CANADA EXCELLENCE RESEARCH CHAIR

DATA SCIENCE FOR REAL-TIME DECISION-MAKING

STOCHASTIC ORDERING IN DATA ENVELOPMENT ANALYSIS

Mostafa Davtalab-Olyaie
Masoud Asgharian
Vahid Partovi Nia

Février 2016

DS4DM-2017-002

POLYTECHNIQUE MONTRÉAL
DÉPARTEMENT DE MATHÉMATIQUES ET GÉNIE INDUSTRIEL
Pavillon André-Aisenstadt
Succursale Centre-Ville C.P. 6079
Montréal - Québec
H3C 3A7 - Canada
Téléphone: 514-340-5121 # 3314

Stochastic Ordering in Data Envelopment Analysis

Mostafa Davtalab-Olyaie*, Masoud Asgharian†, and Vahid Partovi Nia‡

Abstract

Since its inception, stochastic Data Envelopment Analysis (SDEA) has found applications in vast and diverse areas of management science. Performance evaluation of DMUs is at the heart of DEA, both stochastic and deterministic. We present ranking methods for performance evaluation in SDEA using the notion of stochastic ordering that takes random fluctuations of the efficiency score into account. We apply the empirical Bayes approach for estimating the DEA efficiency score distribution and use stochastic ordering to rank the DEA efficiency score distributions. The stochastic ordering provides a notion for stochastic dominance using which one can define admissibility as a minimal performance requirement. We demonstrate how the proposed ranking method can be implemented and illustrate the method using a real data set.

Keywords: Data Envelopment Analysis; Stochastic Ordering; Admissibility; Efficiency Score Distribution; Empirical Bayes Approach.

1 Introduction

Efficiency evaluation of units is often a question of prime interest in many areas of application ranging from banking, business and economy to health care. Efficiency

*Correspondence: Mostafa Davtalab-Olyaie, Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Kashan, Kashan, Iran, m.davtalab-olyaie@kashanu.ac.ir.

†Dept of Math & Stat, McGill Univ, Montreal, Canada masoud@math.mcgill.ca.

‡Dept of Math & Ind Eng, École Poly de Montreal, Canada, vahid.partovinia@polymtl.ca.

analysis concerns with the performance of each unit in transforming their inputs into quantities of outputs. The relative comparison in efficiency analysis is examined against the efficient production frontier. In fact, the efficiency is measured based on the deviation of the position of a specific DMU from the efficient frontier.

In many real applications, the observed data are subject to uncertainty or might have been collected over several time periods such as monthly returns of hedge funds. Stochastic models where the inputs or outputs are considered to be random variables, seem to be the reasonable approach to account for such uncertainties or fluctuations when analyzing such data.

In their work, Aigner *et al.* (1977) and Meeusen and van den Broeck (1977) suggested a parametric approach known as stochastic frontier approach for efficiency analysis of DMUs. In this approach, a known functional form is postulated for the production function beforehand. By taking the statistical noise into account, the stochastic frontier approach allows the separation of deviation into inefficiency and noise terms. The reader can consulate Kumbhakar and Lovell (2000) for a review on theoretical and practical aspects of efficiency analysis using the stochastic frontier approach.

The nonparametric approach, known as the Data Envelopment Analysis (DEA), introduced by Charnes *et al.* (1978, 1979), and extended by Banker *et al.* (1984), offers a method widely used for estimating the efficiency of a set of multi-input multi-output DMUs. Following the criticisms of DEA raised by Schmidt (1985), and recently echoed further by Greene (1993), for the lack of solid statistical foundation, two different approaches were introduced in the literature to fill the gap:

–*Replacing production technology by a random production technology.*

A common approach to handle the uncertainty is via chance constrained models where the random Production Possibility Set (PPS) is replaced by an average PPS where the average is in the sense of Vorob'ev (1984). There has been a surge of articles on chance constrained models over the past two decades. Some of the early work on this subject were carried out by Land *et al.* (1993), Olesen and Petersen

(1995), and Cooper *et al.* (1998), among others. A recent review of this subject can be found in Cooper *et al.* (2011). The efficiency measured with respect to the average PPS is a fixed value. As discussed by Kao and Liu (2009) the inherent random fluctuation of the efficiency score, caused by the random nature of the input and output variables, cannot be captured using the chance constrained models.

–*Specifying a statistical model and a sampling process.*

The methodology developed using this approach can be divided into two categories: 1. Considering measurement error and noise, 2. Ignoring both measurement error and noise. The former assumes that all observations in the sampling process belong to PPS, and so there is no noise in the data generating process (DGP). Therefore the distance from the frontier just indicates the inefficiency term. In his work, Banker (1993) established the first building block of a solid statistical foundation for DEA by showing that the DEA estimators are essentially the maximum likelihood estimators under certain conditions. Gijbels *et al.* (1999) provide the asymptotic distribution of DEA estimator in the case of the single input and output. Kneip *et al.* (1998) generalize this result to the multiple inputs and outputs case. Simar and Wilson (1998) and Simar and Wilson (2000) suggest bootstrap techniques for evaluating the sampling variability of the efficiency estimator. Kneip *et al.* (2008) provide a full theory on the asymptotic properties of DEA estimator and a double-smooth bootstrap technique. Kneip *et al.* (2011) presents a simplified and consistent version of the double-smooth bootstrap method developed by Kneip *et al.* (2008).

In the latter category, it is assumed that there is noise in the DGP, and hence some observations may lie outside of PPS. In this case, the distance from the frontier has two components, noise and inefficiency. Hall and Simar (2002) show a fully nonparametric model with noise and inefficiency is not identifiable. They provide a method that allows for introduction of noise into the model. Simar (2007) extends these ideas to multivariate setting. A recent review of the subject can be found in Simar and Wilson (2015).

In the first category, Kao and Liu (2009) discuss how to obtain the DEA efficiency distributions of each DMU via a simulation technique, and used the mean of these distributions to rank the DMUs. Lamb and Tee (2012) have derived confidence intervals for the DEA efficiency distributions and developed a nonparametric bootstrap technique to rank DMUs. These ranking methods are all based on a summary of the DEA efficiency distributions.

In this manuscript we propose a partial ranking method using the notion of stochastic ordering that encompasses all the information of the efficiency distributions (with respect to the production frontier). This approach leads to the introduction of a minimal requirement that we call *admissibility*, see Definition 2. Using the notion of admissibility one can categorize DMUs into two categories, namely admissible and inadmissible DMUs.

