Estimation of Technical Inefficiency Effects Using Panel Data and Doubly Heteroscedastic Stochastic Production Frontiers

by

K. Hadri*
The University of Liverpool, Liverpool L69 3BX, UK
Hadri@liv.ac.uk
C. Guermat and J. Whittaker
University of Exeter, Exeter EX4 4PU, UK

Corresponding author:
Professor Kaddour Hadri
Department of Economics and Accounting
Eleanor Rathbone Building
Bedford Street south
Liverpool L69 7ZA, UK
E-mail: Hadri@liv.ac.uk

Running title: Technical Inefficiency Effects and Heteroscedasticity.

[0] *The authors are grateful to the Economic and Social Research Council for access to their Data Archive which has provided the data for this research. We are indebted to Badi Baltagi and two anonymous referees for their helpful comments and suggestions. The usual caveat applies.

Abstract

In previous studies, measures of technical inefficiency effects derived from stochastic production frontiers have been estimated from residuals which are sensitive to specification errors. This study corrects for this inaccuracy by extending the doubly heteroscedastic stochastic cost frontier suggested by Hadri (1999) to the model for technical inefficiency effects in a stochastic frontier production function for panel data proposed by Battese and Coelli (1995). The correction for heteroscedasticity is supported by the data. The study uses, for illustration of the techniques, data on 101 mainly cereal farms in England. We provide both point estimates and confidence intervals for technical efficiencies. The confidence intervals are constructed by extending the "Battese-Coelli" method reported by Horrace and Schmidt (1996) by allowing the technical inefficiency to be time varying and the disturbance terms to be heteroscedastic. The confidence intervals reveal the precision of technical efficiency estimates and show the deficiencies of making inferences based exclusively on point estimates.

JEL classification: C23; C24; D24; Q12.
Keywords: stochastic frontier production; heteroscedasticity; technical efficiency; panel data.

1. Introduction

In previous studies of efficiency using stochastic frontier models, Caudill et al. (1995) noted that measures of inefficiency are based on residuals derived from the estimation of a stochastic frontier. They observed that residuals are sensitive to specification errors, particularly in stochastic frontier models, and that this sensitivity will be passed on to the inefficiency measures. To correct for this, they suggested that one should consider testing for and, if present, correcting for heteroscedasticity in the one sided error term. Hadri (1999) argued that the two-sided error term might be expected to be affected by heteroscedasticity as well. Ignoring this likely eventuality leads to inconsistent maximum likelihood (ML) estimators and the usual tests are no longer valid. Hence, in order to obtain correct estimators and valid tests one must test for heteroscedasticity in both error terms and, if indicated, appropriate correction should be made in the estimation procedure. Using the same data as Caudill et al. (1995), Hadri (1999) found that firm-specific inefficiency measures are extremely sensitive to the correction for heteroscedasticity in both random terms. Other forms of known heteroscedasticity have been considered in the literature, see for instance Kumbhakar (1997). For a recent review of stochastic frontier models the reader is referred to Greene (1997).

In this paper, we consider heteroscedasticity of a known form in the stochastic terms in the model for technical inefficiency effects in a stochastic frontier production function for panel data proposed by Battese and Coelli (1995). In their model, Battese and Coelli (1995) explicitly express technical inefficiency effects in terms of appropriate explanatory variables. The parameters of the stochastic frontier and the inefficiency model are then estimated simultaneously using the method of maximum likelihood with panel data. Previous applied papers, including Pitt and Lee (1981) and Kalirajan (1981), 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 effects, under the assumption that these inefficiency effects are identically distributed. The second stage requires the specification of a regression model for the predicted technical inefficiency effects. However, this contradicts the assumption of identically distributed inefficiency effects in the stochastic frontier. The simultaneous estimation avoids this serious problem (see Battese and Coelli (1995) for more details and other related references). Additionally, the consideration, in this paper, of heteroscedasticity in Battese and Coelli model seeks to prevent inconsistency and to obtain valid tests when heteroscedasticity of a known form is present. Three specifications are considered: heteroscedasticity in the one sided term, heteroscedasticity in the symmetrical error term and in both error terms. We also calculate technical efficiency and their confidence intervals for several formulations of the stochastic frontier models. We extend the "Battese-Coelli method" (see Horrace and Schmidt (1996)) for constructing confidence intervals to the case of time varying technical inefficiency and heteroscedastic error terms. The confidence intervals reveal the precision of technical efficiency estimates and show the deficiencies of making inferences based exclusively on point estimates. To illustrate the application of these techniques, we apply them to a set of panel data on 101 mainly cereal farms in England for the harvest years 1982-1987.

The paper is organised as follows. The theoretical models are presented in section 2. In section 3 the models are applied to our data set . Section 4 deals with the construction of confidence intervals and their analysis. Section 5 concludes the paper.

2. Theoretical models

Before presenting the heteroscedastic stochastic production frontier models with technical inefficiency effects, we briefly present the model of Battese and Coelli (1995):

$$y_{it} = x_{it}\beta + w_{it} - v_{it}, \tag{1}$$

where y_{it} denotes the logarithm of the production for the ith sample farm ($i = 1, \ldots, N$) in the tth time period ($t = 1, \ldots, T$); x_{it} is a $(1 \times k)$ vector of the logarithm of the inputs associated with the ith sample farm in the tth time period (the first element is set to one if an intercept term is included); β is a $(k \times 1)$ vector of unknown parameters to be estimated; the w_{it} s are assumed to be iid $N(0, \sigma_w^2)$ error terms, independent of the v_{it}s which are non-negative disturbance terms, associated with the technical inefficiency of production. The v_{it}s are assumed to be independently distributed, such that v_{it} is obtained by truncation at zero of the normal distribution with mean $Z_{it}\delta$, and variance σ_v^2. Z_{it} is a $(1 \times m)$ vector of firm-specific variables which may vary over time. δ is an $(m \times 1)$ vector of unknown coefficients of the firm-specific inefficiency variables. The one-sided disturbance v_{it} reflects the fact that each firm's production must lie on or below its frontier. Such a term represents factors under the firm's control. The two-sided error term represents factors outside the firm's control. Battese and Coelli (1995) assume that the v_{it}s are a function of a set of explanatory variables, the Z_{it}s, an unknown vector of coefficients, δ and a disturbance term, u_{it}, defined by the truncation of the normal distribution with mean zero and variance σ_v^2.

Under the above assumptions, Battese and Coelli (1995), following Weinstein (1964), derive the density function of $\epsilon_{it} = w_{it} - v_{it}$ which we reproduce here with a slightly different notation:

$$f(\epsilon_{it}) = \left[\sigma\Phi\left(\frac{Z_{it}\delta}{\sigma_v}\right)\right]^{-1} \phi\left(\frac{\epsilon_{it} + Z_{it}\delta}{\sigma}\right)\Phi\left(\frac{\mu_{it}^*}{\sigma^*}\right), \quad -\infty < \epsilon_{it} < +\infty, \tag{2}$$

where, $\sigma^2 = \sigma_w^2 + \sigma_v^2$, $\sigma^* = \sigma_v\sigma_w/\sigma$, $\mu_{it}^* = \left(\sigma_w^2 Z_{it}\delta - \sigma_v^2\epsilon_{it}\right)/\sigma^2$ and $\phi(.)$ and $\Phi(.)$ are respectively, the standard normal density and distribution functions. The advantage of stochastic frontier estimation is that it permits the estimation of firm-specific inefficiency. Following Jondrow et al. (1982) it is straightforward to show that the conditional expected value of v given ϵ is given by:

$$E\left[v_{it}|\epsilon_{it}\right] = \mu_{it}^* + \sigma^*\frac{\phi\left(\mu_{it}^*/\sigma^*\right)}{\Phi\left(\mu_{it}^*/\sigma^*\right)}, \tag{3}$$

where μ_{it}^* and σ^* are as defined previously. We use this formula to evaluate efficiencies. More specifically, if y_{it} is the logarithm of output, technical efficiency of the ith firm in the tth time period is $TE_{it} = \exp\left(-\left[v_{it}\right]\right)$ and technical inefficiency is equal to $1 - TE_{it}$.

