Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

A comparative analysis of multi-output frontier models*

Tao ZHANG†, Eoghan GARVEY
(*Department of Economics, National University of Ireland, Galway, Ireland*)
†E-mail: taozhang7608@hotmail.com
Received Feb. 18, 2008; revision accepted June 18, 2008

Abstract: Recently, there have been more debates on the methods of measuring efficiency. The main objective of this paper is to make a sensitivity analysis for different frontier models and compare the results obtained from the different methods of estimating multi-output frontier for a specific application. The methods include stochastic distance function frontier, stochastic ray frontier, and data envelopment analysis. The stochastic frontier regressions with and without the inefficiency effects model are also compared and tested. The results indicate that there are significant correlations between the results obtained from the alternative estimation methods.

Key words: Efficiency, Multi-output, Stochastic distance function frontier, Stochastic ray frontier, Data envelopment analysis (DEA), Correlation
doi:10.1631/jzus.A0820121 **Document code:** A **CLC number:** O212; O22

INTRODUCTION

Hitherto, we have not observed any studies on sensitivity analysis of frontier models for multi-output efficiency measures considering environmental factors. This paper hopes to shed new light on the comparison of empirical results for the selection of efficiency measurement tools and give some sensitivity analysis results to multi-output efficiency measures accounting for environmental factors. Here, we do not debate on the revenue function as it needs the price information, requires behavioral assumptions, and is too simple to discuss. We address multi-product efficiency for Irish farm households by data envelopment analysis (DEA), conditional DEA, multi-output stochastic distance function frontier model, and multi-output stochastic ray function frontier model. In fact, the meaning of results from this paper is outside the specific applications, and the generality of the findings in this paper can provide useful information for researchers who want to measure multi-output efficiency.

* Project supported by the Walsh Fellow Fund from RERC Teagasc

Traditionally, the methods to measure efficiency in production can be divided into two groups: one is linear programming model such as DEA, and the other is stochastic frontier analysis using econometric regression. However, both of these two approaches have a range of advantages and disadvantages probably influencing the results in a particular application. The principal advantage of the DEA approach is that it does not require the specification of a particular functional form for the technology, but it cannot measure the statistical noise. The principal advantage of the stochastic frontier analysis is that it considers the statistical noise and outliers, but it requires the assumed underlying technology and functional form. In addition, the non-parametric nature of the DEA approach makes it easy to handle multiple outputs and multiple inputs, but stochastic frontier analysis is limited by its assumed functional form and cannot be directly used for multi-output production analysis or multi-input cost analysis. Currently, some methods are developed to adjust the functional form of stochastic frontier and make it suitable for multi-product analysis, but they all have to change the functional form and probably make the influence of function assumption on results more strong. Some econometricians proposed that the distance function approach to handling multi-product stochastic frontier might introduce regressor endogeneity and induce estimator inconsistency in estimation (Grosskopf *et al.*, 1997). The principal aim of this study is to contribute to the stock of knowledge of the selection of estimation methods for multi-product efficiency measurement.

To investigate the sensitivity of estimated technical efficiency scores to various estimation methods, most of previous studies focused on one-output stochastic frontier analysis and DEA. Kopp and Smith (1980) compared the estimated results from one-output frontier production function using the parametric linear programming approach developed by Aigner and Chu (1968), a stochastic frontier estimated by corrected ordinary least squares (C-OLS) and a stochastic frontier estimated by maximum likelihood estimation. Fecher *et al*.(1993) compared a single-output DEA and a stochastic frontier estimation of production function for French insurance companies. Sharma *et al*.(1997) also compared a single-output DEA [CRS (constant returns to scale) and VRS (variable returns to scale)] with a stochastic frontier analysis for the productive efficiency of the swine industry in Hawaii. Huang and Wang (2002) compared the stochastic frontier analysis, the distribution-free approach (DFA), and DEA for Taiwanese banks over the period 1982 to 1997. Ruggiero (2007) compared the one-output technical efficiency estimations using DEA, Cobb-Douglas production function estimated by C-OLS method, Cobb-Douglas production function estimated by a stochastic frontier using maximum likelihood method, and Russell measurement of DEA.

Some other studies estimated the efficiency of a multi-output stochastic frontier using cost function frontier approach. Ferrier and Lovell (1990) compared DEA and multi-output cost stochastic frontier estimates. The first and the only observed work which investigated the sensitivity of estimated technical efficiency scores of multi-output production function and output-oriented DEA was the analysis by Coelli and Perelman (1999), which compared the technical efficiency scores from multi-output production function estimated by the parametric linear programming approach, multi-output oriented DEA, and C-OLS production function frontier. However, Coelli and Perelman (1999) did not compare the results from the multi-output stochastic frontier estimated by maximum likelihood approach, or considered the other methods which can be used to estimate multi-output stochastic frontier such as stochastic ray frontier. And they did not consider the multi-output efficiency measurement accounting for environmental factors.

Here, we write this paper to fill this gap, and compare the technical efficiency scores estimated from a multi-output stochastic production function frontier using maximum likelihood estimation and distance function approach, a multi-output stochastic ray production function frontier using maximum likelihood estimation and generalized ray production function approach developed by Lothgren (1997), and output-oriented DEA. This study will also consider the efficiency measurement accounting for environmental factors.