Kneip *et al.* (1998) have shown that the DEA estimator converges to the efficiency score for each DMU at a rate that depends on the smoothness of the production frontier and on the number of inputs and outputs. One can therefore use the DEA efficiency to implement our ranking method. In other words, one can apply our stochastic ranking method on the DEA efficiency distributions of DMUs. For such implementation one needs to study structure of the distribution of the DEA efficiency estimator. As mentioned by Simar and Wilson (2000), Simar and Wilson (2007) and Kao and Liu (2009) the DEA efficiency estimator has a mixture structure with a point mass at 1, i.e., the density function can be returns as $p\delta_1 + (1-p)g$ where δ_1 is Dirc delta function at point 1, g is a continous density on $(0, 1)$ and $0 < p < 1$. We formalize this result in Theorem 2, showing that the DEA efficiency distribution does not have a continuous distribution even if both the random input and output variables are continuous. Using the point mass decomposition of the DEA efficiency distribution (Theorem 2), we then provide some conditions to check stochastic ranking and a sufficient condition for admissibility.

Implementation of our method requires estimation of the DEA efficiency distributions of DMUs. We simulate the input and output data for each DMU and measure the efficiency score of each DMU using the CCR model for each set of simulated data. This approach produces a sample from the DEA efficiency distribution and then, using standard statistical methods, we estimate the DEA efficiency distributions of DMUs. To this end, we use the Bayesian approach (the Markov chain Monte Carlo (MCMC) method (Robert and Casella, 2004)). The Bayesian approach has also attracted the attention of some authors (Tsionas and Papadakis, 2010). Tsionas and Papadakis (2010) have proposed a subjective Bayesian paradigm where a known prior distribution is imposed on the parameters of data distribution. We take a less subjective view in our Bayesian approach by using the empirical Bayes, choosing a data driven prior from a class of priors. This way we maintain both prior robustness and objectivity in our data analysis. As is well known in statistical literature, the empirical Bayes approach produces a statistically more efficient analysis (Carlin and Louis, 2008).

The rest of this manuscript is organized as follows. The notion of stochastic ordering and admissibility are presented in Section 2 where we show that stochastic ordering implies mean and median ordering, among several others. Implementation of stochastic ranking is discussed in Section 3. We start by exploring the structure of the DEA efficiency distribution, showing that it has a mixture structure with a point mass at 1. Using this mixture structure we establish some further results on stochastic ordering using the DEA efficiency distribution and a simple sufficient condition for admissibility in terms of the point-mass magnitude. The Empirical Bayes approach for estimating the DEA efficiency distribution is discussed in Section 3.2. We illustrate our methods using a set of real data in Section 4 where we use the Hasse diagram and graphical tools to visualize the result of our stochastic ranking. The last section, Section 5, includes some closing remarks. Proofs of the theorems are in Appendix I while the details of model calculation for our data analysis are collected in Appendix II.

2 Ranking Using Stochastic ordering

We first recall some basic concepts in efficiency analysis of DMUs. Consider a set of n DMUs, each using m inputs, $x \in R_+^m$, to produce s outputs, $y \in R_+^s$. The

Production Possibility Set (PPS), denoted by Ψ, is the set of all feasible activities,

$$\Psi = \{z = (x, y) \mid \text{the output } y \text{ can be produced with the input } x\}.$$

The frontier of Ψ, denoted by $\partial\Psi$, is called the *production function*. The set Ψ can be described by its x or y sections as follows,

$$X(y) = \{x \in R_+^m \mid (x, y) \in \Psi\} \qquad Y(x) = \{y \in R_+^s \mid (x, y) \in \Psi\}. \tag{1}$$

The Farrell efficiency boundaries are

$$\partial X(y) = \{x \mid x \in X(y), \theta x \notin X(y) \quad \forall 0 < \theta < 1\} \tag{2}$$

$$\partial Y(x) = \{y \mid y \in Y(x), \phi y \notin Y(x) \quad \forall \phi > 1\}, \tag{3}$$

using which Farrell input and output efficiency measures can be defined. For DMU$_j$, $j = 1, \ldots, n$,

$$\theta_j = \inf\{\theta \mid \theta x_j \in X(y_j)\} \tag{4}$$

$$\phi_j = \sup\{\phi \mid \phi y_j \in Y(x_j)\}. \tag{5}$$

When inputs or outputs of DMUs are random variables, the $\theta_j = \theta(x_j, y_j)$ will also be a random variable. To distinguish between random variables and their observed values, we use capital letters for random variables, while retaining small letters for the observed values.

Suppose $\Theta_j = \Theta(x_j, y_j)$ is the efficiency of DMU$_j$. Let $F_{\Theta_j}(\cdot)$ be the cumulative distribution function (cdf) of Θ_j. One may use different measures of central tendency, such as mean, median or quantiles of $F_{\Theta_j}(\cdot)$, for $j = 1, \ldots, n$ to rank DMUs. These ranking methods may be called, mean, median and quantile ranking. While the ranking methods are all based on a summary of $F_{\Theta_j}(\cdot)$, borrowing ideas from reliability theory and decision theory, one can consider the so-called *stochastic ordering* using the whole distribution of Θ_j, i.e., $F_{\Theta_j}(\cdot)$, which encompasses all the information about Θ_j.

Consider two variables Θ_j and $\Theta_{j'}$ as the efficiency variables of two stochastic DMUs, namely DMU_j and $\mathrm{DMU}_{j'}$. It is reasonable to prefer DMU_j over $\mathrm{DMU}_{j'}$ if $\Theta_j > \Theta_{j'}$ is more likely to happen than $\Theta_{j'} > \Theta_j$. Suppose that $S_{\Theta_j}(\cdot) = 1 - F_{\Theta_j}(\cdot)$ and $S_{\Theta_{j'}}(\cdot) = 1 - F_{\Theta_{j'}}(\cdot)$ are survival functions of Θ_j and $\Theta_{j'}$. We have

$$\begin{aligned}\int_{-\infty}^{\infty}\{S_{\Theta_j}(\theta) - S_{\Theta_{j'}}(\theta)\}\mathrm{d}F_{\Theta_j}(\theta) &= \int_{-\infty}^{\infty} S_{\Theta_j}(\theta) f_{\Theta_j}(\theta)\mathrm{d}\theta - \int_{-\infty}^{\infty} S_{\Theta_{j'}}(\theta) f_{\Theta_j}(\theta)\mathrm{d}\theta \\ &= 1/2 S^2_{\Theta_j}(\theta)\mid_{-\infty}^{+\infty} - \int_{-\infty}^{\infty} P(\Theta_{j'} > \Theta_j \mid \Theta_j = \theta) f_{\Theta_j}(\theta)\mathrm{d}\theta \\ &= 1/2 - P(\Theta_{j'} > \Theta_j),\end{aligned}$$

which implies

$$P(\Theta_j > \Theta_{j'}) = 1/2 + \int_{-\infty}^{\infty}\{S_{\Theta_j}(\theta) - S_{\Theta_{j'}}(\theta)\}\mathrm{d}F_{\Theta_j}(\theta). \tag{6}$$