The log-likelihood function of a homoscedastic model given by (1) and (2) for possibly unbalanced panel data is given by:

$$\begin{aligned} L\left(\beta, \delta, \sigma_w^2, \sigma_v^2; y\right) &= \sum_{i=1}^{N}\sum_{t=1}^{T_i} \ln\left[f_{it}(\epsilon_{it})\right] \\ &= -\frac{1}{2}\left(\sum_{i=1}^{N} T_i\right)\left(\ln 2\pi + \ln\sigma^2\right) \\ &\quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(y_{it} - x_{it}\beta + Z_{it}\delta\right)^2/\sigma^2 \\ &\quad -\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(\ln\Phi\left(\frac{Z_{it}\delta}{\sigma_v}\right) - \ln\Phi\left(\frac{\mu_{it}^*}{\sigma^*}\right)\right). \end{aligned} \tag{4}$$

In considering the three possible cases of heteroscedasticity (in the one-sided, two-sided and both error terms), we follow Hadri (1999) in assuming the following multiplicative heteroscedasticity for the one-sided error term

$$\sigma_{vit} = \exp(V_{it}\boldsymbol{\gamma}), \tag{5}$$

where V_{it} is a vector of nonstochastic explanatory variables related to characteristics of firm management and $\boldsymbol{\gamma}$ is a vector of unknown parameters that is assumed to include an intercept parameter. V_{it} is assumed to include an intercept term. The standard deviation of the two-sided error term is also written in exponential form so that $\sigma_w = \exp(\alpha_0)$. To obtain the density function corresponding to the model where only the one-sided error term is assumed heteroscedastic, we replace the variances by their new expressions in (3).

As argued earlier, in the cross-section dimension the two-sided error is likely to be affected by size-related heteroscedasticity. The misspecification resulting from not incorporating heteroscedasticity in the ML estimation of our frontier can cause parameter estimators to be inconsistent as well as invalidating standard techniques of inference, see White (1982). In order to incorporate heteroscedasticity in the two-sided error term we write $\sigma_{wit} = \exp(W_{it}\boldsymbol{\alpha})$, where W_{it} is a vector of nonstochastic explanatory variables related generally to characteristics of firm size and $\boldsymbol{\alpha}$ is a vector of unknown parameters that is assumed to include an intercept parameter. W_{it} is assumed to include an intercept term. The standard deviation of the one-sided error term, assumed here to be homoscedastic, becomes $\sigma_v = \exp(\gamma_0)$. The density function remains as in (3) but the variances are replaced by their new expressions.

Last but not least, the most general specification is the one where both error terms are assumed to be concurrently heteroscedastic. The density in (3) is still appropriate but, now we have to use $\sigma_{wit} = \exp(W_{it}\boldsymbol{\alpha})$ and $\sigma_{vit} = \exp(V_{it}\boldsymbol{\gamma})$.

3. Illustrative application

As an illustration, a set of panel data on 101 English farms, classified as 'mainly cereal' under the nationally organised Farm Business Survey, was used for the years 1982-1987 to estimate the four stochastic frontier production functions. Data on output and input are collected only in value and cost terms, and are here deflated by the appropriate price index to proxy physical output and inputs. The characteristics of the data are summarised in Table 1. One feature of the sample is variability. In all variables, the standard deviation is large compared to the mean. The range of the data in Table 1 also indicate the diversity of the sample. Another feature is size dispersion; a farm that is one standard deviation above the mean is more than 6 times larger than a farm that is one standard deviation below the mean.

[Table 1 here]

In this study, we consider the flexible translog function

$$\ln C_{it} = \beta_0 + \sum_{j=1}^{5} \beta_j x_{jit} + \sum_{j \leq k}^{5} \sum_{k=1}^{5} \beta_{jk}\, x_{jit} x_{kit} + w_{it} - v_{it}, \tag{6}$$

where

C represents the cereal output;

x_1 represents the logarithm of labour;

x_2 represents the logarithm of the total amount of land (in acres) on which cereals were grown;

x_3 represents the logarithm of chemicals (fertilizers and crop protection);

x_4 represents the logarithm of other inputs (machinery, energy, seeds, and others);

x_5 represents the year of observation.

The model for the technical efficiency effects in the stochastic frontier of equation (6), is defined by

$$v_{it} = \delta_0 + \delta_1 z_{1it} + \delta_2 z_{2it} + \delta_3 z_{3it} + \delta_4 z_{4it} + u_{it}, \tag{7}$$

where

z_1 and z_2 are dummy variables for business type (sole trader, partnership, and farm company); where $z_1 = 1$ for partnership and zero otherwise, and $z_2 = 1$ for farming company and zero otherwise.

z_3 is a dummy representing whether the farm produces cereals only ($z_3 = 1$ for specialised farms, zero otherwise);

z_4 represents the year of observation.

$$\sigma_{wit} = \exp(\alpha_0 + \alpha_1 TA_{it}), \tag{8}$$

where TA is the logarithm of total area. The two-sided error term is likely to be affected by size-related heteroscedasticity. We consider that the best measure of size is the total area. This last measure includes all size-related heterogeneities, the one due to cereals production as well as other productions.

$$\sigma_{vit} = \exp(\gamma_0 + \gamma_1 x_{3it}), \tag{9}$$

where x_{3it} represents the logarithm of chemicals . The usual assumption here is that the heteroscedasticity in the one-sided stochastic term is associated with factors under the farm's control. We believe that the variable, chemicals, fulfills adequatly this requirement.

The elasticity of mean production is given by:

$$\frac{\partial \ln[E(C_{it})]}{\partial x_k} = \beta_k + 2\beta_{kk} x_{kit} + \sum_{j \neq k}^{5} \beta_{kj}\ x_{jit}. \tag{10}$$

For models M2 and M3 and only for the elasticity with respect to x_2,we should add the following term:

$$\left(\exp\left[\alpha_0 + \alpha_1 TA_{it}\right]\right)^2 \alpha_1 \exp\left[x_{2it} - TA_{it}\right]. \tag{11}$$

This last term is due to the fact that the variance of the two-sided stochastic term, in the heteroscedastic case, depends indirectly on x_2 for these two models.

The year of observation is included in the model to account for technological change (Hicksian neutral) even though the time period considered is short.

Although we use the flexible translog function, we note that there are other potential flexible functional forms, including the use of neural network (for more details see Guermat and Hadri (1999)).

Generally, management characteristics such as age and education are taken as important determinants of technical inefficiency. Unfortunately, we do not have such data. Instead, we use business type and specialisation dummies to represent management characteristics. The time variable is included to see whether inefficiency effects change linearly with respect to time as in Battese and Coelli (1995).

We estimated seven stochastic frontier production functions using GQOPT/PC version 6.01 routines for the optimisation of the likelihood functions. Model M1 is the most general model with a translog functional form and both error terms accounting for heteroscedasticity. Model M2 is the same as M1 except that it has a Cobb-Douglas functional form. All the remaining models have a translog functional form. Model M3 excludes explanatory variables for the technical inefficiency effects. Model M4 assumes both error terms to be homoscedastic. In model M5 only v is assumed to be homoscedastic, whereas in model M6 sole w is homoscedastic. Model M7 is the selected model. It assumes a translog functional form, technical inefficiency effects, w homoscedastic, v heteroscedastic and no time effect in the technical inefficiency component. Maximum-likelihood estimates of the seven models are reported in Table A1.

