Then, $y_1$ and $y_2$ may be obtained by solving the two equations, viz. Equations (3) and (4):

$$y_1 = a U_1^b \qquad \ldots(3)$$

$$y_2 = a U_2^b \qquad \ldots(4)$$

Solving these equations, we get

$$a = y_1 \ / \ U_1^{\ \ln(y_1 / y_2) / \ln (U_1/U_2)} \qquad \ldots(5)$$

and

$$b = \ln (y_1/y_2) \ / \ \ln (U_1/U_2) \qquad \ldots(6)$$

Substituting these values of $a$ and $b$ in original Equation (1), one gets:

$$Y = y_1 \ (U/U_1)^{\ln(y_1/y_2) \ / \ \ln(U_1/U_2)} \qquad \ldots(7)$$

which may be rewritten as Equation (8):

$$Y = y_1 \ (y_1/y_2)^{\ln(U/U_1) \ / \ \ln(U_1/U_2)} \qquad \ldots(8)$$

It may be easily verified that when $U = U_i$ , the value of $\hat{Y} = \hat{y}_i$, and $i$ = 1, 2 . This procedure has eliminated the original parameters $a$ and $b$ in favour of new parameters $y_1$ and $y_2$. Equation (8) appears to be more cumbersome in appearance than the original Equation (1). However, it has three advantages. First, rapid convergence is ensured as the new model is close-to-linear. Second, initial parameter estimates can easily be obtained. Third, the expected-value parameters are more suitable for inference as resultant estimators are close to being unbiased, normally distributed, minimum variance estimators.

The above procedure can be extended in a straightforward manner when there are two or more explanatory variables. For the case of two explanatory variables $U$ and $V$, Cobb-Douglas production function is given by Equation (9):

$$Y = a U^{b_1} V^{b_2} \qquad \ldots(9)$$

Choose three values ($u_i$ , $v_i$ ), $i$ = 1,2,3, of the two explanatory variables $U$ and $V$; the first pair of values may be in the beginning, second in the middle, and third towards the end of the values of $U$ and $V$ in the data set. Corresponding expected–value parameters $y_i$ , $i$ = 1,2,3, satisfy the Equation (10):

$$y_i = a\, u_i^{b_1}\, v_i^{b_2},\ i = 1, 2, 3 \qquad \ldots(10)$$

On solving these equations, final expressions for original parameters in terms of expected-value parameters $y_i$ , $i$=1,2,3, are obtained as:

$$a = y_1 \,/\, (u_1^{b_1}\, v_1^{b_2}),$$

$$b_1 = [\ln (y_1/y_2) - b_2 \ln (v_1/v_2)] \,/\, \ln (u_1 \,/\, u_2)$$

$$b_2 = [\ln (u_2/u_3) \ln (y_1/y_2) - \ln (u_1/u_2) \ln (y_2/y_3)] \,/\, [\ln (u_2/u_3) \ln (v_1/v_2) - \ln (u_1/u_2) \ln (v_2/v_3) \qquad \ldots(11)$$

Evidently, the models with more than two explanatory variables can be handled in a similar manner but the expressions become more cumbersome.

## An Illustration

As an illustration, annual time-series data for wheat for the state of Punjab for the period 1971-2000 was considered. The response variable ($Y$) was 'Yield' (quintals/ha), while explanatory variable ($U$) was 'quantity of fertilizer' (kg/ha). The SAS, Ver. 9.1 software package was used for the entire data analysis carried out in this paper. In the first instance, Equation (2) was fitted to data using method of least squares. It was noticed that the assumption of multiplicative errors was not justified for this data set as variability of response variable $Y$ did not increase with increasing values of explanatory variable $U$. Subsequently, Equation (1) was fitted to data through 'nonlinear estimation procedures'. However, magnitude of correlation coefficient between parameter estimates, ignoring sign, was found to be as high as 0.999, thus necessitating the need for re-parameterization.

Subsequently, Equation (8) was fitted to data using 'nonlinear estimation procedures'. To this end, the initial values of the parameters ($y_1$, $y_2$) were needed. $U_1$ was assumed to be some value towards the beginning of data set, say the fifth value, viz. 112.54 and so the initial value of parameter $y_1$ was the corresponding value of response variable, viz. 24.87. Similarly, $U_2$ was assumed to be some value towards the end of the data set, say the twenty-fifth value, viz. 212.26 and so the initial value of parameter $y_2$ was the corresponding value of response variable, viz. 42.46. Thus, the initial values of the parameters ($y_1$, $y_2$) were taken as (24.87, 42.46). On fitting nonlinear Equation (8) to data, it was found that the absolute value of correlation coefficient between parameter estimates was reduced to 0.234, implying thereby that the two expected-value parameters $y_1$ and $y_2$ had been estimated almost independently. Further, residual analysis indicated that the assumption of independence of error-terms was not rejected at 5 per cent level, as the calculated value of $|Z|$ using Run test, computed as 0.088, was not greater than the tabulated value of 1.96. Parameter estimates of $y_1$ and $y_2$ along with their standard errors (within the brackets) were obtained as follows:

$$\hat{y}_1 = 23.881(1.230),\ \hat{y}_2 = 39.587(1.057) \quad \ldots(12)$$

It was noted that both the standard errors were much less in comparison with the respective estimates and so the parameters had been estimated efficiently. Finally, the value of $MSE$ = 15.099, not being very high, indicated that Cobb-Douglas production function with single explanatory variable provided a reasonably good fit to the data. The final model was obtained as Equation (13):

$$Y = 23.881\ (0.603)^{1.576\ \ln(112.54/U)} \qquad \ldots(13)$$

As indicated earlier, it is better to use the above parameterization. However, for academic interest, the equivalent form of Equation (13), on using Equations (5) and (6), was obtained as Equation (14):

$$Y = 0.555U^{0.797} \qquad \ldots(14)$$

Attempts were then made to fit Cobb-Douglas production function with two explanatory variables. The second explanatory variable, viz. $V$ represented 'Human labour' (hours). When Equation (9) was fitted through 'nonlinear estimation procedures', the correlation coefficients between various parameter estimates were computed as follows:

$$r(\hat{a}, \hat{b}_1) = -0.921,\ \ r(\hat{a}, \hat{b}_2) = -0.965,$$

$$r(\hat{b}_1, \hat{b}_2) = -0.786, \qquad \ldots(15)$$

Large values of the magnitude of above correlation coefficients reflect the need for re-parameterization.

Equation (10), which is nonlinear, was then fitted to data using three expected-value parameters. To this end, the values of $(U_i, V_i)$, $i = 1,2,3$ were respectively taken as (109.53, 450.46), (168.35, 356.86), (217.84, 301.15) and so the initial estimates of parameters $(y_1, y_2, y_3)$ were (22.60, 25.20, 48.34). The correlation coefficients between various parameter estimates were then computed as:

$$r(\hat{y}_1, \hat{y}_2) = -0.206, \quad r(\hat{y}_1, \hat{y}_3) = -0.062,$$

$$r(\hat{y}_2, \hat{y}_3) = 0.065, \qquad \ldots(16)$$

All the above values were considerably reduced in magnitude. Residual analysis indicated that the assumption of independence of error-terms was not rejected at 5 per cent level, as the calculated value of $|Z|$, using Run test, was 1.301. Parameter estimates of $y_1$, $y_2$, and $y_3$ along with their standard errors (within the brackets) were computed as follows:

$$\hat{y}_1 = 24.334\ (1.324)$$

$$\hat{y}_2 = 33.904\ (0.858)$$

$$\hat{y}_3 = 41.849\ (1.325) \qquad \ldots(17)$$

Again, it was noted that all the standard errors were much smaller in comparison with the respective estimates and so the parameters had been estimated efficiently. The fitted model in terms of $(\hat{y}_1, \hat{y}_2, \hat{y}_3)$ can be written down in a straightforward manner but the expression is very cumbersome. However, for academic interest, its equivalent form, on using Equation (11), was Equation (18):

$$Y = 18.756\ U^{0.561}\ V^{0.389} \qquad \ldots(18)$$

Finally, the low value of $MSE$ = 3.797 implied that Cobb-Douglas production function with two explanatory variables provided a good fit to the data.

## References


Anupama, J., Singh, R. P. and Kumar, Ranjit (2005) Technical efficiency in maize production in Madhya Pradesh: Estimation and implications. *Agricultural Economics Research Review*, **18**: 305-15.

Draper, N. R. and Smith, H. (1998) *Applied Regression Analysis,* 3rd Edn, New York, USA: John Wiley & Sons.

Gujarati, D. N. (2004) *Basic Econometrics*. New Delhi: Tata McGraw-Hill, p. 465-67.

Mandal, S., Datta, K. K., Dayal, Bhu, Minhas, P. S and Chauhan, C. P. S. (2005) Resource use efficiency in saline irrigated environment. *Indian Journal of Agricultural Economics*, **60**: 494-509.

Mruthyunjaya, Kumar, Sant, Rajashekarappa, M. T., Pandey, L. M., Ramanarao, S. V. and Narayan, Prem (2005) Efficiency in Indian edible oilseed sector: Analysis and implications. *Agricultural Economics Research Review*, **18**: 153-66.

Pouchepparadjou, A., Kumaravelu, P. and Achoth, L. (2005) An econometric analysis of green technology adoption in irrigated rice in Pondicherry union territory. *Indian Journal of Agricultural Economics*, **60**: 660-76.

Prajneshu and Chandran, K. P. (2005) Computation of compound growth rates in agriculture. *Agricultural Economics Research Review*, **18**: 317-24.

Ratkowsky, D. A. (1990) *Handbook of Nonlinear Regression Models*. New York, USA: Marcel Dekker, p. 12.

Shaheen, F. A. and Shiyani, R. L. (2005) Water-use efficiency and externality in the groundwater exploited and energy subsidised regime. *Indian Journal of Agricultural Economics*, **60**: 445-57.

Srinivas, T. and Ramanathan, S. (2005) A study on economic analysis of elephant foot yam production in India. *Agricultural Economics Research Review*, **18**: 241-52.