

## *Note*

# **A Methodological Note on a Stochastic Frontier Model for the Analysis of the Effects of Quality of Irrigation Water on Crop Yields**

GEORGE E. BATTESE

A stochastic frontier model is proposed for analysis of crop yields, which considers the effects of differing quality of irrigation water, in addition to different inputs and factors associated with technical inefficiency of production. The parameters of the production frontier involved are assumed to be a function of other variables, which measure the quality of the irrigation water.

## **1. INTRODUCTION**

In recent years in Pakistan, there has been concern about the quality of the irrigation water used by farmers on their crops. Frequently, canal and tubewell water are mixed, or used in sequence, in the irrigation of various crops in the Punjab, because the quality of the tubewell water has been of poor quality and has led to reduced crop yields and salinity of the soils. Research by soil scientists has led to the measurement and classification of soils, according to their chemical and structural characteristics. However, no empirical research work has been done on the analysis of farm-level data on the effects of poor-quality irrigation water on crop yields.

This paper outlines a possible model for analysis of crop yields, which considers the effects of inputs and farmer characteristics, together with the effects of the quality of irrigation water used. The approach involves a stochastic frontier production function model, in which the technical inefficiency effects are assumed to be a function of various observable variables. In addition, the parameters of the stochastic frontier function are assumed to be a function of variables, which are associated with the quality of the irrigation water used in the cropping season.

George E. Battese works at the Centre for Efficiency and Productivity Analysis, School of Economic Studies, University of New England, Armidale, NSW, Australia.

*Author's Note:* This paper arises from the author's three-month assignment at the University of Agriculture, Faisalabad, Pakistan (UAF), as part of a twinning agreement between the University of New England and the University of Agriculture, Faisalabad.

## **2. SAMPLING AND DATA**

In order to investigate the effects of varying quality of irrigation water on crop yields, it is necessary to obtain an appropriate sample of farms from a given population. Unless random sampling methods are used in obtaining the sample of farms, the analysis of the data obtained may be of no benefit in making inference to the whole population of farms of interest. In order to obtain a sample of some design (simple random, stratified, cluster, etc.) some information on the population must be available or obtained, so that the appropriate sample of farms can be selected. For example, if a stratified random sample is to be obtained, where several villages in a district comprise the strata, then lists of all farmers in each village must be obtained before the sampling can proceed. If unequal probability sampling is used in the selection of the sample farms, then these probabilities of selection should be used in the analysis of the sample data. Such analyses are not often taken into account in empirical analysis of input and output data in agricultural economics [see Fuller (1975)]. The analysis suggested below assumes that the sample farms obtained are selected by simple random sampling and so the weights applied to all observations are the same.

In order to be able to conduct a response function analysis, yields of crops on the sample farms must be collected, together with the relevant input data for each crop to be considered. In order to account for the possible effect of varying quality of irrigation water used by the farmers, quantitative measures for the quality of the water must also be obtained, involving laboratory analysis of samples of water. Further, data on variables, which are likely to influence the level of the technical inefficiency of production, are required, given that stochastic frontier production functions are to be estimated, rather than the traditional crop response functions. These data requirements imply that the survey procedures must be carefully designed, and carried out, for the project to result in appropriate data for a satisfactory empirical analysis to be conducted. Battese (1998) comments on these procedures.

## **3. ECONOMETRIC MODEL**

The model, which is proposed for the analysis of farm-level data, involves a stochastic frontier production function, in which the parameters of the production function are specified to be a function of the variables associated with the quality of the irrigation water. The quality of the irrigation water is not to be considered as an input variable, like land, labour, fertiliser, etc. However, the quality of the water applied in the cropping season may affect the responsiveness of a crop to the different factors of production involved. The model is presented in terms of a Cobb-Douglas production function. If data from a large number of sample farmers are available, then the parameters of a more flexible functional form, such as a translog

production function, may be estimated with reasonable levels of precision.