Likelihood ratio statistics were used to test the above specifications and are reported in Table 2. All the tests were carried out using 5% significance level. Model M1 nests all the other models. We started by testing Cobb-Douglas versus Translog, obtaining a likelihood ratio of 84.07 indicating the rejection of the Cobb-Douglas specification. We then tested the hypothesis that there are no technical inefficiency effects. This hypothesis is also rejected on the basis of a likelihood ratio of 53.476. The null hypothesis of v and w being homoscedastic is rejected with a likelihood ratio of 15.91. Next, we tested the hypothesis of a homoscedastic v. This hypothesis is rejected. However, the null hypothesis of w homoscedastic is not rejected. Finally, there is insufficient evidence to reject the restrictions entailed by model M7.

[Table 2 here]

Technical efficiency was estimated for the selected model (M7) and three other models for comparison purpose. The aim was to examine the effects of using the Cobb-Douglas functional form (M2), ignoring technical inefficiency effect (M3) and ignoring heteroscedasticity in both error terms (M4).

The histograms of technical efficiencies from the four models are shown in Figure 1. The various models clearly produce different empirical distributions.

[Figure 1 here]

Table 3 displays the yearly mean technical efficiency of all farms for the four models. Model M2 has the most monotonic (same tendency) yearly mean technical efficiency over time. The Cobb-Douglas model is unstable, giving the highest mean efficiency (0.918) in 1982 and the lowest mean efficiency (0.758) in 1987. The highest average of efficiency is produced by model M3, which suggests that in our case, using the wrong model might overestimate the mean efficiency.

[Table 3 here]

The difference between the four models is more obvious when we compare the ranking of yearly mean efficiency. Table 4 gives the highest, median and lowest three efficiencies for each model respectively. Both M4 and M7 place farm 21 and 51 at the top three farms, and farm 70 and 76 at the bottom three farms. The two models give different farms at the middle ranking. The other models produce very different ranking of farms. We notice, however, that farm 21 is classified in the top three farms in three out of four models, while farm 76 is classified in the bottom three farms in three out of four models. The median farms show a clear cut difference among the models, as no model suggest the same median farms.

[Table 4 here]

In addtion to the absolute measures of efficiency, the relations between the firm-specific efficiency ranking are of interest to establish the extent to which the four models affects the determination of relative rather than absolute efficiency. Table 5 gives the correlation between the rankings of these models. It is clear that the rankings are generally sensitive to the assumption adopted.

[Table 5 here]

The elasticities, shown in Table 6, are all positive as expected. Models M7 and M4 produce similar elasticities and return to scale. However, although the estimated returns to scale are similar in M2 and M3, the individual elasticities are different. Models M4 and M2 produce the lowest and the highest return to scale respectively. All these elasticities are significant at the 5% level except the mean labour elasticities in model M3 ($t = 1.59$). The standard errors have been calculated using the delta method as suggested by one of the referees.

[Table 6 here]

4. Technical efficiency and confidence intervals

The construction of confidence intervals for panel data assuming technical inefficiency time invariant is reported by Horrace and Schmidt (1996) who refer to it as the "Battese-Coelli" method . We extend this method to allow for the technical inefficiency to be time varying. The (1-λ)100% confidence interval (L_{it}, U_{it}) for $TE_{it} = \exp(-E[v_{it}|\epsilon_{it}])$ is given by:

$L_{it} = \exp(-\mu^*_{it} - z_{L_{it}}\sigma^*_{it})$,

$U_{it} = \exp(-\mu^*_{it} - z_{U_{it}}\sigma^*_{it})$,

where

$z_{L_{it}} = \Phi^{-1}\{1 - (\lambda/2)[1 - \Phi(-\mu^*_{it}/\sigma^*_{it})]\}$,

$z_{U_{it}} = \Phi^{-1}\{1 - (1 - \lambda/2)[1 - \Phi(-\mu^*_{it}/\sigma^*_{it})]\}$,

and Φ is the standard normal cdf and the other parameters have been defined previously. Due to the large numbers of farms only nine are reported for each year; these are the three farms with the highest technical efficiencies, the three with median technical efficiencies and the three farms with the lowest technical efficiencies. The efficiency levels are not estimated as precisely as one might have desired. For the year 1987 the farm with the highest estimated efficiency level had an estimated efficiency of 0.9566 but a 95% confidence interval ranging from 0.8646 to 0.9987. The median farm for the same year had an estimated efficiency of 0.8846, with a 95% confidence interval of (0.7272, 0.9942). The worst farm in the sample had an estimated efficiency of 0.4785 with a 95% confidence interval of (0.3959, 0.5784). For the other years we obtain similar patterns. The confidence intervals are relatively wide and overlapping. If we consider the year 1982 for example, all the highest and median farms reported in Table 6 overlap. Only the three lowest farms do not overlap with the first six farms. This would lead us to deduce that the six farms (highest and median) are efficient and that the three bottom farms are less efficient. This contrasts with the point estimates of technical efficiency for which we would conclude that farm 21 is the most efficient farm and that the following farms would be ranked in decreasing order of efficiency. As noted by Horrace and Schmidt (1996), the method used to construct confidence intervals considers the parameters of the model to be known and therefore the confidence intervals do not reflect uncertainty about these parameters. For large N this is not significant since the variability in the parameters estimates is small relative to the variability intrinsic to the distribution of the v_{it} given ϵ_{it}. On the other hand, the lack of precision in the estimation of technical inefficiency is mainly due to the relative variability of v_{it} and w_{it}. In our case we obtain for the estimate of var$(w) = \sigma^2_w$, 0.0445 and for the estimate of $\overline{var(v_{it})} = 0.0186$. This means that on average the variance of w (statistical noise) is almost two and half times as large as the variance of v. This makes it very difficult to estimate v_{it} precisely, hence the wide confidence intervals obtained. To evaluate $\overline{var(v_{it})}$, we used the formula derived by Stevenson (1980).

[Table 7 here]

Figure 2 shows the estimated technical efficiencies for 1987 along with their corresponding 95% confidence intervals for models M4 and M7 respectively. Due to space, the graphs for models M2 and M3 are not reported. In all the models, farms were ranked according to model M7 ranking. The smoothness of the plot for M7 is clearly disturbed in the other models. This indicates that different specifications lead to different ranking of farms as found for instance in Hadri (1999). It is not easy to compare the confidence intervals of the different models. However, it seems that the Cobb-Douglas model has the widest confidence intervals for almost all farms.

[Figure 2 here]

In figure 3 the technical efficiencies for the four models, M2, M3, M4 and M7 are plotted concurrently using M7 ranking. It is clear that the Cobb-Douglas (M2) specification consistently and greatly underestimates technical efficiency. The homoscedastic model M4 also underestimates technical efficiency for most farm, but the magnitude of this underestimation is generally less than that of M2. Ignoring technical efficiency effects (M3) leads to an overestimation of technical efficiency for a great majority of farms, especially those at the lower end.

[Figure 3 here]

figures 4 and 5 show for model M4 and M5 respectively, the evolution of technical efficiencies during 1982 to 1987 for three top farms, three middle farms and three bottom farms. The figures for models M2 and M3 are not reported here for the same reason as above. The ranking is based on the average technical efficiencies during the six years. Overall, the top farms are the most stable during the period concerned followed by the middle farms. The least stable during the same period are the least efficient farms.

