Empirical Economics (1995) 20:325-332

EMPIRICAL ECONOMICS

A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data

G. E. BATTESE AND T. J. COELLI[1]

Department of Econometrics, The University of New England, Armidale, NSW 2351, Australia

Abstract: A stochastic frontier production function is defined for panel data on firms, in which the non-negative technical inefficiency effects are assumed to be a function of firm-specific variables and time. The inefficiency effects are assumed to be independently distributed as truncations of normal distributions with constant variance, but with means which are a linear function of observable variables. This panel data model is an extension of recently proposed models for inefficiency effects in stochastic frontiers for cross-sectional data. An empirical application of the model is obtained using up to ten years of data on paddy farmers from an Indian village. The null hypotheses, that the inefficiency effects are not stochastic or do not depend on the farmer-specific variables and time of observation, are rejected for these data.

JEL Classification System-Numbers: C12, C13, C23, C24, C87

1 Introduction

Since the stochastic frontier production function was independently proposed in Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977), there has been considerable research to extend and apply the model. Reviews of much of this research are provided in Førsund, Lovell and Schmidt (1980), Schmidt (1986), Bauer (1990), Battese (1992) and Greene (1993).

The stochastic frontier production function postulates the existence of technical inefficiencies of production of firms involved in producing a particular output. Most theoretical stochastic frontier production functions have not explicitly formulated a model for these technical inefficiency effects in terms of appropriate explanatory variables. Early empirical papers, in which the issue of the *explanation* of these inefficiency effects was raised, include Pitt and Lee (1981) and Kalirajan (1981). These papers adopt a two-stage approach, in which the first stage involves the specification and estimation of the stochastic frontier production function and the prediction of the technical inefficiency

[1] The authors are Associate Professor and Senior Lecturer, respectively, in the Department of Econometrics, University of New England, Armidale, NSW 2351, Australia. We thank Manolito Bernabe for assistance with data compilation. We gratefully acknowledge the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) for providing the panel data from its Village Level Studies.

effects, under the assumption that these inefficiency effects are identically distributed. The second stage involves the specification of a regression model for the *predicted* technical inefficiency effects, which contradicts the assumption of identically distributed inefficiency effects in the stochastic frontier.

Kumbhakar, Ghosh and McGuckin (1991), Reifschneider and Stevenson (1991) and Huang and Liu (1994) recently proposed models for the technical inefficiency effects involved in stochastic frontier functions. The parameters of the stochastic frontier and the inefficiency model are estimated simultaneously, given appropriate distributional assumptions associated with cross-sectional data on the sample firms.

The present paper proposes a model for technical inefficiency effects in a stochastic frontier production function for panel data. Provided the inefficiency effects are stochastic, the model permits the estimation of both technical change in the stochastic frontier and time-varying technical inefficiencies.

2 Inefficiency Frontier Model for Panel Data

Consider the stochastic frontier production function for panel data,

$$Y_{it} = \exp(x_{it}\beta + V_{it} - U_{it}) \tag{1}$$

where Y_{it} denotes the production at the t-th observation ($t = 1, 2, \ldots, T$) for the i-th firm ($i = 1, 2, \ldots, N$);[2]

x_{it} is a $(1 \times k)$ vector of values of known functions of inputs of production and other explanatory variables associated with the i-th firm at the t-th observation;

β is a $(k \times 1)$ vector of unknown parameters to be estimated;

the $V_{it}s$ are assumed to be iid $N(0, \sigma_V^2)$ random errors, independently distributed of the $U_{it}s$;

the $U_{it}s$ are non-negative random variables, associated with technical inefficiency of production, which are assumed to be independently distributed, such that U_{it} is obtained by truncation (at zero) of the normal distribution with mean, $z_{it}\delta$, and variance, σ^2;

z_{it} is a $(1 \times m)$ vector of explanatory variables associated with technical inefficiency of production of firms over time; and

δ is an $(m \times 1)$ vector of unknown coefficients.

Equation (1) specifies the stochastic frontier production function in terms of the original production values. However, the technical inefficiency effects, the

---

[2] Although it is assumed that there are T time periods for which observations are available for at least one of the N firms involved, it is not necessary that all the firms are observed for all T periods in this model specification.

U_{it}s, are assumed to be a function of a set of explanatory variables, the z_{it}s, and an unknown vector of coefficients, δ. The explanatory variables in the inefficiency model may include some input variables in the stochastic frontier, provided the inefficiency effects are stochastic. If the first z-variable has value one and the coefficients of all other z-variables are zero, then this case represents the model specified in Stevenson (1980) and Battese and Coelli (1988, 1992).[3] If all elements of the δ-vector are equal to zero, then the technical inefficiency effects are not related to the z-variables and so the half-normal distribution originally specified in Aigner, Lovell and Schmidt (1977) is obtained. If interactions between firm-specific variables and input variables are included as z-variables, then a non-neutral stochastic frontier, proposed in Huang and Liu (1994), is obtained.