In addition, the current study makes a new contribution to the knowledge on sensitivity analyses, in that it appears to be the first study to consider the influence of dropped-variable choice on distance function approach. Actually, the distance function approach applied in multi-output stochastic frontier has one disadvantage which is not widely recognized. When economists define the production function which can be estimated directly by stochastic frontier model, they have to arbitrarily choose one output (or one input in input distance function analysis) which is used to change functional form and become the denominator for distance function and all other outputs. However, there is no test or method to help economists choose this output from multiple outputs. Obviously, arbitrarily chosen output will probably influence the results in a particular application. This paper will also compare the results from different output chosen functions to see if the results are sensitive to the arbitrarily chosen output as a denominator in the distance function.

FRONTIER MODEL

Stochastic frontier model

The stochastic frontier approach specifies the relationship between output and input levels using two error terms: one is the traditional normal error term with zero mean and constant variance, and the other represents technical inefficiency and can be expressed as a half-normal, truncated normal, exponential, or two-parameter gamma distribution. Technical efficiency is subsequently estimated via maximum likelihood of the production function subject to the two error terms. The stochastic frontier typically permits assessment of maximal output subject to input levels; as such, it appears to be an output-oriented measure. The inefficiency error term, and subsequently the maximal output, is specified as a function of inputs. Thus, it is possible to consider the input reduction coinciding with a fixed maximum or frontier output.

According to (Aigner *et al*., 1977), the original stochastic frontier model was proposed as a production function specified for cross-sectional data which had an error term with two components separately accounting for random error and technical inefficiency. The technical efficiency of the *i*th sample, denoted by TE_i, is defined in terms of the ratio of the observed output to the corresponding frontier output, conditional on the levels of inputs. The production function can be defined as

$$\boldsymbol{Y}_i=f(\boldsymbol{X}_i, \boldsymbol{\beta})\exp(V_i-U_i),\ \ i=1, 2, ..., n, \tag{1}$$

where $\boldsymbol{Y}_i$ denotes the production of the *i*th unit in the sample, $\boldsymbol{X}_i$ is a $1\times k$ vector of input quantities used by the *i*th unit, $\boldsymbol{\beta}$ is a $k\times 1$ vector of parameters to be estimated, $f(\boldsymbol{X}_i, \boldsymbol{\beta})$ is an appropriate parametric form for the underlying technology, V_i is assumed to be an independently and identically distributed $N(0, \sigma_v^2)$ random error, independent of U_i, and U_i is a non-negative random variable, associated with technical inefficiency in production. This model estimates the variance parameters of the likelihood function in terms of $\sigma^2=\sigma_u^2+\sigma_v^2$ and $\gamma=\sigma_u^2/\sigma^2$.

The technical efficiency score can be calculated as

$$TE_i=\exp(-U_i)=\boldsymbol{Y}_i/[f(\boldsymbol{X}_i, \boldsymbol{\beta})\exp(V_i)], \tag{2}$$

where $f(\boldsymbol{X}_i, \boldsymbol{\beta})\exp(V_i)$ is the stochastic frontier production. The prediction of technical efficiencies is based on the conditional expectation in the above expression, given the model specifications.

In recent years, the stochastic frontier model for the technical inefficiency effects accounting for the effects of various environmental variables on technical efficiency has been popular. As opposed to a traditional two-step procedure which is inconsistent with the assumption of independently and identically distributed technical inefficiency effects in the stochastic frontier, this model is in an econometrically consistent manner. The main advantage of this technique over the two-stage technique is that it incorporates environmental factors in the estimation of the production frontier because these factors may have a direct impact on efficiency. Battese and Coelli (1995)'s model for the cross-sectional data is defined in two equations. The first equation is the same as Eq.(1), with the exception that U_i is assumed to be independently distributed with truncations (at zero) of the normal distribution $N(\mu_i, \sigma_u^2)$. Under these assumptions, μ_i can more formally be expressed as follows:

$$\mu_i=\boldsymbol{Z}_i\boldsymbol{\delta}, \tag{3}$$

where $\boldsymbol{Z}_i$ is a $1\times m$ vector of observable farm-specific variables hypothesized to be associated with technical inefficiency, and $\boldsymbol{\delta}$ is an $m\times 1$ vector of unknown parameters to be estimated. Again, technical efficiency is given as $TE_i=\exp(-U_i)$.

Stochastic distance function frontier for multi-output technical efficiency

A major criticism of the stochastic frontier approach is that it cannot adequately handle multiple outputs. The framework of a stochastic distance function has been developed in the literature to respond to the criticisms. Fare *et al.*(1993) introduced the concept of 'distance functions' in expressing the output bundle of a multi-product technology. The distance function is specified as a function for fixed input and output levels. The technology is specified as a translog function. Coelli and Perelman (1996; 2000) discussed econometric estimation of stochastic distance functions with multiple outputs. Multi-output stochastic distance functions suffer from input-output separability and linear homogeneity in outputs.