[Figures 4 and 5 here]

5. Conclusion

This study has demonstrated that heteroscedasticity within an estimation can have significant effects on results. The models developed in this paper allow for heteroscedasticity in the Battese and Coelli (1995) stochastic production frontier with technical inefficiency effects model. Having tested for heteroscedasticity, we found that only the two-sided error was homoscedastic. The results provide further support for the Hadri (1999) argument that the correction for heteroscedasticity is essential in order to obtain correct estimates, valid tests and satisfactory measures of efficiency. We also extended the "Battese-Coelli" method for constructing confidence intervals to the case of time varying technical inefficiency and heteroscedastic error terms. Finally, our results points to the importance of (a) using a flexible functional form for the unknown production technology, (b) considering technical efficiency effects and (c) allowing for heteroscedasticity in the error terms.

References

Battese GE, Coelli, TJ (1995) A Model for Technical Inefficiency Effects in a Stochastic Frontier Production for Panel Data. Empirical Economics 20: 325-332

Caudill SB, Ford JM, Gropper DM (1995) Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity. Journal of Business & Economic Statistics 13: 105-111

Guermat C, Hadri K (1999) Backpropagation Neural Network Versus Translog in Stochastic Frontiers: A Monte-Carlo Analysis. Discussion paper in Economics 99/16 Exeter University

Greene WH (1997) Frontier Production Functions. In: Pesaran H, Schmidt P (eds.) Handbook of Applied Econometrics, Vol. II - Microeconomics. Basil Blackwell London, pp. 81-166

Hadri K, Whittaker J (1999) Efficiency, Environmental Contaminants and Farm Size: Testing for Links using Stochastic Production Frontiers. Journal of Applied Economics, 2: 337-356

Hadri K (1997) A Frontier Approach to Disequilibrium Models. Applied Economic Letters 4: 699-701

Hadri K (1999) Estimation of a Doubly Heteroscedastic Stochastic Frontier Cost Function. Journal of Business & Economic Statistics, 17: 359-363

Horrace WC, Schmidt P (1996) Confidence Statements for Efficiency Estimates from Stochastic Frontier Models. Journal of Productivity Analysis, 7: 257-282

Jondrow J, Lovell CAK, Materov I, Schmidt P (1982) On the Estimation of Technical Inefficiency in Stochastic Production Function Model. Journal of Econometrics 19: 233-238

Kalirajan K (1981) An Econometric Analysis of Yield Variability in Paddy Production. Canadian Journal of Agricultural Economics 29: 283-294

Kumbhakar SC (1997) Efficiency Estimation with Heteroscedasticity in a Panel Data Model. Applied Economics 29: 379-386

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

Stevenson RF (1980) Likelihood Functions for Generalized Stochastic Frontier Estimation. Journal of Econometrics 13: 57-66

Weinstein MA (1964) The Sum of Values From a Normal and a Truncated Normal Distribution. Technometrics 6: 104-105 (with some additional material, 469-470)

White H (1982) Maximum Likelihood Estimation of Misspecified Models. Econometrica 50: 1-25

$$\overline{var(v_{it})} = \left[\overline{Z_{it}}\delta\right]^2 \left[2\Phi\left(\frac{\overline{Z_{it}}\delta}{\overline{\sigma_v}}\right)\right]^{-1} \left(1 - \left[2\Phi\left(\frac{\overline{Z_{it}}\delta}{\overline{\sigma_v}}\right)\right]^{-1}\right)$$
$$+\overline{\sigma_v^2} \left[2\Phi\left(\frac{\overline{Z_{it}}\delta}{\overline{\sigma_v}}\right)\right]^{-1} \left(\pi - \left[\Phi\left(\frac{\overline{Z_{it}}\delta}{\overline{\sigma_v}}\right)\right]^{-1}\right) / \pi$$

The elasticity of mean production with respect to x_2 when heterocedasticity is assumed to be present in the two-sided disturbance term (models M2 and and M3):

$$\frac{\partial \ln[E(C_{it})]}{\partial x_2} = \beta_2 + 2\beta_{22} x_{2it} + \sum_{j\neq 2}^{5} \beta_{2j}\ x_{jit} + \left(\exp\left[\alpha_0 + \alpha_1 TA_{it}\right]\right)^2 \alpha_1 X_2$$

where X_2 represents the total amount of land (not the logarithm) on which cereals were grown.

Table 1. Summary statistics for variables in the stochastic frontier model.

	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
C	6098.51	1278502.46	211068.08	167984.19	1.897	5.439
LABOUR	656.88	115959.93	22352.58	16987.90	2.212	6.920
CAREA	7.30	882.62	134.92	97.62	2.113	8.494
CHEM	574.96	166491.58	27080.09	22947.91	2.110	6.575
OTHER	3801.69	267854.71	43333.90	34240.50	2.491	9.979
TOTAREA	34.10	1107.21	174.69	127.32	2.274	9.123

Size dispersion = (mean+sd)/(mean-sd)=8.80. (for C).

Cereal Area and Total Area are in hectares. All other variables are in Sterling Pounds at 1985 prices.

Table 2. Likelihood-ratio tests.

Null Hypothesis	Log Likelihood	LR	Critical Value (5%)	Decision
M1 (Translog)	771.446			
M2 H_0 : β_{ij}=0, $i \leq j$=1,…,5	729.410	84.072	24.99	Reject H_0 (Cobb-Douglas)
M3 H_0 : δ_i=0, i=1,…,4	745.131	52.630	9.48	Reject H_0 (No ineff. effect)
M4 H_0 : $\alpha_1 = \gamma_1 = 0$	763.493	15.906	5.99	Reject H_0 (*v* and *w* homosc.)
M5 H_0 : $\alpha_1 = 0$	771.214	0.464	3.84	Accept H_0 (*v* homosc.)
M6 H_0 : $\gamma_1 = 0$	768.311	6.270	3.84	Reject H_0 (*w* homosc.)
M7 H_0 : $\delta_4 = \alpha_1 = 0$	771.188	0.516	5.99	Accept H_0

Table 3. Yearly Mean Technical Efficiency.

	1982	1983	1984	1985	1986	1987	Average
M2	0.9188	0.8860	0.8907	0.8452	0.8388	0.7584	0.856314
M3	0.8976	0.8639	0.8962	0.8702	0.9036	0.8770	0.884741
M4	0.8655	0.8165	0.8621	0.8284	0.8723	0.8332	0.846319
M7	0.8777	0.8368	0.8754	0.8475	0.8866	0.8599	0.863988

Table 4. Mean efficiency ranking. (Mean of each firm over 6 years)

	M2		M3		M4		M7	
	Firm	Efficiency	Firm	Efficiency	Firm	Efficiency	Firm	Efficiency
Highest	40	0.9517	38	0.9491	21	0.9446	21	0.9550
	21	0.9424	51	0.9481	51	0.9409	63	0.9481
	54	0.9379	28	0.9446	24	0.9407	51	0.9475
Median	65	0.8715	45	0.9006	47	0.8588	45	0.8795
	52	0.8698	96	0.8995	74	0.8494	92	0.8789
	49	0.8684	81	0.8980	96	0.8473	72	0.8783
Lowest	73	0.6824	76	0.7266	76	0.6533	85	0.6864
	89	0.6557	46	0.7230	70	0.6495	70	0.6827
	85	0.6292	85	0.6674	79	0.6111	76	0.6721

Table 5.Elasticities

	Labour	Cereal Area	Chemicals	Other Inputs	Return to Scale
M2	0.040	0.379	0.359	0.328	1.106
M3	0.035	0.360	0.372	0.334	1.101
M4	0.063	0.369	0.301	0.329	1.062
M7	0.064	0.380	0.301	0.326	1.071

Table 6. Confidence intervals (95%) for the selected model (M7).