The technical inefficiency effect, U_{it}, in the stochastic frontier model (1) could be specified in equation (2),

$$U_{it} = z_{it}\delta + W_{it} \tag{2}$$

where the random variable, W_{it}, is defined by the truncation of the normal distribution with zero mean and variance, σ^2, such that the point of truncation is $-z_{it}\delta$, i.e., $W_{it} \geq -z_{it}\delta$. These assumptions are consistent with U_{it} being a non-negative truncation of the $N(z_{it}\delta, \sigma^2)$-distribution. The inefficiency frontier production function (1)–(2) differs from that of Reifschneider and Stevenson (1991) in that the W-random variables are not identically distributed nor are they required to be non-negative, as in the latter paper. Further, the mean, $z_{it}\delta$, of the normal distribution, which is truncated at zero to obtain the distribution of U_{it}, is not required to be positive for each observation, as in Reifschneider and Stevenson (1991).

The assumption that the U_{it}s and the V_{it}s are independently distributed for all $t = 1, 2, \ldots, T$, and $i = 1, 2, \ldots, N$, is obviously a simplifying, but restrictive, condition. Alternative models are required to account for possible correlated structures of the technical inefficiency effects and the random errors in the frontier.

The method of maximum likelihood is proposed for simultaneous estimation of the parameters of the stochastic frontier and the model for the technical inefficiency effects. The likelihood function and its partial derivatives with respect to the parameters of the model are presented in Battese and Coelli (1993). The likelihood function is expressed in terms of the variance parameters, $\sigma_S^2 \equiv \sigma_V^2 + \sigma^2$ and $\gamma \equiv \sigma^2/\sigma_S^2$.

The technical efficiency of production for the i-th firm at the t-th observation is defined by equation (3),

$$TE_{it} = \exp(-U_{it}) = \exp(-z_{it}\delta - W_{it}) \ . \tag{3}$$

[3] Not including an intercept parameter, δ_0, in the mean, $z_{it}\delta$, may result in the estimators for the δ-parameters associated with the z-variables being biased and the shape of the distributions of the inefficiency effects, U_{it}, being unnecessarily restricted.

The prediction of the technical efficiencies is based on its conditional expectation, given the model assumptions. This result is also given in the Appendix of Battese and Coelli (1993).

3 Empirical Application

Data on paddy farmers from the Indian village of Aurepalle are considered for an empirical application of our model defined above. These data were collected by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). Information on the age and years of schooling for 14 paddy farmers from Aurepalle are used to explain the differences in the inefficiency effects among the farmers. Data on variables, such as the frequency of contacts with agricultural extension officers, access to credit, the use of high-yielding varieties, etc., were not available. The use of age, years of formal schooling and year of observation illustrate the methodology involved. A total of 125 observations are involved for a ten-year period from 1975–76 to 1984–85.

The stochastic frontier production function to be estimated is

$$\begin{aligned}\ell n(Y_{it}) = \beta_0 &+ \beta_1\, \ell n(\text{Land}_{it}) + \beta_2(\text{PILand}_{it}) + \beta_3\, \ell n(\text{Labour}_{it})\\ &+ \beta_4\, \ell n(\text{Bullocks}_{it}) + \beta_5\, \ell n[\text{Max}(\text{Costs}_{it}, 1 - D_{it})]\\ &+ \beta_6(\text{Year}_{it}) + V_{it} - U_{it} \qquad (4)\end{aligned}$$

where the technical inefficiency effects are assumed to be defined by

$$U_{it} = \delta_0 + \delta_1(\text{Age}_{it}) + \delta_2(\text{Schooling}_{it}) + \delta_3(\text{Year}_{it}) + W_{it} \qquad (5)$$

where ℓn denotes the natural logarithm (i.e., logarithm to the base e);

Y is the total value of output (in Rupees) for the farmer involved;[4]
Land is the total area of irrigated and unirrigated land operated (in hectares);
PILand is the proportion of the operated land that is irrigated;
Labour is the total hours of family and hired labour used on the farm;
Bullocks represents the hours of bullock labour used;
Costs refers to the value of fertilizer, manure, pesticides, machinery, etc.;
D is a variable which has value one if Costs are positive and zero, otherwise;
Age is the age of the primary decision maker in the farming operation;
Schooling is the years of formal schooling of the primary decision maker,
Year indicates the year of the observation involved; and
V_{it} and W_{it} are as defined in the previous section.

