SENSITIVITY OF TECHNICAL EFFICIENCY ESTIMATES TO ESTIMATION METHODS: AN EMPIRICAL COMPARISON OF PARAMETRIC AND NON-PARAMETRIC APPROACHES

Henry de-Graft Acquah

Department of Agricultural Economics and Extension
University of Cape Coast, Ghana
Email: henrydegraftacquah@yahoo.com
Telephone: 00233 245 543956

Abstract: This paper highlights the sensitivity of technical efficiency estimates to estimation approaches using empirical data. Firm specific technical efficiency and mean technical efficiency are estimated using the non parametric Data Envelope Analysis (DEA) and the parametric Corrected Ordinary Least Squares (COLS) and Stochastic Frontier Analysis (SFA) approaches. Mean technical efficiency is found to be sensitive to the choice of estimation technique. Analysis of variance and Tukey's test suggests significant differences in means between efficiency scores from different methods. In general the DEA and SFA frontiers resulted in higher mean technical efficiency estimates than the COLS production frontier. The efficiency estimates of the DEA have the smallest variability when compared with the SFA and COLS. There exists a strong positive correlation between the efficiency estimates based on the three methods.

Keywords: .Technical Efficiency, Stochastic Frontier Analysis, Deterministic Frontier Analysis, Data Envelope Analysis, Tukey's Test.

Introduction

Advances in the productivity and efficiency literature have led to the development of various methods of measuring efficiency. The two most widely used approaches to evaluate the efficiency of decision making units (DMU) are the non-parametric Data Envelope Analysis (DEA) and Parametric Stochastic Frontier Analysis (SFA).

There is an on-going debate over these approaches and productivity researchers tend to have strong preference over which method to use for efficiency estimation. Majority of studies measuring technical efficiency using frontier methodology usually uses only one of the above methods at a time to estimate the production function and efficiency. Although there is a considerable amount of literature in the field of efficiency in production, only a small proportion of this literature is dedicated to comparison of measurement methods of technical efficiency. Furthermore, the findings of the few studies that investigated sensitivity of technical efficiency estimates to different methods are mixed; hence there is the need for more research to focus on comparing technical efficiency measurements from alternative models in order to determine the robustness of estimates from a particular model.

In effect, this study is by no means the first to investigate the sensitivity of technical efficiency estimates to estimation methods. However this study is significant, in the sense that it appears to be the first comparative study of frontier estimation methodologies using a well-known data set (i.e. frontier 4.1 data set). This is a departure from the numerous previous studies using lesser known empirical data for such a comparison. Furthermore, this contribution adds up to the few existing studies that shed light on the sensitivity of empirical results to the selection of the estimation techniques. It is against this background that this study investigates the sensitivity of technical efficiency predictions to estimation technique. The primary objective of this paper is to investigate the sensitivity of technical efficiency estimates to estimation techniques using the frontier 4.1 data set.

Methodology

Researchers have developed several approaches to measure technical efficiency. Based on Farrell's (1957) pioneering article, both parametric and non-parametric techniques measuring efficiency has been developed. Among the numerous approaches, the Stochastic Frontier Analysis (SFA) and the Data Envelope Analysis (DEA) are two approaches that have been heavily used in the estimation of technical efficiency in production. Preceding the stochastic frontier model are the deterministic parametric frontier models of which the corrected ordinary least squares (COLS) approach is widely used.

The stochastic frontier approach was developed by Aigner, Lovell and Schmidt (1977) and Meeusen and Van den Broeck (1977) and the Data Envelope Analysis (DEA) was developed by Charnes, Cooper and Rhodes (1978). Subsequently, numerous authors (Kumbhaker and Lovell, 2000), Seiford and Thrall (1990), Fried et al (1993), Coelli, Rao and Battese (1998), Bravo-Ureta and Pinheiro (1993), Coelli (1995) and Cooper et al. (2000) have reviewed these approaches in the economic literature.

A major difference between the parametric and non-parametric approaches is the estimation principle. The DEA method relies on the idea of minimal extrapolation. The main advantage of the DEA is that it does not require specification of a functional form of the production function. The DEA may be applied to multiple outputs and multiple inputs with each being stated in different units. DEA is deterministic and attributes all deviations from the frontier to inefficiencies. The main disadvantage is that it is not possible to estimate parameters for the model and hence impossible to test hypothesis concerning the performance of the model. However, recently, bootstrap methods have been employed to obtain measures of statistical precision in the DEA model.