Year		Farm No.	T.eff.	Lbnd	Ubnd
1982	Highest	21	0.9701	0.9006	0.9991
		51	0.9621	0.8761	0.9989
		35	0.9616	0.8781	0.9989
	Median	78	0.9007	0.7790	0.9936
		23	0.8998	0.7786	0.9933
		61	0.8995	0.7670	0.9944
	Lowest	100	0.7091	0.5573	0.8990
		46	0.6735	0.5663	0.8009
		73	0.6456	0.5333	0.7815
1983	Highest	52	0.9704	0.9044	0.9991
		23	0.9592	0.8742	0.9987
		51	0.9516	0.8503	0.9985
	Median	11	0.8577	0.6883	0.9911
		56	0.8571	0.7203	0.9850
		77	0.8548	0.6963	0.9891
	Lowest	79	0.6439	0.5006	0.8281
		85	0.6193	0.5287	0.7254
		76	0.5596	0.4544	0.6893
1984	Highest	63	0.9670	0.8931	0.9990
		52	0.9616	0.8832	0.9988
		64	0.9591	0.8737	0.9987
	Median	49	0.9007	0.7563	0.9954
		42	0.9007	0.7464	0.9959
		92	0.8993	0.7705	0.9940
	Lowest	89	0.6822	0.5947	0.7826
		79	0.6742	0.5229	0.8679
		76	0.6446	0.5282	0.7867

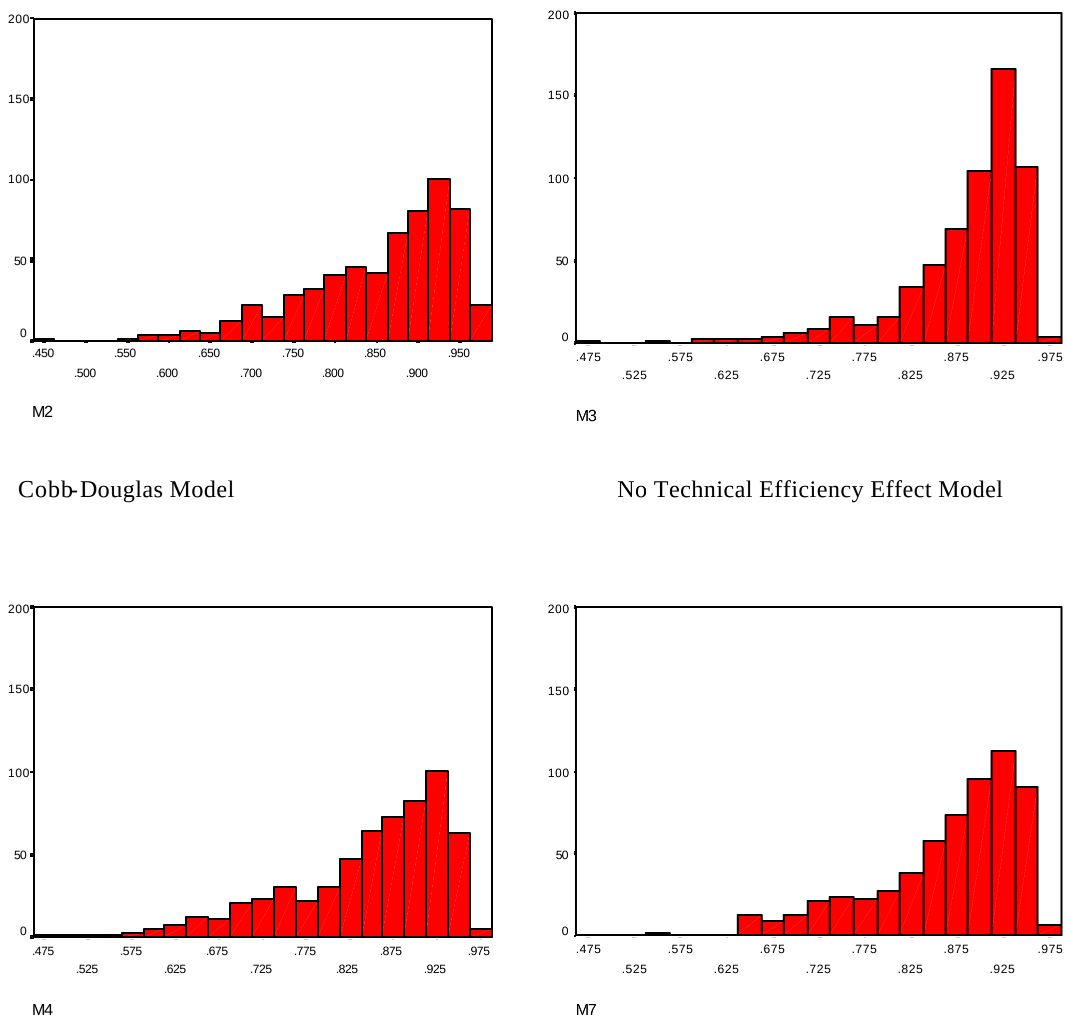
Table 6. Continued.

Year		Farm No.	T.eff.	Lbnd	Ubnd
1985	Highest	63	0.9582	0.8709	0.9987
		20	0.9571	0.8691	0.9987
		21	0.9544	0.8591	0.9986
	Median	43	0.8766	0.7277	0.9922
		66	0.8757	0.7484	0.9885
		19	0.8745	0.7543	0.9862
	Lowest	79	0.6469	0.5034	0.8309
		70	0.6413	0.5130	0.8016
		85	0.5903	0.5032	0.6926
1986	Highest	52	0.9734	0.9124	0.9992
		51	0.9661	0.8878	0.9990
		84	0.9651	0.8853	0.9990
	Median	6	0.9095	0.7716	0.9960
		101	0.9068	0.7576	0.9962
		42	0.9065	0.7562	0.9963
	Lowest	9	0.6980	0.5570	0.8737
		18	0.6683	0.5333	0.8374
		7	0.5799	0.4593	0.7322
1987	Highest	24	0.9566	0.8646	0.9987
		84	0.9526	0.8537	0.9985
		21	0.9465	0.8410	0.9983
	Median	8	0.8846	0.7272	0.9942
		48	0.8778	0.7160	0.9936
		72	0.8754	0.7348	0.9909
	Lowest	22	0.6575	0.5352	0.8079
		70	0.5548	0.4429	0.6949
		82	0.4785	0.3960	0.5784

Table A1. Estimation Results for the Seven Models.