For purposes of exposition, the model is presented in terms of output of wheat, involving the three input variables, land, fertiliser and irrigation, as follows:

$$\log(Y_i) = \beta_{0i} + \beta_{1i} X_{1i} + \beta_{2i} X_{2i} + \beta_{3i} X_{3i} + V_i - U_i \quad \dots \quad \dots \quad \dots \quad (1)$$

where  $\log$  denotes the natural logarithm (base,  $e$ ); the subscript,  $i$ , denotes the  $i$ th farmer in the sample,  $i = 1, 2, \dots, n$ , and  $n$  is the number of farmers selected in the random sample;

$Y_i$  denotes the total yield of wheat for the  $i$ th farmer;

$X_{1i}$  denotes the logarithm of the land area under wheat;

$X_{2i}$  denotes the logarithm of the quantity of fertiliser applied;<sup>1</sup>

$X_{3i}$  denotes the logarithm of the quantity of irrigation water applied;<sup>2</sup>

$\beta_{ki}$   $K = 0, 1, 2, 3$ , are unknown parameters for the production function for the  $i$ th farmer (to be defined below in terms of other variables);

$V_{iS}$  are random errors associated with measurement errors in the yields of wheat reported, or the combined effects of input variables not included in the production function, where the  $V_{iS}$  are assumed to be independent and identically distributed  $N(0, \sigma_v^2)$ -random variables;

$U_{iS}$  are non-negative random variables, associated with the technical inefficiency of production of the farmers in the population, assumed to be independently distributed, such that the technical inefficiency effect for the  $i$ th farmer,  $U_i$ , is obtained by truncation (at zero) of the normal distribution with mean,  $\mu_i$ , and variance,  $\sigma^2$ , such that

$$\mu_i = \delta_0 + \delta_1 Z_{1i} + \delta_2 Z_{2i} + \delta_3 Z_{3i} + \delta_4 Z_{4i} \quad \dots \quad \dots \quad \dots \quad \dots \quad (2)$$

where  $Z_{1i}$ ,  $Z_{2i}$ ,  $Z_{3i}$  and  $Z_{4i}$  are values of explanatory variables for the technical inefficiency effects for the  $i$ th farmer, say, his age and education, logarithm of the total land operated by the farmer,<sup>3</sup> and the proportion of the tubewell water used in the irrigations over the past season(s).

<sup>1</sup>The quantity of fertiliser could be in terms of a particular fertiliser, such as urea, or in terms of NPK. If some farmers did not apply fertiliser, then a dummy variable should be included in the model to permit the intercepts to be different for farmers with positive and zero fertiliser, see Battese (1997).

<sup>2</sup>The quantity of irrigation water applied can be approximated by the number of irrigations *times* the area under the crop, given that the volume of water applied is effectively constant for each occasion of irrigating. If this is not the case then a different measure of the irrigation water would be needed.

<sup>3</sup>When the inputs (or a variable closely corresponding to them, as suggested here) are also involved as explanatory variables for the inefficiency effects, the stochastic frontier model is called a *non-neutral* model, as proposed by Huang and Liu (1994) and further considered by Battese and Broca (1997). These models have important bearing upon the estimation of the elasticity of the mean output with respect to an input variable, which is also an explanatory variable for the inefficiency effects.

The above stochastic frontier model is a development of the original model, proposed by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977). The inefficiency effects are modelled in terms of other variables, as suggested by Battese and Coelli (1995) in the context of panel data on sample firms. The model can be estimated using the computer programme, FRONTIER 4.1, written by Tim Coelli from the University of New England.<sup>4</sup>

The final component of the model, for considering for the effects of differing quality of irrigation water, is defined as follows:

$$\beta_{ki} = \beta_k + \beta_{k1} Q_{1i} + \beta_{k2} Q_{2i}, k = 0, 1, 2, 3, \dots \dots \dots \dots (3)$$

where  $Q_{1i}$  and  $Q_{2i}$  are the values of variables, which measures the quality of the irrigation water applied by the  $i$ th farmer. These measures may be a function of the levels of electrical conductivity (EC), sodium absorption ratio (SAR), or residual sodium carbonate (RSC), which are important measures by which soil and water quality are classified by soil scientists, e.g., see Ghafoor, Qadir and Qureshi (1991). Alternatively, the model in Equation (3) may include second-order terms involving one or more of these variables.<sup>5</sup>

The specification of the parameters of the frontier model, in terms of the quality of the irrigation water, implies that the frontier model to be estimated is obtained by substitution of Equation (3) into Equation (1), which yields the model

$$\begin{aligned} \log(Y_i) = & \beta_0 + \beta_{01}Q_{1i} + \beta_{02}Q_{2i} + \beta_1X_{1i} + \beta_2X_{2i} + \beta_3X_{3i} \\ & + \beta_{11}X_{1i} \times Q_{1i} + \beta_{12}X_{1i} \times Q_{2i} + \beta_{21}X_{2i} \times Q_{1i} + \beta_{22}X_{2i} \times Q_{2i} \\ & + \beta_{31}X_{3i} \times Q_{1i} + \beta_{32}X_{3i} \times Q_{2i} + V_i - U_i, \dots \dots \dots (4) \end{aligned}$$

Thus the stochastic frontier production function includes the quality variables for irrigation water as intercept shifters in the function, together with their interactions with the input variables, as additional explanatory variables in the production function. An implication of this model is that the elasticities of output with respect to the different inputs are a linear function of the quality variables for irrigation water.