The value of the distance function cannot be observed or directly estimated. Lovell *et al*.(1994) suggested a convenient approach circumventing this problem using the linear homogeneity of the distance function. That is,

$$D_i(\boldsymbol{X}, \beta\boldsymbol{Y})=\beta D_i(\boldsymbol{X}, \boldsymbol{Y}),\ \ \text{for any } \beta>0. \tag{4}$$

Here, $\boldsymbol{Y}$ is a vector of outputs of dimension L. Setting $\beta=1/Y_L$, where the arbitrarily chosen Y_L denotes the first component of $\boldsymbol{Y}$, Eq.(4) can be expressed in logarithmic form as

$$\ln D_i(\boldsymbol{X}, \boldsymbol{Y})=\ln Y_L+\ln D_i(\boldsymbol{X}, \boldsymbol{Y}/Y_L), \quad (5)$$

or

$$\ln D_i(\boldsymbol{X}, \boldsymbol{Y})-\ln Y_L=\ln D_i(\boldsymbol{X}, \boldsymbol{Y}/Y_L). \quad (6)$$

It is assumed that

$$\ln D_i(\boldsymbol{X}, \boldsymbol{Y})=-U_i. \quad (7)$$

Combining Eqs.(6) and (7) gives

$$-\ln Y_L=\ln D_i(\boldsymbol{X}, \boldsymbol{Y}/Y_L)+U_i. \quad (8)$$

In the empirical application, the translog production function is chosen. The translog production function of $\ln D_i(\boldsymbol{X}, \boldsymbol{Y})$ can be written as

$$\ln D_i(\boldsymbol{X},\boldsymbol{Y})=\alpha_0+\sum_{k=1}^{K}\alpha_k\ln X_k+\sum_{l=1}^{L}\beta_l\ln Y_l+\frac{1}{2}\sum_{k=1}^{K}\sum_{k'=1}^{K}\alpha_{kk'}\ln X_k\ln X_{k'}+\frac{1}{2}\sum_{l=1}^{L}\sum_{l'=1}^{L}\beta_{ll'}\ln Y_l\ln Y_{l'}+\sum_{k=1}^{K}\sum_{l=1}^{L}\gamma_{kl}\ln X_k\ln Y_l, \quad (9)$$

where i refers to the ith unit; $\boldsymbol{X}$ is the input and $\boldsymbol{Y}$ is the output; α, β, γ are parameters to be estimated.

Here, the homogeneity restrictions are

$$\sum_{l=1}^{L}\beta_l=1,\quad \sum_{l=1}^{L}\beta_{ll}=0,\quad \sum_{l=1}^{L}\gamma_{kl}=0,$$

and the required conditions for symmetry are

$$\alpha_{kk'}=\alpha_{k'k},\ \beta_{kk'}=\beta_{k'k}.$$

Then according to Eq.(4), we can give

$$\ln(D_i(\boldsymbol{X},\boldsymbol{Y})/Y_L)=\alpha_0+\sum_{k=1}^{K}\alpha_k\ln X_k+\sum_{l=1}^{L-1}\beta_l\ln(Y_l/Y_L)+\frac{1}{2}\sum_{k=1}^{K}\sum_{k'=1}^{K}\alpha_{kk'}\ln X_k\ln X_{k'}+\sum_{k=1}^{K}\sum_{l=1}^{L-1}\gamma_{kl}\ln X_k\ln(Y_l/Y_L)+\frac{1}{2}\sum_{l=1}^{L-1}\sum_{l'=1}^{L-1}\beta_{ll'}\ln(Y_l/Y_L)\ln(Y_{l'}/Y_L). \quad (10)$$

Finally, the functional form can be rewritten as

$$-\ln Y_L=\alpha_0+\sum_{k=1}^{K}\alpha_k\ln X_k+\sum_{l=1}^{L-1}\beta_l\ln(Y_l/Y_L)+\frac{1}{2}\sum_{k=1}^{K}\sum_{k'=1}^{K}\alpha_{kk'}\ln X_k\ln X_{k'}+\sum_{k=1}^{K}\sum_{l=1}^{L-1}\gamma_{kl}\ln X_k\ln(Y_l/Y_L)+U_i+\frac{1}{2}\sum_{l=1}^{L-1}\sum_{l'=1}^{L-1}\beta_{ll'}\ln(Y_l/Y_L)\ln(Y_{l'}/Y_L). \quad (11)$$

Again, the technical efficiency can be calculated as $TE_i=\exp(-U_i)=D_i(\boldsymbol{X}, \boldsymbol{Y})$, where TE_i changes from 0 to 1, and $\ln D_i(\boldsymbol{X}, \boldsymbol{Y})\leq 0$.

Stochastic ray frontier for multi-product efficiency measurement

Lothgren (1997) generalized the multi-output ray production function using a polar-coordinate angle output vector. Lothgren used the Euclidean norm of the output vector, which can be written as

$$\boldsymbol{Y}=\|\boldsymbol{Y}\|\cdot m(\theta), \quad (12)$$

where $\|\boldsymbol{Y}\|=\left(\sum_{i=1}^{P}Y_i^2\right)^{1/2}$, and P is the number of outputs.

