Tropical Agricultural Research Vol. 15: 217-225 (2003)

# Comparison of Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DEA) to Evaluate Technical Efficiency: Illustrated by Efficiency Analysis of Cattle Farming Systems in Up-Country Wet-Zone of Sri Lanka

R.N.Serasinghe, M.B.P. Mahipala and L.H.P. Gunaratne

Department of Physical Science
Faculty of Applied Sciences
Rajarata University of Sri Lanka, Polgolla, Sri Lanka.

*ABSTRACT. During the last two decades technical efficiency measurements were developed along two competing paradigms, namely parametric stochastic frontier analysis (SFA) and non-parametric data envelopment analysis (DEA). However, the consistency and compatibility of these two approaches were not investigated by any of these studies. In this study, technical efficiency scores obtained from the above two techniques are compared using production information related to cattle farming systems in the Up-Country Wet Zone of Sri Lanka. Maximum likelihood estimates of the stochastic frontier were obtained and tested for returns to scale. These individual technical efficiencies were then compared with the technical efficiency values estimated from constant returns to scale, output-oriented, data envelopment analysis. Of the two cattle farming systems studied, both approaches revealed that integrated vegetable-based system is more efficient compared to milk-based system, in terms of milk revenue and total revenue. However, the technical efficiency scores obtained from the two approaches depicted slightly different distribution patterns, as data envelopment analysis ignores the random error.*

## INTRODUCTION

Efficiency of a production unit may be defined as how effectively it uses variable resources for the purpose of profit maximization, given the best production technology is available. From an applied perspective, measuring efficiency is important because this is the first step in a process that might lead to substantial resource savings. These resource savings have important implications for both policy formulation and firm management.

The concept of productive efficiency is further decomposed into two components, technical and allocative efficiency. Technical efficiency refers to the ratio of actual output to the maximum attainable level of output, for a given level of production inputs. Allocative efficiency refers only to the adjustment of inputs and outputs to reflect relative prices, having chosen the production technology.

Farrell (1957) suggested two approaches, (i.e., parametric and non-parametric) to measure the technical efficiency and during the past few decades the studies developed along these paradigms. Of these, the econometric (parametric) approach has

[1] Department of Animal Science, Faculty of Agriculture, University of Peradeniya, Sri Lanka.
[2] Department of Agricultural Economics, Faculty of Agriculture, University of Peradeniya, Sri Lanka.

been motivated to develop stochastic frontier models based on the deterministic parameter frontier of Aigner and Chu (1968). The Stochastic Frontier Analysis (SFA), which was independently proposed by Aigner *et al.* (1977) and Meeusen and van den Broeck (1977), acknowledges the random noise around the estimated production frontier. The main strength of the stochastic frontier analysis is that it can deal with stochastic noise. The need for the underlying technology and an explicit distributional assumption are the main weaknesses of the SFA.

The non-parametric approach or mathematical programming method developed by Charnes *et al.* (1978) mainly focuses on the development of piece-wise linear function using Data Envelopment Analysis (DEA). The main advantage of the DEA approach is that no explicit functional form is imposed on the data, and DEA can easily accommodate multiple outputs.

Both SFA and DEA are used in different fields such as agriculture, education and industry to compare various systems. The compatibility and consistency of efficiency scores estimated from these two approaches were not evaluated properly. Therefore, it is vital to examine the behaviour of technical efficiencies generated by these two techniques using a real data set.

Nuwara Eliya district, which is situated in Up-Country Wet-Zone, plays a significant role in Sir Lankan dairy industry with respect to cattle density and productivity (Dept. of Census and Statistics, 2001). Two distinct cattle farming systems that are practiced in this area are: milk-based systems in tea estates and vegetable-based integrated systems in villages (Ibrahim *et al.*, 1999). Though several workers (Ibrahim *et al.*, 1999; Mahipala, 2001) compared these two major cattle farming systems, none have attempted to study the production economics of Up-Country cattle farming systems. Such a comparison is deemed to be appropriate since identification of best-practice farmers and information on individual technical efficiencies possesses important policy implications.

Therefore, the objectives of this study were to estimate the technical efficiencies of two major farming systems in Up-Country Wet Zone using SFA and DEA and to compare the results of these two approaches.