For practical purposes one needs to estimate $P(\Theta_{j'} > \Theta_j)$. Using (6), we can readily estimate $p(\Theta_{j'} > \Theta_j)$ if we have an estimate of the survival functions of Θ_j and $\Theta_{j'}$.

Definition 1. *We say DMU_j is stochastically more efficient than $DMU_{j'}$ on Δ, denoted by $\Theta_j \succ_\Delta \Theta_{j'}$, if*

$$S_{\theta j}(\theta) \geq S_{\theta j'}(\theta), \text{ for all } \theta \in \Delta.$$

In particular, if $\Delta = [0, 1]$, we write $\Theta_j \succ \Theta_{j'}$, and say $DMU_{j'}$ is inadmissible.

We note that using equation (6), $P(\Theta_j > \Theta_{j'}) > 1/2$ if $\Theta_j \succ \Theta_{j'}$.

Figure 1 illustrates the notion of stochastic ordering. It depicts the probability density functions (pdf), $f(\theta)$, and the survival functions, $S(\theta)$, of the efficiency of two DMUs. We notice that while the pdfs are overlapping (left panel), the survival function of DMU_1, the solid curve, is always below the survival function of DMU_2, the dashed curve. The survival function of DMU_1 is dominated by the survival function of DMU_2. In other words, for any given efficiency level ξ, the efficiency of DMU_2 has a greater chance to be above ξ than the efficiency of DMU_1. That is, the

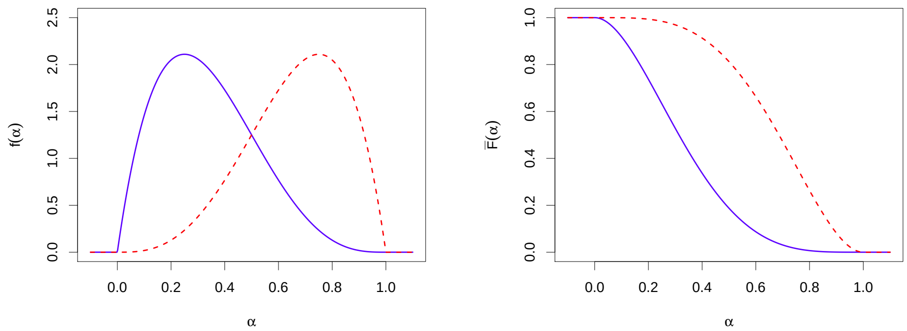

Figure 1: Stochastic ordering of efficiency distributions of DMU_1 (solid curve) versus DMU_2 (dashed curve). Comparison densities, left panel, and survival function, right panel (see Definition 1).

performance of DMU_2 is always superior to that of DMU_1 if superiority is measured by likeliness of being above an efficiency threshold.

Let the mean of the random variable Θ_j be denoted by $\mathbb{E}(\Theta_j)$ and its β-quantile by $\tilde{\mathbb{E}}_\beta(\Theta_j)$, where $0 < \beta < 1$. Each of these quantities can be used for a total (linear) ordering of DMUs. For instance, mean ranking can be performed based on $\mathbb{E}(\Theta_j)$ and β-quantile ranking based on $\tilde{\mathbb{E}}_\beta(\Theta_j)$. As a special case using $\tilde{\mathbb{E}}_{\beta=0.5}(\Theta_j)$, one can order DMUs based on the median of their efficiency distributions. The following result, whose proof follows from Definition 1, shows that ranking using stochastic ordering implies mean, median, quantile, and p-ranking. The converse is not necessarily true.

Theorem 1. *If $\Theta_j \succ \Theta_{j'}$ then $\mathbb{E}(\Theta_j) > \mathbb{E}(\Theta_{j'})$, and $\tilde{\mathbb{E}}_\beta(\Theta_j) > \tilde{\mathbb{E}}_\beta(\Theta_{j'})$ for any $0 < \beta < 1$.*

Remark 1. *A simple partial reverse connection between ranking based on quantiles and stochastic ordering immediately follows. If for all $0 < \beta < 1, \tilde{\mathbb{E}}_\beta(\Theta_j) > \tilde{\mathbb{E}}_\beta(\Theta'_j)$, then $\Theta_j \succ \Theta_{j'}$. This observation can be useful when a sample from both Θ_j and $\Theta_{j'}$ is available.*

The notion of inadmissibility was introduced in Definition 1. To further investigate and distinguish inadmissible DMUs from the admissible ones in Ψ, we need the following definition. Let $\Gamma = \{\Theta_Z \mid Z \in \Psi\}$ where Θ_Z is the efficiency variable of Z and $\mathcal{F} = \{S_\Theta \mid \Theta \in \Gamma\}$.

Definition 2. *An $S \in \mathcal{F}$ is called admissible with respect to $\mathcal{G} \subseteq \mathcal{F}$, if there is no $S_* \in \mathcal{G}$ such that $S_*(\theta) \geq S(\theta)$ for all $\theta \in [0,1]$, and the inequality is strict at least for one value of θ.*

3 Implementing ranking methods

3.1 DEA Efficiency Distribution

To implement the ideas developed in the previous section, one can measure efficiency using DEA.