Eq.(12) can be rewritten as

$$m(\theta)=\boldsymbol{Y}/\|\boldsymbol{Y}\|. \quad (13)$$

The polar coordinate angle can be easily obtained as

$$\theta_i(\boldsymbol{Y})=\arccos\left(Y_i\Big/\left(\|\boldsymbol{Y}\|\prod_{j=0}^{i-1}\sin\theta_j\right)\right), \quad (14)$$

where $i=1, 2, \ldots, P-1$.

Based on the assumption that the error term affects the output vector multiplicatively, according to (Lothgren, 1997; 2000), the stochastic frontier model can be specified as

$$\boldsymbol{Y}=f(\boldsymbol{X}, \theta)m(\theta)\exp(V-U), \quad (15)$$

where $\boldsymbol{X}$ is the input, θ denotes the polar-coordinate angle, V is the error component, and U is the inefficiency component.

The frontier output given by the ray production function can be written as

$$\boldsymbol{Y}^{\mathrm{f}}=f(\boldsymbol{X},\theta)m(\theta). \tag{16}$$

Here, $\boldsymbol{Y}/\boldsymbol{Y}^{\mathrm{f}}=\exp(V-U)$ denotes traditional output distance function $D(\boldsymbol{X}, \boldsymbol{Y})$.

We can combine Eqs.(13) and (15) to yield

$$\|\boldsymbol{Y}\|=f(\boldsymbol{X},\theta)\exp(V-U). \tag{17}$$

Then, in empirical analysis, the natural log-linear stochastic ray translog production frontier function can be written as

$$\begin{aligned}\ln\|\boldsymbol{Y}\| &= \alpha_0+\sum_{k=1}^{K}\alpha_k\ln X_k+\sum_{l=1}^{L}\beta_l\ln\theta_l \\ &+\frac{1}{2}\sum_{k=1}^{K}\sum_{k'=1}^{K}\alpha_{kk'}\ln X_k\ln X_{k'}+\frac{1}{2}\sum_{l=1}^{L}\sum_{l'=1}^{L}\beta_{ll'}\ln\theta_l\ln\theta_{l'} \\ &+\sum_{k=1}^{K}\sum_{l=1}^{L}\gamma_{kl}\ln X_k\ln\theta_l-U,\end{aligned} \tag{18}$$

where θ is the polar-coordinate angle, K denotes the number of inputs, $L=P-1$, and P is the number of outputs.

DEA method

DEA, the non-parametric approach to measuring efficiency, was first introduced in the literature as a linear programming model by Charnes *et al.*(1978), following on Farrell (1957)'s posing of the question of relative technical efficiency in the form of a unit isoquant model. Generally, the DEA approach defines the technical efficiency in terms of a minimum set of inputs needed to produce a given output known as 'input-oriented model' or maximum output obtainable from a given set of inputs known as 'output-oriented model' (Charnes *et al.*, 1994).

The measurement of efficiency is the value of the objective function, which measures CRS technical efficiency. If a further constraint, i.e., the sum of the output weights must be equal to 1, is added, VRS is allowed. This ensures that technical efficiencies of units are compared with efficiencies of other units in a similar size. Therefore, the data envelope fits closer, and VRS technical efficiency measures are higher than or equal to CRS technical efficiencies. In the actual calculations, we will use the dual function of the DEA model. The dual VRS output-oriented DEA technical efficiency can be rewritten as

$$\theta_{\mathrm{DEA}}(\boldsymbol{X},\boldsymbol{Y})=\sup\left\{\theta\,\middle|\,\sum_i\lambda_i X_{i,m}\le X_m;\ \sum_i\lambda_i Y_{i,j}\ge\theta Y_j;\ \text{s.t.}\ \sum_i\lambda_i=1;\ \lambda_i\ge 0\right\}. \tag{19}$$

Here, λ_i are output and input weights; $1/\theta$ is the output-oriented technical efficiency; i denotes the number of observations; j and m denote the number of outputs and inputs, respectively.

To provide a more 'fair' assessment of DEA, we also consider the conditional DEA accounting for environmental factors. Daraio and Simar (2005; 2007) developed the conditional FDH (free disposal hull) by extending it to conditional DEA, and clearly described the input-oriented conditional DEA, which can be easily extended to output-oriented conditional DEA. The VRS output-oriented conditional DEA function can be written as

$$\begin{aligned}&\theta_{\mathrm{DEA}}(\boldsymbol{X},\boldsymbol{Y}\mid\boldsymbol{Z})\\&=\sup\left\{\theta\,\middle|\,\sum_{\{i\mid z-h\le Z_{ii}\le z+h\}}\lambda_i X_{i,m}\le X_m;\ \sum_{\{i\mid z-h\le Z_{ii}\le z+h\}}\lambda_i Y_{i,j}\ge\theta Y_j;\right.\\&\qquad\left.\text{s.t.}\sum_{\{i\mid z-h\le Z_{ii}\le z+h\}}\lambda_i=1;\ \lambda_i\ge 0\right\}.\end{aligned} \tag{20}$$

Here, i still denotes the number of observations, h denotes the chosen bandwidth, and Z_i denotes the external factors which may influence efficiency. The bandwidth is estimated by a k-nearest neighbour method. According to (Daraio and Simar, 2005; 2007), we can evaluate the leave-one-out kernel density estimation of Z_i and find the value of k as well as local bandwidth.