**Theoretical framework**

In the econometric approach to efficiency measurement, production frontiers are characterized by smooth, continuous, twice differentiable, quasi-concave production transformation functions. The frontier is the limit to the range of possible productions. Hence, we have $TE_i = \dfrac{y_i}{f(x_i;\beta)}$, where $y_i$ is the scalar output of producer $i$, $i$=1,2...N, $x_i$ is a vector of K inputs used by producer $i$, $f(x_i;\beta)$ is the production frontier, $\beta$ is a vector of technology parameters to be estimated, and $TE_i$ is the output-oriented technical efficiency of producer $i$. This defines technical efficiency as the ratio of observed output to maximum output feasible under the current technology used where $y_i$ achieves its maximum value of $f(x_i;\beta)$ if, and only if, $TE_i$ =1. The amount by which an observation lies below the frontier is called inefficiency when $TE_i$ <1. The stochastic production frontier Aigner *et al.* (1977), Battese and Cora

(1977) and Meeusen and van den Broeck (1977) incorporate producer-specific random shocks into the analysis.

These models allow for technical inefficiency, but they also acknowledge the fact that random shocks outside the control of producers can affect output. An appropriate formulation of a stochastic frontier model in terms of a general production function for the i^th^ production unit is $y_i = f(x_i;\beta) + v_i - u_i = f(x_i;\beta) + \varepsilon_i$, where $v_i$ is the two-sided "noise" component, and $u_i$ is the non-negative technical inefficiency component of the error term. Thus the error term $\varepsilon_i = v_i - u_i$ is not symmetric, since $u_i \geq 0$. The parameters of this model are estimated by maximum likelihood estimation procedure, given suitable distributional assumptions for the error terms. Although, there are various distributional assumptions, we make only the distributional assumptions, $v_i \sim iid\,N(0,\sigma_v^2)$ and $u_i \sim iid\,N(0,\sigma_u^2)$. Moreover, it is assumed that $u_i \geq 0$ (i.e. non negative half normal), and $v_i$ and $u_i$ are distributed independently of each other. The estimates of technical inefficiency for the i^th^ firm, $u_i$ can be obtained from $TE_i = \exp\{-u_i\}$.

DEA is the non-parametric mathematical programming approach to frontier estimation. DEA is an extreme point method and compares each producer with only the 'best' producers. Extreme point methods are not always the right tool for a problem, but are appropriate in certain cases, based on a set of *n* linear programming problems.

If there is data on K inputs and M outputs on each of N producers or decision making units (DMUs) then for i^th^ producer these are represented by the vectors $x_i$ and $y_i$, respectively. The KxN input matrix, $X$, and the MxN output matrix, $Y$, represent the data of all N decision making units or producers. The purpose of DEA is to construct a non-parametric envelopment frontier over the data points such that all observed points lie on or below the production frontier. The output-oriented CRS model is given below;

$$
\begin{aligned}
&\max_{\phi,\lambda}\phi,\\
&\text{s.t.}\quad -\phi y_i + Y\lambda \geq 0,\\
&\qquad\quad x_i - X\lambda \geq 0,\\
&\qquad\quad \lambda \geq 0,
\end{aligned}
$$

where, $1 \leq \phi < \infty$, and $\phi - 1$ is the proportional increase in outputs that could be achieved by the i^th^ DMU, with input quantities held constant. Note that $1/\phi$ defines a technical efficiency (TE) score, which varies between zero and one.

A few of the characteristics that make it powerful are, ability to handle multiple input and multiple output models and free from assumption of a functional form relating inputs to outputs. However the same characteristics that make DEA a powerful tool can also create problems. Some of these limitations are; sensitivity to statistical noise, slow convergence towards the absolute efficiency and inability to test statistical hypotheses.

# MATERIALS AND METHODS

## Data

The data used in this study was collected from cattle farming systems in Nuwara Eliya district. In this study, we considered monthly milk revenue and monthly total revenue as outputs from both systems. Milk–based system is taken as system 1 while vegetable - based system is taken as system 2.