Alternatively, parametric stochastic frontier models assume that deviations from the model can be due to both noise and inefficiency. It also assumes the production function has a functional form. The principal advantage of the SFA is that it allows the test for hypothesis concerning the goodness of fit of the model. Whilst, the main disadvantage is that it requires pre-specification of the functional form and a distributional assumption for technical inefficiency.

Alternative Approaches to Technical Efficiency Estimation

This section discusses parametric and non-parametric frontier estimation. The non-parametric methods emphasize the Data Envelope Analysis (DEA) and the parametric approach emphasizes the two most commonly employed parametric alternatives: Deterministic Frontier Analysis (DFA) and Stochastic Frontier Analysis (SFA).

Deterministic Frontier Analysis (DFA)

Consider the production function below:

$$y^k = f\left(x^k;\beta\right) - u^k,\ \ u^k \overset{iid}{\sim} H,\ \ k = 1,\ldots,k, \tag{1}$$

Where H is some probability distribution with support only on R_+. The above model assumes that all deviations are the result of inefficiency. Noticeably, the deterministic model assumes that there is no noise in the data like a DEA model. The above equation can be estimated using the OLS. However there are problems in using OLS to estimate this production function. Greene (1980), notes that the OLS estimator is biased downwards in this estimation. As a result of this problem, it is possible for the estimated residuals of the model to have the incorrect signs. However, since the calculations of technical efficiency relies on these residuals being non negative, Greene (2008) suggest a correction for this bias by shifting beta hat, $\hat{\beta}$ the OLS estimator of β_0 upward by the largest positive residual. This two-step procedure is known as the Corrected Ordinary Least Squares (COLS) method. The COLS involves two steps.

First is to make an ordinary least square estimate of the value of β

$$\min_{\beta} \sum_{k=1}^{K} \left(y^k - f\left(x^k;\beta\right)\right)^2 \tag{2}$$

Second, find the smallest possible correction of the intercept $\beta_0$to β_{00} to ensure that all observations are below the production frontier. In short adjust β_0 upward with the maximum error term.

$$\beta_{00} = \max\left\{ y^k - f(x^k;\hat{\beta}) \middle| k = 1,\ldots,K \right\} \tag{3}$$

Stochastic Frontier Analysis (SFA)

The stochastic frontier model includes both a stochastic error term and a term that can be characterized as inefficiency. The model can be specified as follows:

$$y^k = f(x^k;\beta) + v^k - u^k$$

$$v^k \sim N(0,\sigma_v^2),\ \ u^k \sim N_+(0,\sigma_u^2),\ \ k = 1\ldots,K \tag{4}$$

The v term takes care of the stochastic nature of the production process and possible measurements errors of inputs and outputs and the u term is the possible inefficiency of the firm. We assume that the term v and u are independent. If u=0, the firm is 100% efficient and if u>0 there is some inefficiency. The N_+ denotes a half normal distribution. That is a truncated normal distribution where the point of truncation is 0 and the distribution is concentrated on the half-interval.

Data Envelopment Analysis

DEA is a linear programming based technique for measuring the relative performance of organizational units where the presence of multiple inputs and outputs make comparisons difficult.

Assuming that there are n DMU, each with m inputs and s outputs, the relative efficiency score of a test DMU p is obtained by solving the following model proposed by Charnes et al. (1978):

$$\max \frac{\sum_{k=1}^{s} v_k y_{kp}}{\sum_{j=1}^{m} u_j x_{jp}}$$

$$\frac{\sum_{K=1}^{s} v_k y_{ki}}{\sum_{j=1}^{m} u_j x_{ji}} \leq 1 \forall i$$

$$v_k, u_j \geq 0 \forall k, j \tag{5}$$

Where

K = 1 to s, j = 1 to m, i = 1 to m

y_{ki} = amount of output k produced by DMU I,

x_{ji} = amount of input j utilized by DMU i

v_k= weight given to output k,

u_j= weight given to input j

In order to solve the model, we need to convert it into a linear programming formulation.