---

[4] Defining the production variable as the total *value of output*, rather than physical output of a crop, has implications for the interpretation of the inefficiency effect, U_{it}, in the frontier. In fact, this random variable accounts for any factors associated with inefficiency of production, including technical inefficiency. Use of value of output is required, given that the Indian farmers involved engaged in other agricultural activites, including mixed cropping, in addition to growing paddy.

The stochastic frontier production function in (4) can be viewed as a linearized version of the logarithm of the Cobb-Douglas production of function in which the land variable is a weighted average of the number of irrigated and unirrigated hectares of land used in the production of paddy and other crops. The variable, PILand, accounts for the differences in the productivities of irrigated and unirrigated land.

The inefficiency frontier model (4)–(5) accounts for both technical change and time-varying inefficiency effects. The *Year* variable in the stochastic frontier (4) accounts for Hicksian neutral technological change. However, the *Year* variable in the inefficiency model (5) specifies that the inefficiency effects may change linearly with respect to time. The distributional assumptions on the inefficiency effects permit the effects of technical change and time-varying behaviour of the inefficiency effects to be identified, in addition to the intercept parameters, β_0, and δ_0, in the stochastic frontier and the inefficiency model.

Maximum-likelihood estimates of the parameters of the model are obtained using a modification of the computer program, FRONTIER 2.0 (see Coelli, 1992). These estimates, together with the estimated standard errors of the maximum-likelihood estimators, given to two significant digits, are as follows:

Stochastic Frontier:

$$\ell n\, Y = \underset{(0.60)}{2.86} + \underset{(0.12)}{0.37}\, \ell n(\text{Land}) + \underset{(0.21)}{0.38}\, \ell n(\text{PILand}) + \underset{(0.13)}{0.85}\, \ell n(\text{Labour})$$

$$- \underset{(0.11)}{0.33}\, \ell n(\text{Bullocks}) + \underset{(0.031)}{0.071}\, \ell n(\text{Costs}) + \underset{(0.013)}{0.014}\, \text{Year}$$

Inefficiency Model:

$$U = \underset{(2.8)}{-1.5} + \underset{(0.034)}{0.035}\, \text{Age} - \underset{(0.077)}{0.006}\, \text{Schooling} - \underset{(0.60)}{0.57}\, \text{Year}$$

Variance Parameters: $\hat{\sigma}_S^2 = \underset{(0.75)}{0.74}$, $\hat{\gamma} = \underset{(0.047)}{0.952}$

Log(likelihood) $= -22.595$.

The signs of the coefficients of the stochastic frontier are as expected, with the exception of the negative estimate of the bullock-labour variable. The negative elasticity for bullock labour may be due to the fact that it is used more extensively in years of poorer rainfall (for weed control, levy bank improvements, etc.) when yields are lower. Thus bullock labour may be an inverse proxy for rainfall. The positive coefficient of the proportion of irrigated land confirms the expected positive relationship between the proportion of irrigated land and total value of production. The estimated coefficients for the land and labour variables, 0.37 and 0.85, respectively, are highly significant, while that

Table 1. Tests of hypotheses for parameters of the inefficiency frontier model for paddy farmers in Aurepalle

Null Hypothesis	Log(Likelihood)	$\chi^2_{0.95}$-value	Test statistic*
$H_0: \gamma = \delta_0 = \cdots = \delta_3 = 0$	−37.588	12.59	29.99*
$H_0: \gamma = 0$	−36.082	7.82	26.97*
$H_0: \delta_1 = \delta_2 = \delta_3 = 0$	−27.941	7.82	10.69*

* An asterisk on the value of the test statistic indicates that it exceeds the 95th percentile for the corresponding χ^2-distribution and so the null hypothesis is rejected.

for costs of other inputs is relatively small, but significant. The coefficient of *Year* indicates that the value of output has tended to increase by a small, but insignificant, rate over the ten-year period.

The estimated coefficients in the inefficiency model are of particular interest to this study. The *Age* coefficient is positive, which indicates that the older farmers are more inefficient than the younger ones. The negative estimate for *Schooling* implies that farmers with greater years of schooling tend to be less inefficient. However, the relationship is very weak, because the coefficient is very small relative to its estimated standard error. The negative coefficient for *Year* suggests that the inefficiencies of production of the paddy farmers tended to decline throughout the ten-year period.