Under the standard assumption of *inclusion of observations* and *return to scale*, n observations construct the unique non-empty PPS as follows:

$$\Psi^{\mathrm{DEA}} = \left\{ (x,y) \mid x_i \geq \sum_{j=1}^{n} \lambda_j x_{ij}, \forall i; y_r \leq \sum_{j=1}^{n} \lambda_j y_{rj}, \forall r; L \leq \sum_{j=1}^{n} \lambda_j \leq U; \lambda_j \geq 0, j = 1, \ldots, n \right\}, \tag{7}$$

where $L (0 \leq L \leq 1)$ and $U (U \geq 0)$ are lower and upper bounds for the sum of λ_j . Setting $L = 0$ and $U = \infty$, constant return to scale assumption, gives Ψ^{CCR} (Charnes *et al.*, 1978); while setting $L = U = 1$, variable return to scale assumption, gives Ψ^{BCC} (Banker *et al.*, 1984). The frontier of Ψ^{DEA}, $\partial\Psi^{\mathrm{DEA}}$, provides an estimate of $\partial\Psi$, the production function. Should we take Ψ^{CCR}, for instance, we can evaluate the relative efficiency by solving the CCR model

$$\begin{aligned} \tilde{\theta}_o = \min \quad & \theta \\ \text{s.t.} \quad & \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{io}, i = 1, \ldots, m \\ & \sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro}, r = 1, \ldots, s \\ & \lambda_j \geq 0, j = 1, \ldots, n. \end{aligned} \tag{8}$$

If $\tilde{\theta}_o = 1$, then DMU_o is CCR-efficient. We confine our attention to the CCR model to simplify our discussion in the sequel, though our approach is equally applicable to other conventional DEA models.

Note that the DEA efficiency $\tilde{\theta}$ is an estimate of the true efficiency θ. When the inputs or outputs are random variables, then $\tilde{\theta}$ is going to be a random variable itself too. It is clear that $\tilde{\theta}$ depends on n, the number of DMUs. We therefore use $\tilde{\Theta}_j^n$ to denote the random variable obtained using DEA which is an estimate of Θ_j. To have an idea about the relationship between Θ_j and $\tilde{\Theta}_j^n$, consider the case where there is only one output. In this case, $\tilde{\Theta}_j^n = \epsilon_j^n \Theta_j$ where $0 \leq \epsilon_j^n \leq 1$. Having assumed some regularity conditions, Kneip et al (1998) have shown that $\epsilon_j^n \to 1$ in probability as $n \to \infty$ at a rate that depends on the number of inputs and outputs, and the smoothness of the production function. For the ease of presentation we drop the superscript n and use $\tilde{\Theta}_j$ in the sequel whenever there is no danger of confusion.

The following result shows that there is at least one DMU whose $\tilde{\Theta}$ has a mixture structure with a point mass at 1 irrespective of the input and output variables being discrete or continuous. This point has also been implicitly mentioned by Simar and Wilson (2007) and Kao and Liu (2009). This structure of DEA efficiency distribution can be used to introduce a further simple ranking method using the point mass at 1. DMUs can be ranked according to the point mass of their DEA efficiencies at 1, the greater the point mass of a DMU at 1, the higher the ranking of the DMU. We call this ranking method *p-ranking*. Theorem 2 can also be used to establish a simple sufficient condition for admissibility. The proof of the theorem can be found in Appendix I.

Theorem 2. *Let $\tilde{\Theta}_j$ be the efficiency score of $Z_j = (X_j, Y_j)$, for $j = 1, \ldots, n$. Then there is at least one $\tilde{\Theta}_j$ with a positive mass at 1.*

Let $F_{\tilde{\Theta}_j}$ be the cumulative distribution function of $\tilde{\Theta}_j$, which is an estimate of F_{Θ_j}. Using Theorem 2 we have the following decomposition,

$$S_{\tilde{\Theta}_j}(\theta) = p_j + (1 - p_j) S^{<}_{\tilde{\Theta}_j}(\theta), \tag{9}$$

where $p_j = P(\tilde{\Theta}_j = 1)$, and $S_{\tilde{\Theta}_j}(\theta) = 1 - F_{\tilde{\Theta}_j}(\theta)$. Similarly we define $S^{<}_{\tilde{\Theta}_j}(\theta) = 1 - F^{<}_{\tilde{\Theta}_j}(\theta)$, where $F^{<}_{\tilde{\Theta}_j}$ is the cdf of the inefficiency component of the DEA efficiency score distribution. In the other words, $F^{<}_{\tilde{\Theta}_j}$ is the cdf of $\tilde{\Theta}_j$ when $\tilde{\Theta}_j < 1$.

The p-ranking method is perhaps the simplest method of ranking among the methods suggested above. Theorem 1 shows that ranking DMUs using stochastic ordering implies p-ranking. The following result establishes a partial reverse.

Theorem 3. *Let $\Upsilon = [\upsilon, 1]$, $\upsilon \in [0,1)$ and $p_j \geq p_{j'}$. Then $\tilde{\Theta}_j \succ_\Upsilon \tilde{\Theta}_{j'}$, if any of the following three conditions holds*

1. $\inf_{\theta \in \Upsilon} \left\{ S^{<}_{\tilde{\Theta}_j}(\theta) - S^{<}_{\tilde{\Theta}_{j'}}(\theta) \right\} > 0$,
2. $\inf_{\theta \in \{\theta | S^{<}_{\tilde{\Theta}_{j'}}(\theta) < 1\} \cap \Upsilon} \left\{ \frac{S^{<}_{\tilde{\Theta}_j}(\theta) - S^{<}_{\tilde{\Theta}_{j'}}(\theta)}{1 - S^{<}_{\tilde{\Theta}_{j'}}(\theta)} \right\} > \frac{p_{j'} - p_j}{1 - p_j}$,
3. $\inf_{\theta \in \{\theta | S^{<}_{\tilde{\Theta}_j}(\theta) < 1\} \cap \Upsilon} \left\{ \frac{S^{<}_{\tilde{\Theta}_j}(\theta) - S^{<}_{\tilde{\Theta}_{j'}}(\theta)}{1 - S^{<}_{\tilde{\Theta}_j}(\theta)} \right\} > \frac{p_{j'} - p_j}{1 - p_{j'}}$.

See Appendix I for the proof.

The following corollary follows immediately.

Corollary 1. *If $S^{<}_{\tilde{\Theta}_j}(\theta) = S^{<}_{\tilde{\Theta}_{j'}}(\theta)$ for all $\theta \in \Upsilon$, then $\tilde{\Theta}_j \succ_\Upsilon \tilde{\Theta}_{j'}$ if and only if $p_j > p_{j'}$.*

Direct verification of admissibility using Definition 2 is cumbersome. Using the mass point decomposition of the efficiency distribution, equation (9), we can present a simple sufficient condition for admissibility.