$$\max \sum_{k=1}^{s} v_k y_{kp}$$

$$s.t \sum_{j=1}^{m} u_j x_{jp} = 1$$

$$\sum_{k=1}^{s} v_k y_{ki} - \sum_{j=1}^{m} u_j x_{ji} \le 0 \forall i$$

$$v_k, u_j \ge 0 \forall k, j \tag{6}$$

The dual problem can be specified as follows:

$$\min \theta$$

$$\sum_{i=1}^{n} \lambda_i x_{ji} - \theta x_{jp} \le 0 \forall j$$

$$\sum_{i=1}^{n} \lambda_i y_{ki} - y_{kp} \ge 0$$

$$\lambda_i \ge 0 \forall i \tag{7}$$

Where

θ efficiency score, and λs = dual variables

Analysis Of Variance and Tukey's Test

The analysis of variance is employed to compare three or more means for statistical significance. It involves simultaneous comparison of means using the F test. Fundamentally, variances are analyzed to make inferences about population means. Tukey's test is used in conjunction with an analysis of variance to find means that are significantly different from each other.

Results

Comparing DEA, SFA and COLS efficiencies

In order to calculate firm specific technical efficiency using alternative methods, we use the well known data sets provided with Tim Coelli's frontier 4.1. The data consist of the output of 60 firms (y) and variables labour and capital as the input variables of these firms. Subsequently, the technical efficiency of the 60 firms is computed using alternative methods namely, Data Envelope Analysis (DEA), Stochastic Frontier Analysis (SFA) and Corrected Ordinary Least Squares (COLS). For the purpose of brevity, I will denote DEA technical efficiency, SFA technical efficiency and COLS technical efficiency by teDEA, teSFA and teCOLS respectively in the rest of this paper.

The results in Table 1 indicate that the efficiency scores of the firms derived using the 3 methods, ranged between 20 to 100%. At lower levels of efficiency (<50%), SFA and COLS obtained 3 and 38 firms respectively whilst DEA recorded no firm. At moderate levels of efficiency (50 to 79%), SFA reported 33 firms, DEA reported 20 and COLS reported 21 firms. At higher levels of efficiency (>80%), SFA recorded 24 firms, DEA recorded 40 firms and COLS reported 1 firm.

Table 1. Frequencies and Cumulative Frequencies of Technical Efficiency Estimates of Firms obtained with DEA, SFA and COLS

Percent	teSFA Freq F	C. F	teDEA- Freq F	C. F	teCOLS Freq F	C. F
10–19	0	0	0	0	1	1
20–29	0	0	0	0	8	9
30–39	1	1	0	0	11	20
40–49	2	3	0	0	18	38
50–59	6	9	3	3	14	52
60–69	9	18	4	7	6	58
70–79	18	36	13	20	1	59
80–89	22	58	21	41	0	59
90–99	2	60	14	55	0	59
100	0	60	5	60	1	60

The average efficiencies of the three methods are presented in Table 2. The average efficiencies tend to differ among the three methods studied. The teDEA approach provided a higher mean efficiency of 83.37%, this is followed by teSFA and teCOLS approaches with 74.08% and 45.45% respectively. The coefficient of variation (CV) which is defined as the standard deviation expressed as a percentage of the mean is also investigated. The teCOLS method tends to have the largest CV of 31.72 %. This followed by teSFA and teDEA methods with CVs of 17.36% and 13.37% respectively as indicated in Table 2.

In order to investigate whether there is a significant difference in means between the efficiency scores from different methods, the analysis of variance (ANOVA) and Tukey's HSD (Honest Significance Difference) test were applied. The anova test (p-value=2e-16) suggest a significant difference among the scores from the three efficiency techniques as illustrated in Table 3. Using Tukey's HSD follow up test indicates that significant differences exist between teSfa and teDae, teCols and teDae, and teCols and teSfa as shown in Table 4.