In this paper the following variables are taken as input variables for efficiency analysis:

$X_1$ - Number of cows per herd( % cow units ).
$X_2$ - Monthly expenditure for concentrated feed and minerals ( Rs. ).
$X_3$ - Monthly expenditure for veterinary drugs and breeding (Rs.).
$X_4$ - Monthly payments for labor ( Rs. ).

The dependent variables, $Y_1$ and $Y_2$ of cattle farming are considered under two different cases:

Case I: $Y_1$ - Monthly milk revenue (Rs.), i.e., the income of milk sales and the value of household milk consumption.

Case II: $Y_2$ - Monthly total revenue ( Rs.), i.e., the income of milk, cattle and manure sales and the value of household milk consumption and manure used.

## Estimation of stochastic production frontier and technical efficiency

The stochastic frontier model suggested by Aigner *et al.*, (1977) and Meeusen and van den Broeck (1977) is given by,

$$Y_i = \beta_0 \prod_{k=1}^{4} X_{ik}^{\beta_k} \exp(v_i - u_i) \quad : \quad i = 1,2,\ldots,n \text{ ---------------- } (1)$$

The Cobb-Douglas functional form has been incorporated into the model (1) due to its computational feasibility (Heady and Dillon, 1969). It is also the most widely used algebraic form in farm efficiency studies (Bravo-Ureta and Pinheiro, 1993). Using log transformation the model can be written in the form;

$$\mathrm{Ln}\, Y_i = \mathrm{Ln}\, \beta_0 + \beta_1 \mathrm{Ln}\, X_{i1} + \beta_2 \mathrm{Ln}\, X_{i2} + \beta_3 \mathrm{Ln}\, X_{i3} + \beta_4 \mathrm{Ln}\, X_{i4} + v_i - u_i,$$

where $X_{i1}$, $X_{i2}$, $X_{i3}$, and $X_{i4}$ are the independent variables as defined above and $v_i$ and $u_i$ are as explained in the introduction. Using Battese and Coelli (1992) parametric specification, the maximum likelihood estimation provides the estimates for $\beta$ coefficients, $\sigma^2$ and $\gamma$;

where $\sigma^2 = \sigma_v^2 + \sigma_u^2$, and $\gamma = \sigma_u^2 / (\sigma_u^2 + \sigma_v^2)$.

The Stochastic Production Frontiers and TE's for case I and case II were estimated by the maximum likelihood method using FRONTIER 4.1 (Coelli, 1994 ).

The results include. estimates of ordinary least square production function (OLS), maximum likelihood estimates of frontier production function (MLE), log-likelihood ratio and technical efficiencies of individual milk farmers. A 't' test was employed to detect whether the production operates on constant returns to scale, where the restriction for constant returns to scale (CRS) with Cobb- Douglas specification is $\sum_{k=1}^{4} \beta_k = 1$, which was taken as the null hypothesis. Since DEA estimations vary with the returns to scale parameter, this will be quite useful in obtaining comparable results in DEA estimation, which will be given in the next section.

### Estimation for DEA models and technical efficiency

The DEA production frontiers for milk revenue (case I) and total revenue (case II) were specified for the same output and input variables that were used in estimation of Stochastic Production Frontiers. Among various DEA models, standard output-oriented CRS envelopment surfaces were estimated since SFA has identified returns to scale parameter is not that different from one. The model was estimated using DEAP 2.1 (Coelli, 1996).

## RESULTS AND DISCUSSION

### Results of the stochastic production frontier estimation

Table 1 depicts the results of the SFA estimation for milk-revenue and total -revenue.

**Table 1. ML estimates of SFA model for milk- revenue and total- revenue.**

| Description | Parameter | Estimated Coefficient | |
|---|---|---|---|
| | | Milk-Revenue | Total-Revenue |
| Intercept | $\beta_0$ | 3.724* (0.741) | 4.179* (0.705) |
| Number of cows per herd | $\beta_1$ | 1.013* (0.197) | 1.087* (0.558) |
| Monthly expenditure for concentrated feed and minerals | $\beta_2$ | 0.664* (0.020) | 0.630* (0.065) |
| Monthly expenditure for veterinary drugs and breeding | $\beta_3$ | -0.092** (0.103) | 0.020** (0.063) |
| Monthly payments for labor | $\beta_4$ | 0.136** (0.210) | 0.094** (0.120) |
| Random error | $\sigma^2$ | 0.869* (0.262) | 0.880* (0.144) |
| Log likelihood ratio | | 5.511 | 6.280 |
| t – statistic (test for CRS) | | 1.661 | 0.793 |