DATA AND VARIABLES

All the raw data come from farm survey managed by Teagasc. There are cross-sectional data for year 2004. The number of observations is 606 farms. We consider two outputs—farm output and off-farm work income, and two groups of inputs—farm inputs and off-farm inputs. The farm output chosen is the total farm output. The subsidies which are directly related to the production are also included in the total farm output (cattle and dairy subsidies, sheep subsidies, and crop subsidies). The off-farm output is off-farm income. The farm input includes utilized land, labour input, total livestock direct costs, and total crops direct costs. The off-farm input is only off-farm work time.

Because the dataset in this study involves 606 observations at one time point under the similar environmental conditions, it is suitable to illustrate the sensitivity of technical efficiency estimation methods. We will also estimate the frontier model with inefficiency effects function considering environmental factors. It should be noted that the empirical analysis included in this paper is mainly to illustrate the sensitivity of technical efficiency predictions to the selection of various methodologies and is not a complete and thorough analysis of technical efficiency. The environmental variables used in inefficiency effects model include farm-type dummies (cattle and sheep; other farm-type dummies are dropped), soil code (here, a higher soil code value indicates a lower soil quality), insurance, and consultant fees paid by farmers.

RESULTS

Table 1 depicts the estimated coefficients from stochastic ray frontier regression. The likelihood ratio test statistic can be given by $\lambda=2(\ln L_1-\ln L_0)$, where $\ln L_1$ and $\ln L_0$ denote the log-likelihood values for the regression results. From Table 1, the likelihood ratio test statistic for the above two stochastic ray frontier regressions is 74.7 and the critical value with q+1 (q is the number of parameters constrained to zero) degrees of freedom is 16.81 for P=0.01, indicating that the regression result of stochastic frontier not considering environmental variables should be rejected.

The coefficient estimates of the polar-coordinate angles are statistically significant at P=0.01 for both regression results. The coefficient of labor-day input is significant at P=0.01, but that of quadratic labor-day input is insignificant. The estimated coefficients for the other direct inputs are all statistically insignificant, but those for the quadratic values of the other direct inputs are all statistically significant at P=0.05 or 0.01. For the technical inefficiency effects model, all the coefficients are statistically significant. The estimated farm-type dummy coefficients for cattle and sheep are positive, indicating that these two farm types are relatively technically inefficient compared with other farm types. The soil code coefficient estimate is positive, indicating that the high quality of soil can reduce the technical inefficiency. The coefficient estimates for insurance and consultant fees are negative, suggesting that the insurance and consultants have reduced the technical inefficiency.

In Table 2, the coefficients estimated by stochastic distance function frontier regressions are reported. It should be noted that, in this paper, the different results from two arbitrarily chosen outputs as the denominators in distance functions are compared to see if the results are sensitive to the chosen output. In this paper, the arbitrarily chosen outputs are farm output and work income for two stochastic distance function frontier regressions. Because no pertinent methods are available to test the two regression results estimated from different output-chosen distance functions, we simply calculate the correlation between them. The estimated correlations between two sets of parameters for stochastic frontier with and without inefficiency effects model are 0.169 and 0.241, respectively, showing that there is enormous difference between the results from two regressions choosing different outputs as the denominators. Although the log likelihood values of work-income-chosen stochastic frontier regressions both considering and not considering environmental variables are higher than those of two stochastic frontier regressions choosing farm output as the denominator, this cannot give any direct information about which regression is better because one of them is not nested within the other. Furthermore, an interesting thing is that the estimated coefficients of farm size (farm land input) are always highly significant while the coefficients of the only input for off-farm work (work hours) are always insignificant for the stochastic frontier regressions choosing farm output as the denominator. However, as for the regression results of two stochastic frontier models choosing work income as the denominator, the status is in the opposite way. Thus, it is found that the estimated coefficients of stochastic distance function frontier are sensitive to the distance functional form change and arbitrarily output-chosen process. However, although the stochastic distance function frontier for multi-output model has been widely used, it is rarely noted that we should carefully choose the denominator in distance functional form change.