		Translog (M1)		Cobb-Douglas (M2)		No T. Eff. Effect (M3)		Homoscedastic *v* & *w* (M4)		Homoscedastic *v* (M5)		Homoscedastic *w* (M6)		Selected Model (M7)	
		Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value
Frontier	Constant (β_0)	4.429	1.810	2.979	12.861	11.045	5.174	4.305	2.060	4.155	1.572	3.882	2.758	3.863	2.794
	Labour (β_1)	-1.796	-4.944	0.040	2.051	-2.614	-8.387	-1.315	-2.748	-1.443	-2.632	-1.610	-6.626	-1.618	-9.449
	Cereal Area (β_2)	0.923	0.990	0.379	8.801	2.923	7.837	1.660	3.183	1.095	1.940	0.853	2.042	0.855	3.427
	Chemicals (β_3)	0.466	0.468	0.359	9.913	-0.871	-1.286	-0.487	-0.890	-0.010	-0.030	0.360	2.051	0.369	1.187
	Other Inputs (β_4)	1.517	2.149	0.328	10.707	1.343	1.906	1.640	5.045	1.628	4.147	1.583	21.753	1.584	8.924
	Year (β_5)	-0.096	-0.838	0.003	0.205	-0.093	-0.788	-0.001	-0.005	-0.050	-0.466	-0.094	-0.836	-0.090	-0.837
	Labour^2 (β_{11})	0.066	3.084			0.079	4.339	0.061	2.734	0.058	2.600	0.061	8.150	0.061	3.385
	Cereal Area^2 (β_{22})	-0.116	-0.903			0.042	0.690	-0.000	-0.004	-0.046	-0.640	-0.102	-2.949	-0.102	-1.861
	Chemicals^2 (β_{33})	-0.386	-6.647			-0.290	-3.960	-0.283	-3.641	-0.322	-7.337	-0.363	-18.270	-0.364	-12.328
	Other Inputs^2 (β_{44})	-0.258	-3.635			-0.263	-11.787	-0.239	-7.947	-0.249	-4.344	-0.250	-15.369	-0.249	-7.502
	Year^2 (β_{55})	-0.015	-5.343			-0.013	-4.494	-0.013	-4.425	-0.013	-4.751	-0.015	-5.243	-0.015	-5.276
	Labour x Cer. Area (β_{12})	-0.305	-5.088			-0.433	-7.191	-0.264	-3.078	-0.263	-2.786	-0.275	-8.550	-0.278	-6.775
	Labour x Chem (β_{13})	0.242	4.267			0.315	4.366	0.195	4.195	0.204	2.224	0.230	11.773	0.230	6.211
	Labour x Other (β_{14})	-0.035	-0.440			0.004	0.149	-0.047	-1.454	-0.038	-0.350	-0.045	-1.570	-0.045	-0.922
	Labour x Year (β_{15})	-0.008	-0.696			-0.009	-0.775	-0.007	-0.633	-0.005	-0.406	-0.006	-0.580	-0.006	-0.534
	CerArea x Chem (β_{23})	0.227	1.291			-0.012	-0.112	0.093	0.662	0.149	1.537	0.199	7.425	0.202	3.180
	CerArea x Other (β_{24})	0.138	1.209			0.147	1.202	0.042	0.477	0.083	1.188	0.128	3.425	0.127	2.417
	CerArea x Year (β_{25})	-0.038	-1.710			-0.035	-1.563	-0.021	-0.989	-0.029	-1.349	-0.036	-1.908	-0.036	-1.689
	Chem x Other (β_{34})	0.383	3.533			0.372	3.813	0.389	3.571	0.381	8.842	0.375	12.379	0.372	13.505
	Chem x Year (β_{35})	0.019	0.969			0.023	1.169	-0.002	-0.105	0.000	0.025	0.015	0.759	0.014	0.782
	Other x Year (β_{45})	0.022	1.255			0.017	0.947	0.024	1.409	0.027	1.656	0.024	0.924	0.024	1.513
Efficiency	Constant (δ_0)	0.188	1.750	-0.072	-0.558	-0.301	-0.518	0.149	1.322	0.167	1.683	0.143	1.127	0.129	1.192
	D(partnership) (δ_1)	-0.151	-1.942	-0.099	-2.509			-0.186	-2.644	-0.201	-2.934	-0.192	-1.961	-0.191	-2.088
	D(farm company) (δ_2)	-0.176	-1.771	-0.162	-1.864			-0.258	-1.912	-0.191	-1.556	-0.183	-1.603	-0.183	-1.665
	D(cereal only) (δ_3)	-0.305	-1.789	-0.150	-3.777			-0.302	-2.552	-0.454	-2.627	-0.405	-1.562	-0.396	-1.711
	Year (δ_4)	-0.004	-0.216	0.069	2.444			0.005	0.255	-0.006	-0.309	-0.005	-0.230		
σ_w	Constant (α_0)	-2.407	-2.082	-4.371	-4.446	-2.694	-3.041	-1.527	-9.372	0.058	0.111	-1.555	-7.838	-1.555	-8.221
	Total Area (α_1)	0.153	0.745	0.482	2.841	0.265	2.255			-0.316	-2.880				
σ_v	Constant (γ_0)	1.034	1.151	0.704	0.708	1.570	2.492	-2.129	-22.380	-2.116	-26.610	0.401	0.755	0.398	0.675
	Chemicals (γ_1)	-0.316	-3.324	-0.266	-2.508	-0.362	-5.322					-0.250	-4.357	-0.250	-3.967
Log Likelihood		771.446		729.410		745.131		763.493		768.311		771.214		771.188	
Likelihood Ratio				**82.332**		**52.630**		**15.906**		**6.270**		0.464		0.516	

The statistics in bold are significant at the 5% level of significance.



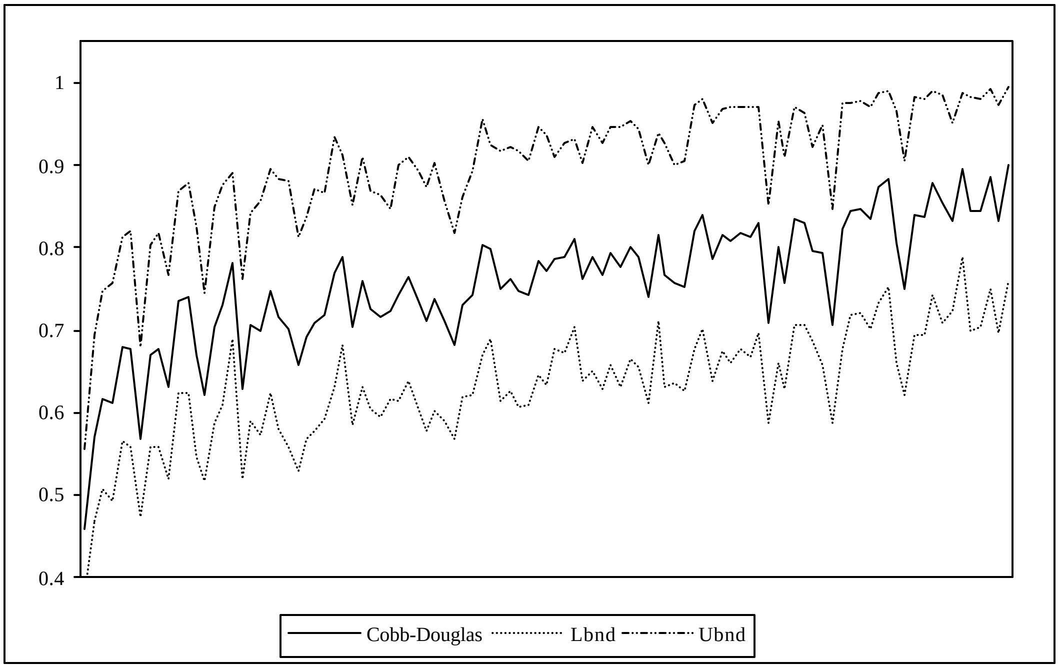
Cobb-Douglas Model

No Technical Efficiency Effect Model

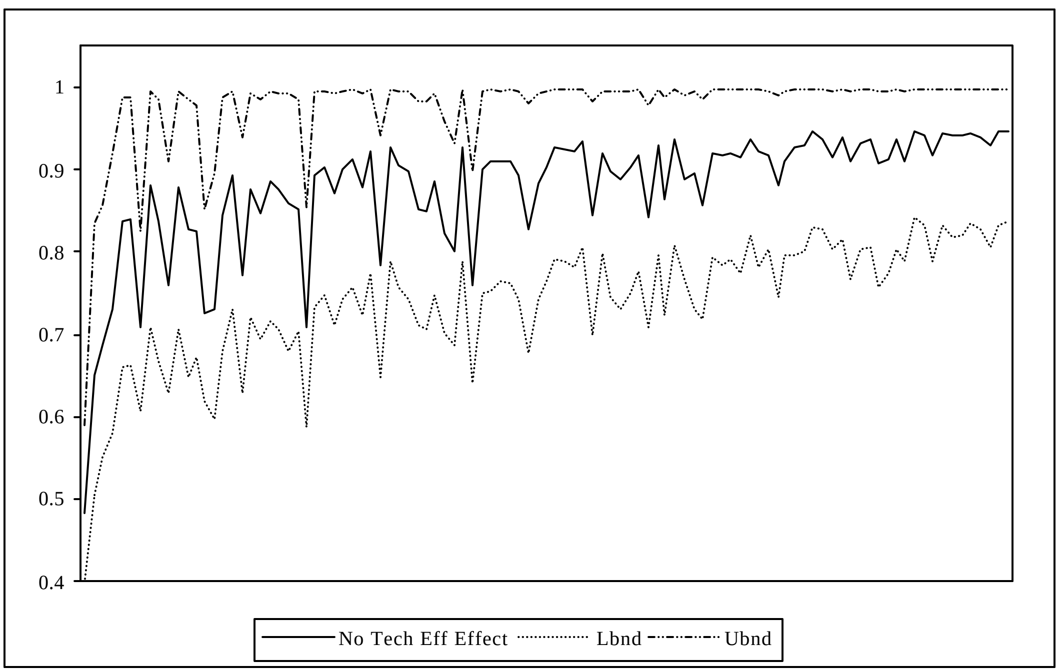
Homoscedastic Model

Selected Model

Figure 1. Empirical Distribution of Technical Efficiencies.