Table 2. Average Efficiencies with Standard Deviation (S.D) and Coefficients of Variation (CV) according to the Different Estimation Methods

Model	Mean	S.d	CV (%)
teSFA	74.0833	12.8621	17.3617
teDEA	83.3667	11.1430	13.3663
teCOLS	45.45	14.4204	31.7279

Table 3. Analysis of Variance of Technical Efficiency Estimates of Firms obtained with DEA, SFA and COLS

	Df	Sum Sq	Mean Sq	F value	Pr (>F)
Method	2	46 874	23 437	141.3	<2e-16 ***
Residuals	177	29 355	166		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 4. Tukey's Honest Significance Difference test of Technical Efficiency Estimates of Firms obtained with DEA, SFA and COLS

method	diff	lwr	upr	p adj
teSfa – teDae	–9.283333	–14.84071	–3.725953	0.0003327
teCols – teDae	–37.916667	–43.47405	–32.359286	0.0000000
teCols – teSfa	–28.633333	–34.19071	–23.075953	0.0000000

Table 5. Correlation Analysis of the Efficiency Estimates between the Different Methods.

	teDEA	teSFA	teCOLS
teDEA	1.0000	0.8169	0.7892
teSFA	0.8169	1.0000	0.9191
teCOLS	0.7892	0.9191	1.0000

Table 5 provides the results of the correlation analysis between the actual values of the efficiency estimates from the three different methods. The correlation between the DEA and SFA efficiencies is 0.82, suggesting that the two kind of efficiencies are highly correlated, but they are not perfectly correlated. Similarly, Bogetoft and Otto (2011) in an empirical study noted that the correlation between DEA and SFA efficiencies as 0.78. They suggest that the two kinds of efficiencies are highly correlated.

The evidence of a strong correlation between teSFA and teDEA can also be seen in Figure 1 where there is a clear positive slope in the connection in the points. It is also obvious that the correlation between parametric methods tends to be stronger (0.92) than correlation between parametric and non parametric methods (0.82). Noticeably, it is also clear that there are several firms with DEA efficiency of 100% that have much lower SFA efficiency. There is even a firm with DEA efficiency of 100% and SFA efficiency of 57%.

Figure 2 shows three different boxplots representing efficiency scores derived from three estimation methods namely: DEA efficiency (teDEA), SFA efficiency (teSFA) and COLS efficiency (teCOLS). It is obvious that the mean and median differ between the three methods and the spread in the DEA efficiencies is much smaller than the spread in the SFA and COLS methods. It can also be noted that the median is lower for COLS efficiency and that, there are only a few firms with very high efficiency.

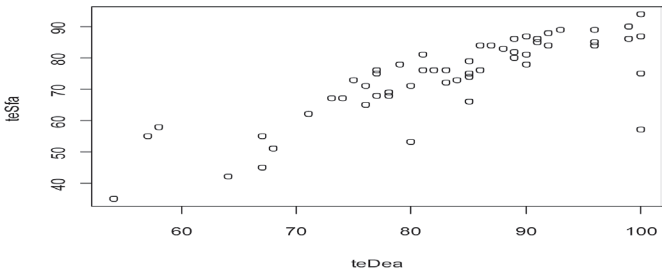

Figure 1. Scatterplot showing the Correlation between teSFA and teDEA

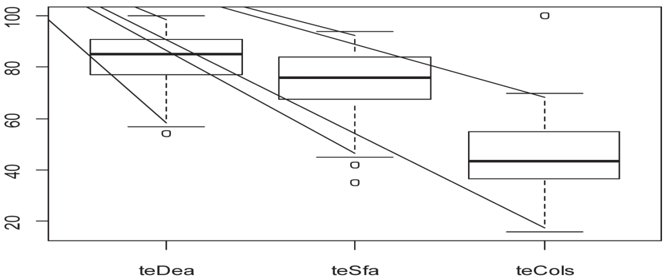

Figure 2. A Boxplot of the Different Estimation Methods

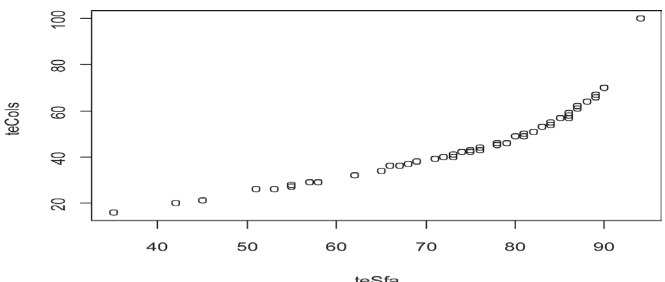

Figure 3. Scatterplot showing the Correlation between teCOLS and teSFA

The relationship between SFA and COLS efficiency is illustrated in Figure 3, where it is clear that for almost all firms the COLS efficiency is lower than the SFA efficiency except a few with very high COLS efficiency. This is expected as the COLS efficiency is constructed such that at least one firm has an efficiency of 1, which corresponds to the firm with the largest OLS error.