The estimate for the variance parameter, γ, is close to one, which indicates that the inefficiency effects are likely to be highly significant in the analysis of the value of output of the farmers. Generalized likelihood-ratio tests[5] of null hypotheses, that the inefficiency effects are absent or that they have simpler distributions, are presented in Table 1. The first null hypothesis, which specifies that the inefficiency effects are absent from the model, is strongly rejected. The second null hypothesis, which specifies that the inefficiency effects are not stochastic,[6] is also strongly rejected. The third null hypothesis, considered in Table 1, specifies that the inefficiency effects are not a linear function of the age and schooling of the farmers and the year of observation. This null hypothesis is also rejected at the 5% level of significance. This indicates that the joint effects of these three explanatory variables on the inefficiencies of production is significant although the individual effects of one or more of the variables may not be statistically significant. The inefficiency effects in the stochastic frontier are clearly stochastic and are not unrelated to the age and level of formal

[5] The likelihood-ratio test statistic, $\lambda = -2\{\log[\text{Likelihood}(H_0)] - \log[\text{Likelihood}(H_1)]\}$, has approximately chi-square distribution with parameter equal to the number of parameters assumed to be zero in the null hypothesis, H_0, provided H_0 is true.

[6] If the parameter, γ, is zero, then the variance of the inefficiency effects is zero and so the model reduces to a traditional mean response function in which the variables, age and schooling of the farmers, are included in the production function. In this case, the parameters, δ_0 and δ_3, are not identified. Hence the critical value for the test statistic for this second null hypothesis is obtained from the χ^2_3-distribution.

schooling of farmers and year of observation. Thus it appears that, in this application, the proposed inefficiency stochastic frontier production function is a significant improvement over the corresponding stochastic frontier which does not involve a model for the technical inefficiency effects.

4 Conclusions

A model for technical inefficiency effects in a stochastic frontier production function is proposed for panel data. An application of the model is presented using data from 14 Indian paddy farmers, observed over a ten-year period. The results indicate that the model for the technical inefficiency effects, involving a constant term, age and schooling of farmers and year of observation, is a significant component in the stochastic frontier production function. The application also illustrates that the model specification permits the estimation of both technical change and time-varying technical inefficiency, given that inefficiency effects are stochastic and have a known distribution.

Further theoretical and applied work is obviously required to obtain better and more general models for stochastic frontiers and the technical inefficiency effects associated with the analysis of panel data.

References

Aigner DJ, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6:21–37

Battese GE (1992) Frontier production functions and technical efficiency: A survey of empirical applications in agricultural economics. Agricultural Economics 7:185–208

Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies: With a generalized frontier production function and panel data. Journal of Econometrics 38:387–399

Battese GE, Coelli TJ (1992) Frontier production functions, technical efficiency and panel data with application to paddy farmers in india. Journal of Productivity Analysis 3:153–169

Battese GE, Coelli TJ (1993) A stochastic frontier production function incorporating a model for technical inefficiency effects. Working Papers in Econometrics and Applied Statistics No 69, Department of Econometrics. University of New England. Armidale

Bauer PW (1990) Recent developments in the econometric estimation of frontiers. Journal of Econometrics 46:39–56

Coelli TJ (1992) A computer program for frontier production function estimation: FRONTIER Version 2.0. Economics Letters 39:29–32

Førsund FR, Lovell CAK, Schmidt P (1980) A survey of frontier production functions and their relationship to efficiency measurement. Journal of Econometrics 13:5–25

Greene WH (1993) The econometric approach to efficiency analysis. In: Fried HO, Lovell CAK, Schmidt SS (Eds) The measurement of productive efficiency: Techniques and applications. Oxford University Press New York 68–119

Huang CJ, Liu J-T (1994) Estimation of a non-neutral stochastic frontier production function. Journal of Productivity Analysis 5:171–180

Kalirajan K (1981) An econometric analysis of yield variability in paddy production. Canadian Journal of Agricultural Economics 29:283–294

Kumbhakar SC, Ghosh S, McGuckin JT (1991) A generalized production frontier approach for estimating determinants of inefficiency in US dairy farms. Journal of Business and Economic Statistics 9:279–286

Meeusen W, van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review 18:435–444

Pitt MM, Lee M-F (1981) The measurement and sources of technical inefficiency in the Indonesian weaving industry. Journal of Development Economics 9:43–64

Reifschneider D, Stevenson R (1991) Systematic departures from the frontier: A framework for the analysis of firm inefficiency. International Economic Review 32:715–723

Schmidt P (1986) Frontier production functions. Econometric Reviews 4:289–328

Stevenson RF (1980) Likelihood functions for generalized stochastic frontier estimation. Journal of Econometrics 13:57–66

First version received: April 1994
Final version received: September 1994