In the empirical application of this model, there would be interest to see if the elasticities are independent of the quality of the irrigation water applied to the crop. This would involve testing if the coefficients of the interaction variables are simultaneously zero (i.e.,  $\beta_{k1} = \beta_{k2} = 0, k = 1,2,3$ ). Testing the null hypothesis, that the quality variables for irrigation water have no effects on the level of yields and the elasticities of output, would involve estimating the traditional Cobb-Douglas

<sup>4</sup>FRONTIER 4.1 can be downloaded from the Internet by accessing the address for the Centre for Efficiency and Productivity Analysis (CEPA) at the University of New England. <http://www.une.edu.au/econometrics/cepa.htm>.

<sup>5</sup>Such a model has more parameters to estimate and so the efficiency of estimation would be reduced.

frontier production function. These tests can be easily performed using a generalised likelihood-ratio test<sup>6</sup> and the FRONTIER 4.1 programme to estimate the appropriate models involved under the null and alternative hypotheses.

#### 4. CONCLUSION

The investigation of the effect of quality of irrigation water on crop production is a subject of considerable importance for the agricultural and economic development of Pakistan. It would require that considerable resources be expended on carefully designing and conducting a sufficiently large random sample of farmers. However, it is a research project that should be given some priority by government departments, in conjunction with researchers in relevant departments of the agricultural universities, which have the appropriate expertise in the areas required.

#### REFERENCES

- Aigner, D. C., A. K. Lovell, and P. Schmidt (1977) Formulation and Estimation of Stochastic Frontier Production Models. *Journal of Econometrics* 6: 21–37.
- Battese, G. E. (1997) A Note on the Estimation of Cobb-Douglas Production Functions When Some Explanatory Variables Have Zero Values. *Journal of Agricultural Economics* 48: 250–252.
- Battese, G. E. (1998) A Stochastic Frontier Model for the Analysis of the Effects of Quality of Irrigation Water on Crop Yields. Centre for Efficiency and Productivity Analysis, University of New England, Armidale. (CEPA Working Papers, No. 9/98.)
- Battese, G. E., and S. S. Broca (1997) Functional Forms of Stochastic Frontier Production Functions and Models for Technical Inefficiency Effects: A Comparative Study for Wheat Farmers in Pakistan. *Journal of Productivity Analysis* 8: 395–414.
- Battese, G. E., and T. J. Coelli (1995) A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Economics* 20: 325–332.
- Coelli, T. J., and G. E. Battese (1996) Identification of Factors which Influence the Technical Inefficiency of Indian Farmers. *Australian Journal of Agricultural*

<sup>6</sup>The generalised likelihood-ratio statistic is calculated as  $\lambda = -2 \log[L(H_0)/L(H_1)]$ , where  $L(H_1)$  and  $L(H_0)$  are the values of the likelihood functions for the original model and that for which the specifications of the null hypothesis,  $H_0$ , are imposed before estimation. This test statistic asymptotically has Chi-square or mixed Chi-square distribution with degrees of freedom equal to the difference between the numbers of parameters estimated in the two models. Coelli and Battese (1996) pointed out that, for the test for absence of the inefficiency effects in the stochastic frontier model,  $\lambda$  has a mixed Chi-square distribution and the appropriate critical values for the test are obtained from Table 1 of Kodde and Palm (1986). For this test, the value of  $\lambda$  is directly calculated by the FRONTIER programme.

- Economics* 40: 103–128.
- Fuller, W. A. (1975) Regression Analysis for Sample Survey. *Sankya: The Indian Journal of Statistics* 57: 117–132.
- Ghafoor, A., M. Qadir, and R. H. Qureshi (1991) Using Brackish Water on Normal and Salt-affected Soils in Pakistan: A Review. *Pakistan Journal of Agricultural Science* 28: 273–288.
- Huang, C. J., and J-T. Liu (1994) Estimation of a Non-neutral Stochastic Frontier Production Function. *Journal of Productivity Analysis* 5: 171–180.
- Kodde, D. A., and F. C. Palm (1986) Wald Criteria for Jointly Testing Equality and Inequality Restrictions. *Econometrica* 54: 1243–1248.
- Meeusen, W., and J. van den Broeck (1977) Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review* 18: 435–444.