(Figures in parentheses denote standard errors)

* Significant at 0.05; ** Significant at 0.01

The pooled t test employed to test the null hypotheses of CRS is not rejected at $p \leq 0.05$ in both cases. This implies that the production systems operate on CRS. Therefore, in DEA estimations for the comparison, CRS assumption will be made. Moreover, statistically significant log likelihood ratio indicates that production frontier with MLE is a better representation than that of average OLS production function.

### Comparison of milk-based system and vegetable-based system

Technical efficiency scores were obtained for the cattle farming based on SFA and output-oriented DEA models using CRS assumption. The summary is shown in table 2.

**Table 2. Comparison of mean technical efficiencies between two systems based on SFA & DEA models.**

| | Technical Efficiency | | | |
|---|---|---|---|---|
| | Milk-revenue (Case I) | | | |
| | SFA | | DEA | |
| Description | System I | System II | System I | System II |
| Minimum | 0.0860 | 0.2450 | 0.1110 | 0.3510 |
| Maximum | 0.8340 | 1.0000 | 0.6950 | 1.0000 |
| Mean | 0.4020 | 0.6568 | 0.3859 | 0.6334 |
| Standard deviation | 0.1892 | 0.2335 | 0.1759 | 0.2295 |
| d.f | 35 | | 35 | |
| Pooled t-statistic | -3.67 | | -3.72 | |
| p-value | 0.0008 | | 0.0007 | |
| | Total-revenue (Case II) | | | |
| | SFA | | DEA | |
| | System I | System II | System I | System II |
| Minimum | 0.0940 | 0.3130 | 0.1420 | 0.4410 |
| Maximum | 0.9990 | 0.9440 | 1.0000 | 1.0000 |
| Mean | 0.4011 | 0.6526 | 0.4467 | 0.7275 |
| Standard deviation | 0.2103 | 0.2034 | 0.2314 | 0.2144 |
| d.f | 35 | | 35 | |
| Pooled t-statistic | -3.65 | | -3.77 | |
| p-value | 0.0008 | | 0.0006 | |

( System I = milk-based; System II = vegetable-based )

As shown in Table 2, the mean technical efficiencies in vegetable-based systems were higher than that of milk-based systems for both the approaches. The mean technical efficiencies of these two systems were statistically compared using a pooled t test. When milk revenue is the dependent variable (case I), the computed t statistics for SFA and DEA estimations were -3.67 ($P = 0.0008$) and -3.72 ($P = 0.0007$), respectively. Similarly, when the total revenue (case II) is used as the dependent variable, the calculated t statistics were -3.65 (0.0008) and -3.77 (0.0006), respectively.

This clearly indicates that irrespective of the approach or response variable used, integrated the vegetable-based systems perform better than the milk-based systems.

**Comparison of the results of the two techniques, SFA and DEA for composite system.**

The other objective of this study is to examine the consistency and compatibility between these efficiency estimation techniques. In this case, disregarding the two farming systems, the distribution of the technical efficiencies is observed (Table 3).

**Table 3. Comparison of mean technical efficiencies between two techniques, SFA & DEA, for milk-revenue & total-revenue.**

| Description | Technical Efficiency | | | |
|---|---|---|---|---|
| | Milk- revenue (Case I) | | Total- revenue (Case II) | |
| | SFA | DEA | SFA | DEA |
| Minimum | 0.0860 | 0.1110 | 0.0940 | 0.1420 |
| Maximum | 1.0000 | 1.0000 | 0.9990 | 1.0000 |
| Mean | 0.5122 | 0.4929 | 0.5099 | 0.5681 |
| Standard deviation | 0.2428 | 0.2337 | 0.2403 | 0.2623 |
| Coefficient of correlation | 0.811 | | 0.829 | |
| p-value | 0.0001 | | 0.0001 | |
| d.f | 36 | | 36 | |
| Paired t-statistic | 0.80 | | -2.38 | |
| p-value | 0.43 | | 0.023 | |

As shown in the Table 3, significantly high coefficients of correlation between technical efficiencies of SFA and DEA for case I and case II indicate the consistency of these two approaches. However, paired t test based on individual technical efficiencies revealed that though efficiency scores are associated the implications of the comparisons are not necessarily the same.