Theorem 4. *If $p_o > 3 - \sqrt{6}$, then DMU_o is admissible.*

See Appendix I for the proof.

3.2 Estimating DEA Efficiency Distribution

Implementing the above ranking methods requires estimation of the DEA efficiency score distribution. Suppose $Z_j \sim f_{Z_j}(. \mid \nu_j)$ [1] where the pdf f_{Z_j} is known up to

[1] The common underlying assumption of Kneip *et al.* (1998) and Banker (1993) is that Z_js are independent and identically distributed. While independence assumption can be retained for the purpose of estimating the efficiency distribution, should we assume common distribution for Z_js, distributions of $\tilde{\Theta}_j$ and Θ_j will be independent of j, i.e. all DMUs have the same distribution. We therefore assume that $Z_j \sim f_j$ are independent, but not identically distributed. Relaxing the assumption of common underlying distribution does not come without a price if we want to retain consistency of $\tilde{\Theta}$ for estimating Θ. All $f_j = f_{Z_j}$ should fulfill Assumption 4 of Kneip *et al.* (1998). One needs further assume that the number of DMUs, say n, as well as the size of the samples taken from each DMU, say T, grow to infinity to consistently estimate the production function (Lamb and Tee (2012)). In the setting of panel data, this amounts to assuming the data matrix grows in both dimensions.

finitely many unknown parameters $\nu_j, j = 1, \ldots, n$. To generate a sample from the DEA efficiency score distribution of DMU_j, one needs a sample from each DMU_j, i.e. a sample from Ψ^{DEA}. To this end we take an empirical Bayes approach.

Motivated by our example in Section 4, we consider a DEA analysis in situations where observations on DMU_j, for $j = 1, \ldots, n$ can be made at several discrete points in time, say $t = 1, \ldots, T$. Denote DMU_j at time t by $Z_{jt} = (X_{jt}, Y_{jt})'$ for $j = 1, ..., n$, where $X_{jt} = (X_{1jt}, \ldots, X_{mjt})'$ and $Y_{jt} = (Y_{1jt}, \ldots, Y_{sjt})'$ are respectively the input and output of DMU_j where we use X' for the transpose of X. We therefore have a PPS, Ψ_t^{DEA} at each time t. We further denote the whole data vector by $\mathbf{z}$ with entries z_{ijt}, where

$$\begin{aligned} \mathbf{z} = [z_{ijt}], i &= 1, \ldots, m+s, \\ j &= 1, \ldots, n, \\ t &= 1, \ldots, T, \end{aligned}$$

$\mathbf{z}_{jt} = [z_{ijt}]$, $i = 1, \ldots, m+s$, is the observed value of the random vector Z_{jt}, and the data vector of variable i of DMU_j over time t is denoted by $\mathbf{z}_{ij} = [z_{ijt}], t = 1, \ldots, T$.

3.2.1 Estimation Using Empirical Bayes

Bayesian data analysis involves the assignment of two distributions, the likelihood function being the multivariate distribution of observations given a parameter vector ϑ_o, say $f(\mathbf{z}_{o1}, \ldots, \mathbf{z}_{oT} \mid \vartheta_o)$ for DMU_o and the prior distribution of ϑ_o which itself is parameterized by hyper-parameter φ, say $f(\vartheta_o \mid \varphi)$. We consider a class of prior distributions such that $\int_{-\infty}^{\infty} f(\vartheta_o \mid \varphi) \mathrm{d}\vartheta_o = 1$. We take an empirical Bayes approach, and devise a numerical approximation using sampling from the posterior predictive distribution of data. Having estimated the hyper-parameters from the marginal likelihood (prior predictive distribution), a set of DMUs similar to the one observed is simulated and the efficiency is obtained on the simulated data to produce observations from the efficiency. These observations from the efficiency are used to estimate the efficiency distribution. Next we explain how this approach can be implemented.

Applying the empirical Bayes method, we estimate φ from data by maximizing the prior predictive

$$f(\mathbf{z} \mid \varphi) = \prod_{j=1}^{n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{z}_{j1}, \ldots, \mathbf{z}_{jt}, \ldots, \mathbf{z}_{jT} \mid \vartheta_j) f(\vartheta_j \mid \varphi) \mathrm{d}\vartheta_j. \tag{10}$$

The empirical Bayes estimate of φ is

$$\varphi_{\max} = \operatorname{argmax}_{\varphi} \log f(\mathbf{z} \mid \varphi). \tag{11}$$

We can generate a sample from the distribution of $\mathbf{z}_o$, say $\mathbf{z}_o^*$ by sampling from the posterior predictive distribution

$$f(\mathbf{z}_o^* \mid \mathbf{z}_{o1}, \ldots \mathbf{z}_{oT}, \varphi_{\max}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{z}_o^* \mid \vartheta_o) f(\vartheta_o \mid \mathbf{z}_{o1}, \ldots, \mathbf{z}_{oT}, \varphi_{\max}) \mathrm{d}\vartheta_o, \tag{12}$$

where

$$f(\vartheta_o \mid \mathbf{z}_{o1}, \ldots, \mathbf{z}_{oT}, \varphi_{\max}) = \frac{f(\mathbf{z}_{o1}, \ldots, \mathbf{z}_{oT} \mid \vartheta_o) f(\vartheta_o \mid \varphi_{\max})}{f(\mathbf{z}_{o1}, \ldots, \mathbf{z}_{oT} \mid \varphi_{\max})}$$

is the posterior distribution of ϑ_o.

If direct sampling from the posterior predictive distribution is complicated, one may use an indirect sampling through posterior samples. In other words, sample first from the posterior distribution $f(\vartheta_o \mid \mathbf{z}_{o1}, \ldots, \mathbf{z}_{oT}, \varphi_{\max})$, say $\vartheta_o^{\text{post}}$ and then generate a sample from $\mathbf{z}_o^*$ by sampling from $f(\mathbf{z}_o \mid \vartheta_o^{\text{post}})$.

To generate a sample from the efficiency of DMU$_o$, say θ_o^*, one needs to have a sample from the PPS, say Ψ^{CCR*}, which itself requires a predictive sample of all DMUs. Having produced Ψ^{CCR*}, a sample from the efficiency distribution of each DMU (including DMU$_o$) can be obtained by solving the CCR model on Ψ^{CCR*}. We repeat this procedure B times, for B large enough[2], to find B samples from the efficiency distribution of each DMU.