Table 1 Estimated coefficients from stochastic ray frontier

Parameter	Two-function regression with μ		One-function regression	
	Coefficient	Standard error	Coefficient	Standard error
Labor days d_l	1.699***	0.562	1.744***	0.616
Livestock costs c_l	−0.120	0.167	−0.230	0.194
Crop costs c_c	0.133	0.105	0.212*	0.122
Work hours t_w	−0.170	0.526	−0.596	0.570
Farm size s_f	0.286	0.633	0.294	0.659
s_f^2	−0.624**	0.268	−0.607**	0.254
t_w^2	0.776***	0.159	0.912***	0.174
c_c^2	0.088***	0.019	0.100***	0.022
c_l^2	0.149***	0.037	0.163***	0.045
d_l^2	0.237	0.234	0.368	0.241
$s_f{\cdot}t_w$	0.247	0.189	0.329*	0.199
$s_f{\cdot}c_c$	0.020	0.038	0.047	0.043
$s_f{\cdot}c_l$	−0.277***	0.082	−0.299***	0.090
$s_f{\cdot}d_l$	0.546**	0.223	0.435**	0.221
$t_w{\cdot}c_c$	−0.036	0.031	−0.051	0.036
$t_w{\cdot}c_l$	0.061	0.058	0.080	0.067
$t_w{\cdot}d_l$	−1.034***	0.164	−1.085***	0.180
$c_c{\cdot}c_l$	−0.080***	0.023	−0.084***	0.026
$c_c{\cdot}d_l$	0.063*	0.037	0.039	0.041
$c_l{\cdot}d_l$	0.150*	0.083	0.181**	0.095
θ	0.769***	0.210	0.738***	0.242
θ^2	0.030	0.026	0.038	0.030
$\theta{\cdot}d_l$	0.070	0.093	0.098	0.102
$\theta{\cdot}c_l$	−0.082***	0.030	−0.091***	0.035
$\theta{\cdot}t_w$	−0.112	0.076	−0.099	0.088
$\theta{\cdot}c_c$	0.006	0.013	0.012	0.016
$\theta{\cdot}s_f$	−0.186**	0.088	−0.215**	0.097
Constant coefficient	6.591***	1.111	7.011***	1.213
Parameters for μ				
Farm type−cattle	0.207***	0.070		
Farm type−sheep	0.173*	0.099		
Soil code	0.036**	0.018		
Insurance	−0.026***	0.010		
Consultant fees	−0.026**	0.012		
Log likelihood	106.476		69.124	
σ^2	0.095		0.077	

For *Z*-test, significant at * *P*=0.10, ** *P*=0.05, *** *P*=0.01

Again, the likelihood ratio test is used to test the regression results between stochastic distance function frontiers with and without inefficiency effects model. The likelihood ratio test statistic for two stochastic distance function frontier regressions choosing farm output as the denominator is 46.91, indicating that the regression result of stochastic frontier not considering environmental variables has been rejected. The likelihood ratio test between two stochastic distance function regressions choosing work income as the denominator yields χ^2=72.67, again showing that the regression result of stochasticfrontier with the inefficiency effects model is better than that without the inefficiency effects model. For all four stochastic distance function frontier regressions, about 2/3 or more of the estimated coefficients

Table 2 Estimated coefficients from stochastic distance function frontier regressions

Parameter	Two-function regression with μ		One-function regression	
	Farm output as Y_L	Work income as Y_L	Farm output as Y_L	Work income as Y_L
Farm size s_f	−2.486***	0.105	−2.488***	−0.053
Work hours t_w	0.233	−0.836***	0.071	−1.075***
Crop costs c_c	0.127*	0.092	0.149**	0.138**
Livestock costs c_l	0.400***	−0.312***	0.379***	−0.381***
Labor days d_l	1.816***	1.975***	1.802***	2.122***
s_f^2	−0.395***	−0.255**	−0.420***	−0.278***
t_w^2	0.164***	0.346***	0.206***	0.407***
c_c^2	0.042***	0.042***	0.045***	0.046***
c_l^2	0.080***	0.070***	0.084***	0.074***
d_l^2	0.675***	0.315***	0.682***	0.378***
$s_f \cdot t_w$	0.510***	0.063	0.530***	0.107
$s_f \cdot c_c$	−0.013	0.015	−0.010	0.023
$s_f \cdot c_l$	−0.077*	−0.100***	−0.088**	−0.105***
$s_f \cdot d_l$	−0.320***	0.227***	−0.293***	0.217***
$t_w \cdot c_c$	−0.014	−0.013	−0.015	−0.018
$t_w \cdot c_l$	−0.077***	0.038*	−0.073***	0.048**
$t_w \cdot d_l$	−0.411***	−0.531***	−0.426***	−0.565***
$c_c \cdot c_l$	−0.039***	−0.042***	−0.040***	−0.044***
$c_c \cdot d_l$	0.047***	0.026*	0.039**	0.019
$c_l \cdot d_l$	0.197***	0.048	0.204***	0.051
Y_l/Y_L	0.016	0.083	0.021	0.077
$(Y_l/Y_L)^2$	0.002	0.003	0.006	0.009
$(Y_l/Y_L)\cdot s_f$	−0.708***	0.046	−0.699***	0.069*
$(Y_l/Y_L)\cdot t_w$	0.010	0.244***	0.002	0.255***
$(Y_l/Y_L)\cdot c_c$	0.017	−0.006	0.019	−0.011
$(Y_l/Y_L)\cdot c_l$	0.117***	0.073***	0.115***	0.082***
$(Y_l/Y_L)\cdot d_l$	0.564***	−0.356***	0.564***	−0.396***
Constant coefficient	−0.664	2.269***	−0.477	2.579***
Parameters for μ				
Farm type−cattle	0.059	0.061**		
Farm type−sheep	0.032	0.050		
Soil code	0.014*	0.019***		
Insurance	−0.009**	−0.008**		
Consultant fees	−0.010*	−0.011**		
Log likelihood	584.511	667.752	561.058	631.418
σ^2	0.0188	0.0142	0.0167	0.0128

For *Z*-test, significant at * P=0.10, ** P=0.05, *** P=0.01

are statistically significant. For the technical inefficiency effects model, the estimated farm-type dummy coefficient for cattle is only significant for the model choosing work income as the denominator, and all the other coefficients of farm-type dummies are insignificant. The coefficients of soil code, insurance and consultants are significant for both regressions. The trend of estimated coefficients in inefficiency effects model of stochastic distance function frontier is the same as that of stochastic ray frontier.