The histogram of the efficiency estimates of the 3 methods differ in the shape of their distribution. Noticeably, the histogram of teDEA shows more uniform distribution of efficiency estimates when compared to those of teSFA and teCOLS.

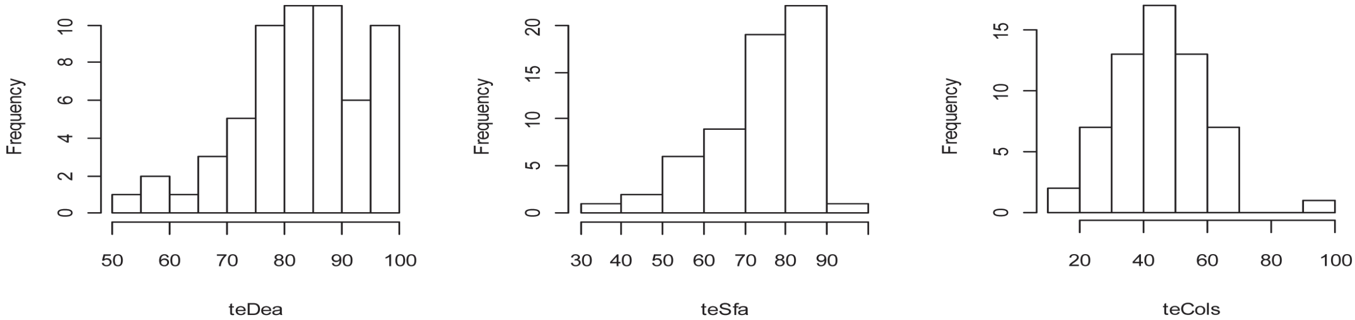

Figure 4. Histogram of Technical Efficiency Scores

 ISSN 1789-7874

Discussion

The results presented shows that the actual values of the efficiency estimates differ between the three competing methods. These differences in estimates may be attributed to the methodological differences in the different methods used. Mean technical efficiency is found to be sensitive to the choice of estimation method. Statistical test such as ANOVA and Tukey's HSD suggest significant difference in means between the efficiency scores from different methods. DEA and SFA frontiers resulted in higher mean technical efficiency estimates than the COLS production frontier. The different technical efficiency estimates provided by the different methods might have different policy implications since they imply different levels of firm capacity. These results suggest that the different methods lead to differences in conclusion. The findings of the current study are consistent with those of some studies. Jaforullah and Premachandra (2003) found that mean technical efficiency of the dairy industry is found to be sensitive to the choice of estimation method. They noted that generally, the SFA and DEA frontiers resulted in higher mean technical efficiency estimates than the COLS production frontier. Similarly, Bravo-Ureta and Rieger (1990) found that estimates of technical efficiencies vary across frontier estimation methods. Fundamentally, using one of the three techniques could lead researchers to entirely different conclusions due to the significant differences in efficiency scores between the methods.

The results from the research further revealed that efficiency estimates derived from the different methods tends to differ in the extent to which they vary. The results suggest that efficiency estimates from teCOLS is more variable when compared with efficiency estimates of teSFA and teDEA methods. Notably, teDEA efficiency estimates has the smallest variability among the three methods.

Though the actual values of the estimates differ between the methods but the estimates based on the three methods are highly correlated. This is consistent with Neff, Garcia and Nelson (1993) who found the correlation between the parametric measures to be higher than the correlation between parametric and non-parametric models. The presence of a strong positive correlation between the different efficiency estimates in this research suggests that the methods can be used concurrently to provide a holistic perspective of firm specific efficiency analysis.