The estimate of p_o, say $\hat{p}_o$, and the non-parametric maximum likelihood estimate

[2]Similar to what is discussed in Kao and Liu (2009), one can investigate how many replications are proper to produce reliable results for his real data set. $B = 10000$ is typical.

of $S^{<}_{\tilde{\Theta}_o}(t)$ based on a sample of size B, $\theta^*_{o1},\ldots,\theta^*_{oB}$, for any $0<t<1$, is given by

$$\hat{p}_o = \frac{1}{B}\sum_{i=1}^{B}\varepsilon_{\{i|\theta_{oi}=1\}}(i), \hat{S}^{<}_{\tilde{\Theta}_{oB}}(t) = \frac{1}{B}\sum_{i=1}^{B}\varepsilon_{\{i|\theta_{oi}>t\}}(i),$$

where $\varepsilon_A(x)=1$ if $x\in A$, and equal to zero otherwise. [3]

4 Data example

We illustrate the methodologies developed in the previous sections using the airline data of Greene (2011)[4]. There are 6 airlines, each with three inputs (x_1 = total cost, x_2 = fuel price, and x_3 = load factor) and one output (y = revenue passenger miles). These information were collected on yearly basis for each airline over a period of 15 years. The left panel of Figure 2 depicts the trend of the output y (revenue passenger miles) over 15 years for each airline. The right panel of Figure 2 shows the output y versus x_3 (the load factor, the average capacity utilization of the fleet). Using Figure 2 one may speculate that Airline 1 and 2 have better performance than the other four airlines. This speculation is not, however, based on all the inputs and as such cannot be conclusive. We use the proposed ranking methods to analyse the performance of these six airlines in this section.

To simplify the computational aspects of our illustration, we first rescale the data so that each variable (input or output) has unit variance and we use the CCR model. Note that the CCR model is scale invariant. Furthermore, in the sequel we assume that the data are independent through time. Let

$$\mathbf{z}_{jt} \mid \mu_j, \mathbf{\Sigma}_j \overset{\text{iid}}{\sim} \mathcal{N}_{(m+s)}(\mu_j, \mathbf{\Sigma}_j),$$

[3] One can show that

$$||\hat{S}^{<}_{\tilde{\Theta}_{oB}} - S^{<}_{\tilde{\Theta}_o}||_\infty = \sup_{x\in[0,1]} |\hat{S}^{<}_{\tilde{\Theta}_{oB}}(x) - S^{<}_{\tilde{\Theta}_o}(x)| = O\left(\sqrt{\frac{\log\log(B)}{B}}\right), \quad \text{almost surely.}$$

When the input and output variables are continuous, the density of the DEA efficiency score can be estimated using the kernel, or other, density estimation method if visualization of the density is required.

[4] The data is available online through `http://people.stern.nyu.edu/wgreene/Text/tables/TableF7-1.txt`.

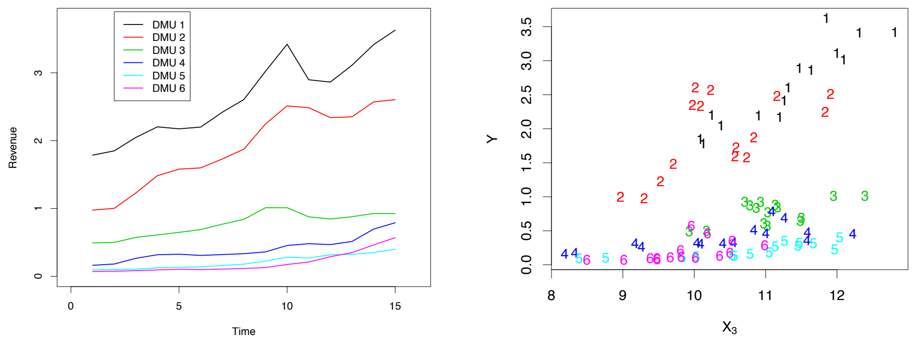

Figure 2: Depiction of the trend of output over 15 years for each DMUs of DMU_1, left panel, and output versus the third input, right panel.

and consider the conjugate prior family,

$$\begin{aligned} \mu_j \mid \tau, \kappa, \boldsymbol{\Sigma}_j &\overset{\text{iid}}{\sim} \mathcal{N}_{(m+s)}(\nu, \kappa\boldsymbol{\Sigma}_j), \\ \boldsymbol{\Sigma}_j &\overset{\text{iid}}{\sim} \mathcal{W}^{-1}(a, \boldsymbol{\Psi}), \end{aligned} \tag{13}$$

where $\mathcal{N}_k(\mu, \boldsymbol{\Sigma})$ denotes the k-variate normal distribution with mean μ and variance-covariance matrix $\boldsymbol{\Sigma}$, $\mathcal{W}^{-1}(a, \boldsymbol{\Psi})$ denotes the inverse Wishart distribution with a degrees of freedom and scaling matrix $\boldsymbol{\Psi}$. The scalar κ is an over-dispersion parameter. This model produces the following marginal distribution for a diagonal matrix $\boldsymbol{\Sigma}_j$, $\nu = \tau\mathbf{1}$, where $\mathbf{1}$ is a vector whose components are all equal 1, and $\boldsymbol{\Psi} = \frac{b}{2}\mathbf{I}$,

$$f(\mathbf{z} \mid \varphi) = \prod_{j=1}^{n} \prod_{i=1}^{m+s} \frac{b^{\frac{a}{2}} \Gamma\left(\frac{a+T}{2}\right)}{\pi^{\frac{T}{2}} |\mathbf{V}|^{\frac{1}{2}} \Gamma\left(\frac{a}{2}\right) \left\{b + (\mathbf{z}_{ij} - \tau\mathbf{1})'\mathbf{V}^{-1}(\mathbf{z}_{ij} - \tau\mathbf{1})\right\}^{\frac{a+T}{2}}}, \tag{14}$$

where each univariate random variable z_{ijt} is marginally Student-t distributed; see Appendix II for details.

As discussed above, we take an empirical Bayes approach, that is we estimate the hyper-parameters by maximizing the marginal distribution of observations given the hyper-parameters.The estimated hyper-parameters and their asymptotic standard errors are as follows: $\tau = 0.755(0.381)$, $\kappa = 34.510(131.869)$, $a = 1.056(0.074)$, $b = 0.100(0.002)$ (see Appendix II for details). We simulated $B = 10000$ data sets with each data point drawn from the predictive density and computed the efficiency for each data set. This gives 10000 efficiency values for each DMU. The inefficiency component in (9), $\tilde{S}^{<}(\theta)$, has a pdf $g(\theta)$.