Table 3 depicts the summary statistics of estimated efficiencies from eight different models. The mean of technical efficiencies estimated from DEA model is 0.749, indicating that there is substantial inefficiency in Irish farm households. The mean technical efficiencies for the sample farm households obtained from the stochastic ray frontier regressions considering and not considering environmental effects are 0.835 and 0.849, respectively. The conditional DEA yields a mean efficiency of 0.890. Thus, the stochastic ray frontier analyses and conditional DEA also reveal considerable inefficiency in Irish farm households. However, the mean technical efficiencies from four stochastic distance function frontier regressions range from 0.920 to 0.930, showing that there is inefficiency in Irish farm households but not too high. The minimum values of technical efficiencies estimated from four stochastic distance function frontier models range from 0.595 to 0.654, which suggests that nearly all the observations have the efficiency higher than 0.6. However, the minimum value of technical efficiencies from DEA is only 0.202, which is closer to the minimum efficiency value estimated from stochastic ray frontier with inefficiency effects model (0.296). The minimum efficiency value estimated from stochastic ray frontier without inefficiency effects model is very close to that from conditional DEA.

The three approaches used above yield eight sets of technical efficiency estimates of producers relative to different production frontiers. As shown in Table 3, there are differences, although not distinct, in technical efficiencies estimated by the different methods. The efficiency estimates from DEA (TE_7) and stochastic ray frontier (TE_6) show a significantly higher variability than the other stochastic efficiency measures. To test the agreement among the different approaches, Spearman correlation coefficients are calculated and reported in Table 4.

All the correlation coefficients are positive and significant. The correlation coefficients between TE_1 and TE_4, between TE_2 and TE_3, and between TE_5 and TE_6 measuring the relationship between efficiencies estimated from stochastic frontier regressions with and without inefficiency effects model are 0.843, 0.866, and 0.823, respectively, showing that there are some differences in efficiency measures estimated from the stochastic frontier with and without inefficiency effects model.

Table 3 Statistics of efficiency predictions from eight specified models for 606 observations

Technical efficiency	Estimated from	Mean	Std. Dev.	Min	Max
TE_1	Stochastic distance function frontier without inefficiency effects model choosing work income as Y_L	0.931	0.034	0.654	0.985
TE_2	Stochastic distance function frontier without inefficiency effects model choosing farm output as Y_L	0.921	0.040	0.652	0.981
TE_3	Stochastic distance function frontier with inefficiency effects model choosing farm output as Y_L	0.924	0.051	0.612	0.993
TE_4	Stochastic distance function frontier with inefficiency effects model choosing work income as Y_L	0.929	0.052	0.595	0.995
TE_5	Stochastic ray frontier without inefficiency effects model	0.849	0.064	0.416	0.967
TE_6	Stochastic ray frontier with inefficiency effects model	0.835	0.112	0.296	0.988
TE_7	DEA	0.749	0.175	0.202	1.000
TE_8	Conditional DEA	0.891	0.128	0.427	1.000

Table 4 Correlation coefficients for eight sets of efficiency estimates

	TE_1	TE_2	TE_3	TE_4	TE_5	TE_6	TE_7
TE_1	1.000						
TE_2	0.877	1.000					
TE_3	0.799	0.866	1.000				
TE_4	0.843	0.727	0.922	1.000			
TE_5	0.907	0.783	0.716	0.771	1.000		
TE_6	0.749	0.626	0.837	0.927	0.823	1.000	
TE_7	0.635	0.523	0.579	0.644	0.667	0.703	1.000
TE_8	0.684	0.603	0.612	0.698	0.716	0.627	0.622

Note: TE_1 to TE_8 are the same as those defined in Table 3

The correlation coefficients between TE_1 and TE_2 and between TE_3 and TE_4 are respectively 0.877 and 0.922, indicating that there are differences between efficiencies estimated from the stochastic frontier choosing farm output as the denominator and those choosing work income as the denominator, although the differences are not very large. Compared with the correlation coefficients for the efficiencies, all estimated from parametric frontier, correlations between DEA and the six sets of efficiencies (ranging from 0.523 to 0.703) from parametric methods are not very robust, but still significant. The correlation between basic DEA and conditional DEA is positive and significant. It is obvious that compared with basic DEA, the conditional DEA has higher estimated efficiency scores and its results have higher correlations with those of stochastic frontier.

CONCLUSION

This paper compares the performance of stochastic distance function, stochastic ray, and DEA production frontiers in estimating technical efficiencies for a sample of Irish farm households. The correlations among the various sets of technical efficiency estimates are all positive and significant, which indicates that these approaches provide similar information on the relative productive performance of the Irish farm households in this study. However, the mean of the technical efficiencies estimated from the stochastic distance function frontier is larger than that obtained from the stochastic ray frontier, and in turn the mean technical efficiency in the stochastic ray frontier is higher than that estimated from the DEA analyses. It is also shown that the mean technical efficiency estimated from the conditional DEA is higher than that estimated from the basic DEA.