It can be noted that the differences in methodology between the DEA, SFA and COLS accounts for the different results that were presented. These results are consistent with Bogetoft and Otto (2011) who found that several firms with a DEA efficiency of 1 that has much lower SFA efficiency. For example using a scatter plot they noted that there was a firm with a DEA efficiency of 1 and SFA efficiency of 0.6. Furthermore, Bogetoft and Otto (2011) using box plots found that the median is lower for COLS efficiency when compared with DEA and SFA efficiency. They found that for most firms they studied, the COLS efficiency was lower than the SFA efficiency. The significant differences in efficiency scores observed from different methods in the current study certainly have implications for the conclusions which can be derived for policy. It remains imperative that researchers employ an integrated approach that takes into consideration competing methods whilst modelling efficiency of decision making units.

Conclusion

Parametric and non parametric approaches of computing technical efficiency of decision making units have been developed. This study investigated the effect of the different methods on efficiency scores, by estimating technical efficiency from parametric and non parametric methods. The results indicates that though the actual values of the efficiency estimates differ between the alternative approaches of estimating technical efficiency, there exists a strong positive correlation between the efficiency estimates based on the three methods. Mean technical efficiency is found to be sensitive to the choice of estimation technique. Statistical test suggest significance difference in means between efficiency scores from different methods. On the basis of these results, this study argues that differences in conclusions are possible when the alternative methods of measuring technical efficiency such as the DEA, SFA and COLS are applied. Importantly, the methodologies in the DEA and SFA are very different and that is an important reason for the different results. Moreover, the differences in technical efficiency estimates provided by the alternative models might have different policy implications since they imply different levels of firm capacity. The presence of a strong positive correlation between the different efficiency estimates, suggest that the methods can be used concurrently to provide a holistic view of firm specific efficiency analysis. In effect, in estimating mean technical efficiency of an industry, it is advisable that one applies different methods of efficiency estimation as opposed to a single approach since the measurement of technical efficiency is sensitive to the choice of estimation method. Thus applying the approaches concurrently will produce better information on the technical efficiency of the industry by producing a range within which the true technical efficiency may lie.

Acknowledgements

The author is grateful to Dr. Edward Ebo Onumah and two anonymous reviewers for valuable comments.

Reference

Aigner, D.J., Lovell, C. A. K. and **Schmidt, P.** (1977): Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6: 21–27.

Bogetoft, P. and **Otto, L** (2011): Benchmarking with DEA, SFA and R. New York: Springer.

Bravo-Ureta B.E and **Pinheiro A.E.** (1993): Efficiency Analysis of Developing Country Agriculture: A Review of Frontier Function, Agricultural and Resource Economics Review 22: 88–101.

Bravo-Ureta, B. E. and **Rieger, L.** (1990): Alternative production frontier methodologies and dairy farm efficiency. Journal of Agricultural Economics 41: 215–226.

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

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

Coelli T.J., Rao D.S.P. and **Battese G.E.** (1998): An introduction to Efficiency and Productivity Analysis. Boston: Kluwer Academic Publishers.

Cooper, W.W., Seiford, M.L. and **Tone, K.** (2000): Data Envelopment Analysis: A Comprehensive Text With Models, Applications, References and DEA-Solver Software. Boston: Kluwer Academic Publishers.

Farrell, M. J. (1957): The measurement of productive efficiency. Journal of the Royal Statistical Society A120: 253–290.

Fried H.O., Knox Lovell C.A. and **Schmidt S.S.** (eds.) (1993): The Measurement of Productive Efficiency: Techniques and Applications. New York: Oxford University Press.

Greene, W. H. (2008): Econometric Analysis. (six edn) Pearson Prentice Hall

Greene, W. H. (1980): Maximum likelihood estimation of econometric frontier functions. Journal of Econometrics 13: 27–56.

Jaforullah, M. and **Premachandra, E.** (2003): Sensitivity of Technical Efficiency to estimation approaches: An Investigation using New Zealand Dairy Industry Data. University of Otago Economic Discussion Papers No. 0306.

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

Neff, D. L., Garcia, P. and **Nelson, C. H.** (1993): Technical efficiency: A comparison of production frontier methods. Journal of Agricultural Economics 44: 479489.

Kumbhakar, S.C. and **Lovell, C.** (2000): Stochastic Frontier Analysis. Cambridge: Cambridge University Press.

Seiford, M.L. and **Thrall, R.M.** (1990): Recent Developments in DEA: The Mathematical Programming Approach to Frontier Analysis, Journal of Econometrics 46: 7–38.