The generated efficiency samples can then be used to estimate the parameters p_j and the density $g_j(.)$ for each $\mathrm{DMU}_j, j = 1, \ldots, 6$. The codes are implemented in the statistical software R (R Development Core Team, 2005) using the package `Benchmarking` (Bogetoft and Otto, 2010). Ranking DMUs with different methods and the estimation of p_j are reported in Table 1. The estimate of $g_j(.)$ for each $\mathrm{DMU}_j, j = 1, \ldots, 6$ is depicted in Figure 4. In the upper panel of Figure 4 the continuous parts are almost uniform. In the lower panel they are concentrated around 0.20. Therefore DMUs of the upper panel are efficient with probability p_o and their efficiency score is anywhere in $(0, 1)$ with probability $1 - p_o$, while in the lower panel a DMU is efficient with probability p_o and has efficiency score close to 0.2 with probability $1 - p_o$.

Our finding in Table 1 indicates that DMU_1 has the best performance according to all the ranking methods introduced in the previous sections. Ranking DMUs using stochastic ordering is feasible first by p-ranking and then by checking the conditions of Theorem 3. The results of the analysis is summarized in Table 1. DMU_1 and DMU_2 are stochastically unordered, but both are superior to DMU_3, DMU_4, DMU_5 and DMU_6. DMU_3 performs better than DMU_4, DMU_5 and DMU_6. DMU_4 is better than DMU_5, while DMU_4 with DMU_6, and DMU_5 with DMU_6 are stochastically unordered. We can depict this partial ranking of DMUs using a Hasse diagram. A Hasse diagram shows an arrow from k to j if $\tilde{\Theta}_k \succ \tilde{\Theta}_j$, and there is no i such that $\tilde{\Theta}_k \succ \tilde{\Theta}_i$ and $\tilde{\Theta}_i \succ \tilde{\Theta}_j$, see Rutherford (1965). Given that ranking using stochastic ordering provides the most comprehensive ranking method, the arrows in the Hasse diagram in Figure 3 show which dominance in DMUs cannot be changed through different specifications of ranking using different measures of central tendencies. It is, for instance, evident that using any measure of central tendency, DMU_1 is superior to DMU_3, but applying different central tendency statistics may reverse the ranking of DMU_1 with DMU_2.

The results reported in Table 1 indicate that DMU_3, DMU_4, DMU_5 and DMU_6 are inadmissible, while using Theorem 4, DMU_1 and DMU_2 are admissible in Ψ^{DEA}.

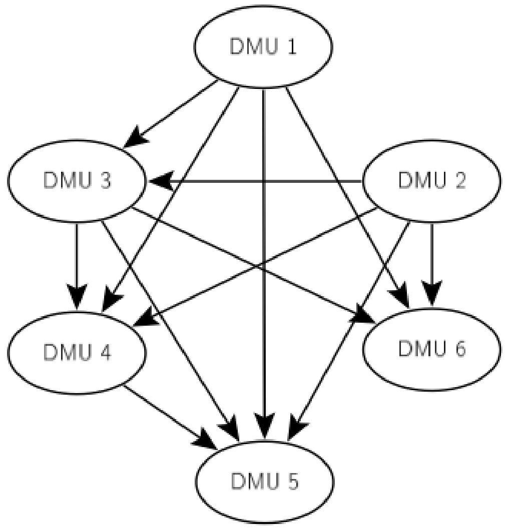

Figure 3: Hasse diagram of DMU domination, visualizing the result of ranking using stochastic ordering in Table 1.

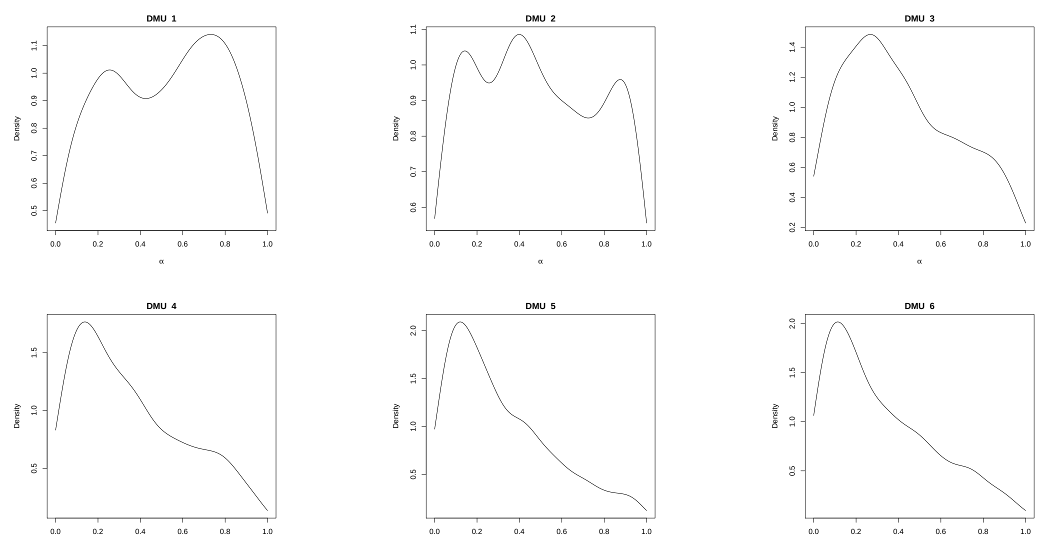

Figure 4: Estimation of $g_j(.)$, the probability density function of $S^{\leq}_{\tilde{\Theta}_j}$, for each DMU$_j, j = 1, \ldots, 6$ using kernel density estimation.

	DMU					
	1	2	3	4	5	6
mean	0.88	0.81	0.68	0.51	0.48	0.45
median	1.00	1.00	0.81	0.44	0.37	0.34
$\hat{p}_j$	0.74	0.63	0.44	0.25	0.23	0.20
stochastically ordered units	$\{3,4,5,6\}$	$\{3,4,5,6\}$	$\{4,5,6\}$	$\{5\}$	$\{\}$	$\{\}$

Table 1: Ranking DMUs using different distribution summaries, mean-ranking, median-ranking, p-ranking, ranking using stochastic ordering, and ranking using interactive ordering.