## CONCLUSIONS

The stochastic frontier analysis revealed that the milk production operates on constant returns to scale and ML frontier function is a better representation compared to average OLS function. The average technical efficiencies of these two systems were found to be 0.402 and 0.657 (milk revenue as the dependent variable) 0.401 and 0.653 (total revenue as the dependent variable) for milk-based systems and vegetable-based systems, respectively. The DEA estimation was based on output oriented CRS model and the respective technical efficiencies obtained were 0.386 and 0.633 (milk revenue as the dependent variable) 0.447 and 0.728 (total revenue as the dependent variable). In both cases, the differences were found to be statistically significant at P=0.05 implying integrated vegetable-based systems are more efficient than milk-based systems, irrespective of the approach used. However, the study indicated that there are substantial inefficiencies prevailing in the cattle farming systems in Up Country Wet Zone and there is ample scope to increase both milk revenue and total revenue without using any additional inputs. Therefore, further studies should be carried out to understand the sources of inefficiencies using approaches such as the Battese and Coelli (1995) model. The overall results corroborate the findings of Mahipala (2002), which found that vegetable-based systems are more economically viable than milk-based systems.

The discrepancy between the distributions of technical efficiencies can be interpreted on theoretical grounds. DEA does not incorporate the measurement error and other statistical noises into the analysis. However, it is free from degrees of freedom problems and model specification errors. In contrast, technical efficiency estimation through SFA is sensitive to model specification errors but free from statistical noise. These theoretical distortions between DEA and SFA are depicted by the results.

## REFERENCES

Aigner, D.J. and Chu. S.F. (1968). On estimating industry production function. American Economic Review. 58:826-839.

Aigner, D.J., Lovell, C.A.K and Schmidt, P.J. (1977). Formulation and estimation of stochastic frontier production function models. Journal of Econometrics.6:21-37.

Battese, G.E. and Corra, G.S. (1977). Estimation of production frontier model: with application to the pastoral zone of Eastern Australia. Australian Journal of Agricultural Economics. 21(3):169-179.

Bravo-Uretra, B.E and Pinheiro, A.E. (1993). Efficiency analysis of developing country agriculture: a review of the frontier function literature. Agricultural and Resource Economics Review. 22:88-101.

Charnes, A., Cooper, W.W. and Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research. 2:429-444.

Coelli, T.J. (1994). A guide to FRONTIER version 4.1: A computer program for stochastic frontier production and cost function estimation. Department of Economics, University of New England, Armidale, Australia.

Coelli, T.J.(1995). Recent developments in frontier modeling and efficiency measurement. *Australian Journal of Agricultural Economics*. 39: 219-245.

Coelli, T.J. (1996). A guide to DEAP version 2.1: A computer program for data envelopment analysis. Department of Economics, University of New England, Armidale, Australia.

Department of Census and Statistics. (2001). Annual statistics handbook. Department of Census and Statistics, Colombo, Sri Lanka.

Farrell, M.J. (1957). The measurement of productive efficiency. Journal of Royal Statistical Society. 120(3): 253-281.

Heady, E.O. and Dillion, J.L. (1969). Agricultural Production Functions, Iowa State University Press, Ames, Iowa.

Ibrahim M.N.M., Staal, S.J.,Daniel. S.L.A and Thorpe, W. (1999). Appraisal of the Sri Lankan Dairy Sector, Synthesis Report and the Main Report, Ministry of Livestock, Colombo Sri Lanka.

Mahipala M.B.P. (2002). Comparison of milk-based and vegetable-based cattle farming systems in up-country wet-zone. Unpublished M.Sc. report. Postgraduate Institute of Agriculture, University of Peradeniya.

Meeusen, W. and van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composite error. International Economic Review. 18(2):435-444.

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