From the correlations of different technical efficiency predictions, it seems that any of these methods can be selected if researchers do not need to concern too much about the potential influence upon results from a different choice. However, according to the implications of the results in this paper, it is suggested that a sensitivity analysis for different efficiency measurement methods should be probably necessary to obtain robust conclusions in specific applications. If the stochastic distance function frontier is chosen, one had better compare the different results from different choices of output as the denominator in case there are some differences in various efficiency predictions.

References

Aigner, D.J., Chu, S.F., 1968. On estimating the industry production function. *Am. Econ. Rev.*, **58**(4):826-839.

Aigner, D., Lovell, K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier models. *J. Econ.*, **6**(1):21-37. [doi:10.1016/0304-4076(77)90052-5]

Battese, G., Coelli, T., 1995. A model for technical efficiency effects in a stochastic frontier production function for panel data. *Emp. Econ.*, **20**(2):325-332. [doi:10.1007/BF01205442]

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. *Eur. J. Oper. Res.*, **2**(6):429-444. [doi:10.1016/0377-2217(78)90138-8]

Charnes, A., Cooper, W.W., Lewin, A.Y., Seiford, L.M. (Eds.), 1994. Categorical Inputs and Outputs, Section 3.3. *In:* Data Envelopment Analysis: Theory, Methodology, and Applications. Kluwer Academic Publishers, Boston/Dordrecht/London, p.52-54.

Coelli, T., Perelman, S., 1996. Efficiency Measurement, Multiple-output Technologies and Distance Functions: With Application to European Railways. CREPP Working Paper, University of Liege, Liege, Wallonia, Belgium.

Coelli, T., Perelman, S., 1999. A comparison of parametric and non-parametric distance functions: with application to European railways. *Eur. J. Oper. Res.*, **117**(2):326-339. [doi:10.1016/S0377-2217(98)00271-9]

Coelli, T., Perelman, S., 2000. Technical efficiency of European railways: a distance function approach. *Appl. Econ.*, **32**(15):1967-1976. [doi:10.1080/00036840050155896]

Daraio, C., Simar, L., 2005. Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *J. Prod. Anal.*, **24**(1):93-121. [doi:10.1007/s11123-005-3042-8]

Daraio, C., Simar, L., 2007. Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. *J. Prod. Anal.*, **28**(1-2):13-32. [doi:10.1007/s11123-007-0049-3]

Fare, R., Grosskopf, S., Lovell, C.A.K., Yaisawarng, S., 1993. Derivation of shadow prices for undesirable outputs: a distance function approach. *Rev. Econ. Stat.*, **75**(2):374-380. [doi:10.2307/2109448]

Fecher, F., Kessler, D., Perelman, S., Pestieau, P., 1993. Productive performance of the French insurance industry. *J. Prod. Anal.*, **4**(1-2):77-93. [doi:10.1007/BF01073467]

Ferrier, G.D., Lovell, C.A.K., 1990. Measuring cost efficiency in banking: econometric and linear programming evidence. *J. Econ.*, **46**(1-2):229-245. [doi:10.1016/0304-4076(90)90057-Z]

Grosskopf, S., Hayes, K., Taylor, L., Weber, W., 1997. Budget constrained frontier measurement of fiscal equality and efficiency in schooling. *Rev. Econ. Stat.*, **79**(1):116-124.

[doi:10.1162/003465397556458]

Huang, T.H., Wang, M.H., 2002. Comparison of economic efficiency estimation methods: parametric and non-parametric techniques. *The Manchester School*, **70**(5): 682-709. [doi:10.1111/1467-9957.00320]

Kopp, R., Smith, V., 1980. Frontier production function estimates for steam electric generation: a comparative analysis. *South. Econ. J.*, **46**(4):1049-1059. [doi:10.2307/1057240]

Lothgren, M., 1997. Generalized stochastic frontier production models. *Econ. Lett.*, **57**(3):255-259. [doi:10.1016/S0165-1765(97)00246-2]

Lothgren, M., 2000. Specification and estimation of stochastic multiple-output production and technical inefficiency. *Appl. Econ.*, **32**(12):1533-1540. [doi:10.1080/000368400418943]

Lovell, C.A.K., Richardson, S., Travers, P., Wood, L.L., 1994. Resources and Functionings: A New View of Inequality in Australia. *In*: Eichhorn, W. (Ed.), Models and Measurement of Welfare and Inequality. Springer-Verlag, Berlin, p.787-807.

Ruggiero, J., 2007. A comparison of DEA and the stochastic frontier model using panel data. *Int. Trans. Oper. Res.*, **14**(3):259-266. [doi:10.1111/j.1475-3995.2007.00585.x]

Sharma, K.R., Leung, P., Zaleski, H.M., 1997. Productive efficiency of the swine industry in Hawaii: stochastic frontier vs. data envelopment analysis. *J. Prod. Anal.*, **8**(4):447-459. [doi:10.1023/A:1007744327504]