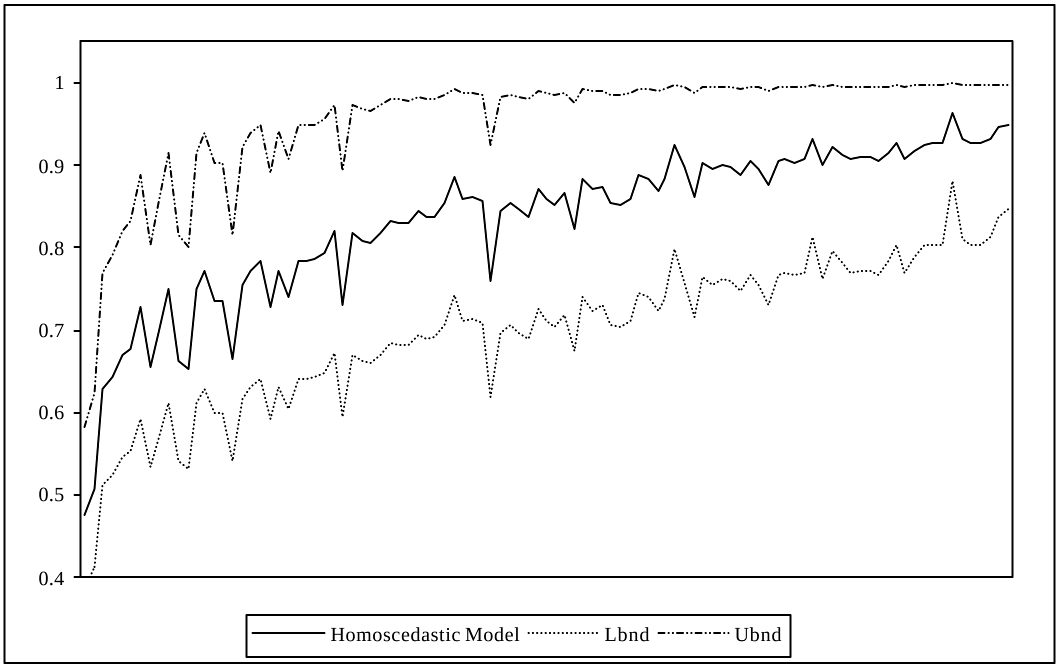
(M2)



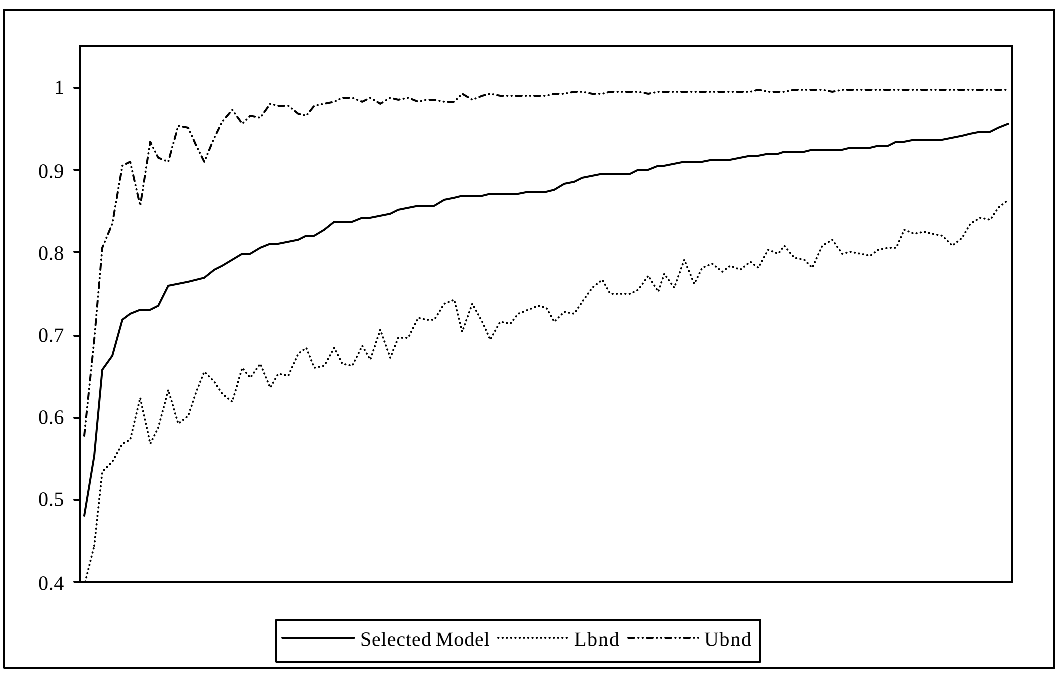
(M3)

Figure 2. Technical efficiencies ranked according to selected model with 95% confidence intervals (1987).

The graphs in figure 2 were produced by ranking farms according to their technical efficiency using the selected model. The other three models then follow the same ranking of the selected model. In this way we can compare the three models with the selected model.



(M4)



(M7)

Figure 2. Continued.

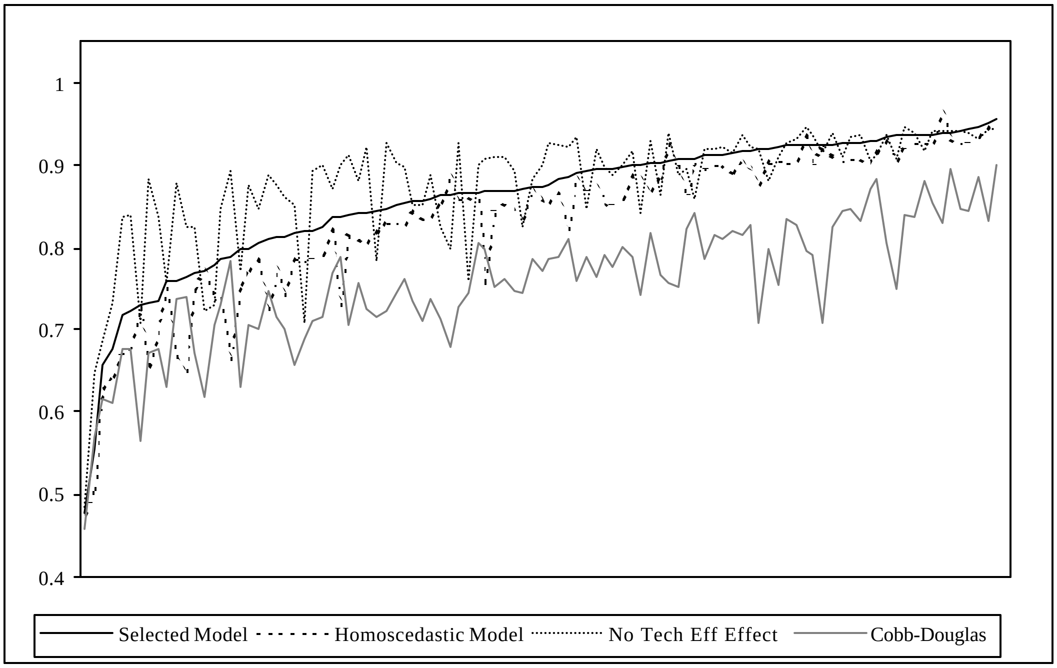


Figure 3. Technical efficiencies based on the ranking from the selected model (1987).

(you will see it better if you print it)

- Cobb-Douglass consistently underestimates technical efficiency.
- Ignoring heteroscedasticity underestimates technical efficiency for about the third of farms. At the upper end the difference is less prominent.
- Ignoring technical efficiency effect over estimates technical efficiency for more than half the farms. The difference is smaller at the very top.

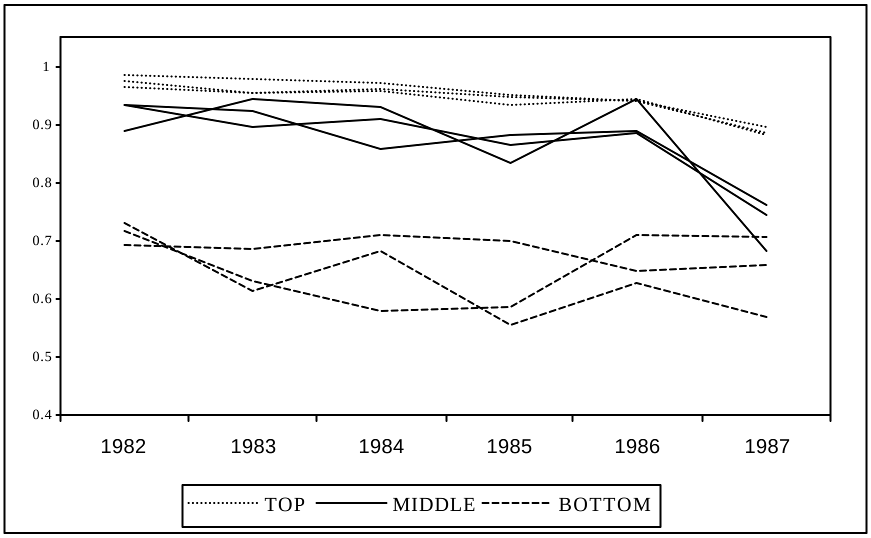


Figure 4. Technical Efficiency of Three Top, Middle and Bottom Farms (Cobb-Douglas, M2).

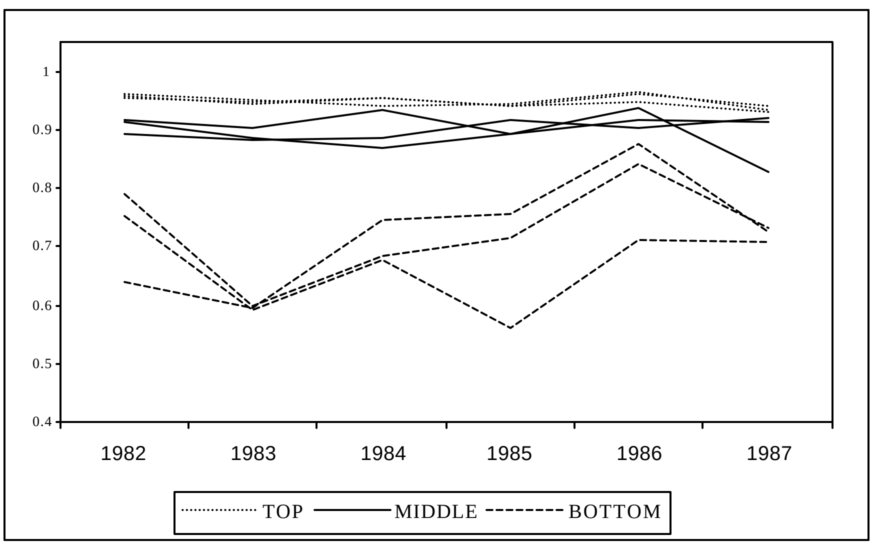


Figure 5. Technical Efficiency of Three Top, Middle and Bottom Farms (No technical Efficiency Effect, M3).

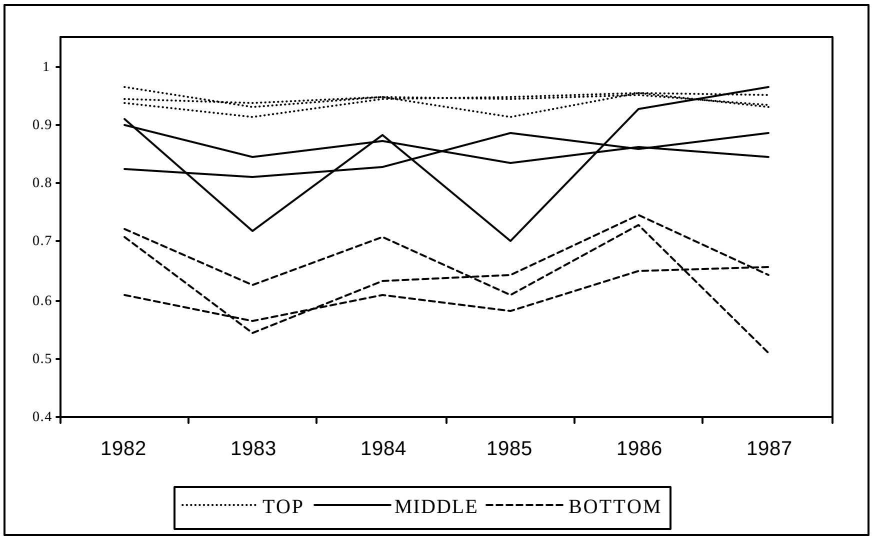


Figure 6. Technical Efficiency of Three Top, Middle and Bottom Farms (Homosecedastic, M4).

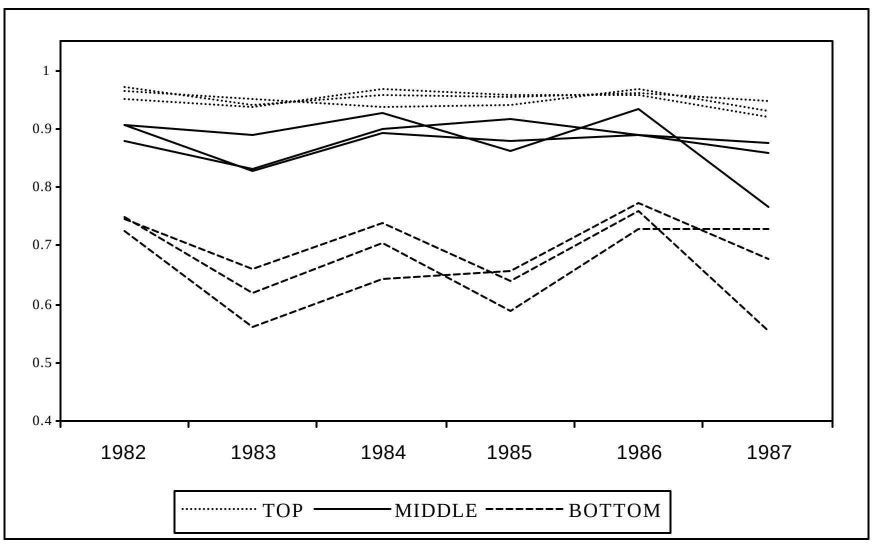


Figure 7. Technical Efficiency of Three Top, Middle and Bottom Farms (Selected Model, M7).