5 Closing Remarks

1. We took an empirical Bayes approach to estimate the DEA efficiency score distribution. One could also take a frequentist approach by estimating the parameters using maximum likelihood, or other, approach and then use a parametric bootstrap to generate data and produce samples from the DEA efficiency score distribution.

2. A more robust approach is possible when T is large. If information is gathered on each DMU over a long period, i.e. T is large, one can use a nonparametric approach to estimate the DEA efficiency score distribution for each DMU using the empirical cumulative distribution function.

3. We assumed that information collected on each DMU through time $(t = 1, 2, \ldots, T)$ are independent. This assumption can be relaxed. One can, for instance, replace this independence assumption by a Markov model. Such modelling, however, lead to more computational cost.

4. To facilitate calculation of our example in section 4 , we considered a conjugate prior family. Our approach can, however, be easily implemented for any prior using MCMC methods implemented in statistical packages such as R. One can therefore easily do a sensitivity analysis with respect to changes in the prior.

5. An R package that implements different ranking methods is under preparation, and will be published on R CRAN in the near future.

Appendix I

Proof of theorems

Proof of Theorem 2: We first note that $\tilde{\Theta}$ is a random variable defined on the probability space $(\Omega, \Im, P)$, where $\Im$ is a σ-algebra of the subsets of Ω and P is the probability measure on $\Im$. We further note that for any $\omega \in \Omega$ we have a Ψ^{CCR}. Let $A_i = \{\omega \in \Omega : \tilde{\Theta}_i(\omega) = 1\}$ for $i = 1, \ldots, n$. Since in any Ψ^{CCR} there is at least one efficient DMU, we have $\Omega = \bigcup\limits_{i=1}^{n} A_i$. Now suppose that there is no mass point at 1 for any DMU, i.e., $p_i = P(\omega : \tilde{\Theta}_i(\omega) = 1) = P(A_i) = 0$ for $i = 1, \ldots, n$. Then using Boole's inequality, $P(\bigcup\limits_{i=1}^{n} A_i) \leq \sum\limits_{i=1}^{n} P(A_i) = 0$. On the other hand, $P(\bigcup\limits_{i=1}^{n} A_i) = P(\Omega) = 1$. This is a contradiction. ■

Proof of Theorem 3: Suppose $p_j > p_{j'}$ and Condition 1 is satisfied; i.e., $S^{<}_{\tilde{\Theta}_j}(\theta) > S^{<}_{\tilde{\Theta}_{j'}}(\theta)$, $\forall \theta \in \Upsilon$. Then

$$(1 - p_{j'}) \left(1 - S^{<}_{\tilde{\Theta}_{j'}}(\theta)\right) - (1 - p_j) \left(1 - S^{<}_{\tilde{\Theta}_j}(\theta)\right) > 0,$$

and hence

$$p_j + (1 - p_j) S^{<}_{\tilde{\Theta}_j}(\theta) > p_{j'} + (1 - p_{j'}) S^{<}_{\tilde{\Theta}_{j'}}(\theta).$$

Therefore $S_{\tilde{\Theta}_j}(\theta) > S_{\tilde{\Theta}_{j'}}(\theta)$, for all $\theta \in \Upsilon$; which implies $\tilde{\Theta}_j \succ_{\Upsilon} \tilde{\Theta}_{j'}$.

Suppose Condition 2 holds, that is, $\forall \theta \in \{\theta \mid S^{<}_{\tilde{\Theta}_{j'}}(\theta) < 1\} \cap \Upsilon$, and

$$\inf_{\theta \in \Upsilon} \left\{ \frac{S^{<}_{\tilde{\Theta}_j}(\theta) - S^{<}_{\tilde{\Theta}_{j'}}(\theta)}{1 - S^{<}_{\tilde{\Theta}_{j'}}(\theta)} \right\} > \frac{p_{j'} - p_j}{1 - p_j}.$$

This is equivalent to

$$\frac{S^{<}_{\tilde{\Theta}_j}(\theta) - S^{<}_{\tilde{\Theta}_{j'}}(\theta)}{1 - S^{<}_{\tilde{\Theta}_{j'}}(\theta)} > \frac{p_{j'} - p_j}{1 - p_j}, \forall \theta \in \Upsilon,$$

hence

$$\left\{1 - S^{<}_{\tilde{\Theta}_{j'}}(\theta)\right\} - \left\{1 - S^{<}_{\tilde{\Theta}_j}(\theta)\right\} - \frac{\left\{1 - S^{<}_{\tilde{\Theta}_{j'}}(\theta)\right\}}{1 - p_j} \{(1 - p_j) - (1 - p_{j'})\} > 0,$$

which, in turn, implies

$$(1 - p_{j'}) \left\{ 1 - S^{<}_{\tilde{\Theta}_{j'}}(\theta) \right\} - (1 - p_j) \left\{ 1 - S^{<}_{\tilde{\Theta}_j}(\theta) \right\} > 0.$$

Thus $p_j + (1 - p_j) S^{<}_{\tilde{\Theta}_j}(\theta) > p_{j'} + (1 - p_{j'}) S^{<}_{\tilde{\Theta}_{j'}}(\theta)$, and therefore $S_{\tilde{\Theta}_j}(\theta) > S_{\tilde{\Theta}_{j'}}(\theta), \forall \theta \in \Upsilon$; yielding $\tilde{\Theta}_j \succ_{\Upsilon} \tilde{\Theta}_{j'}$.

The proof for Condition 3 is similar. ■

To prove Theorem 4, we need to establish the following lemma first.

Lemma 1. *If DMU$_o$ is inadmissible, then there exists* $\tilde{\lambda} = \left(\tilde{\lambda}_1, ..., \tilde{\lambda}_n\right) \geq 0$ *such that* $P(\tilde{\Theta}_{\tilde{\lambda}} \geq \tilde{\Theta}_o) \geq 2p_o - \frac{p_o^2+1}{2}$, *where* $\tilde{\Theta}_{\tilde{\lambda}}$ *indicates the efficiency of the virtual stochastic DMU* $(\sum_{j=1}^n \tilde{\lambda}_j X_j, \sum_{j=1}^n \tilde{\lambda}_j Y_j)$.

Proof. Since DMU$_o$ is inadmissible, then there exists $\tilde{\lambda} = \left(\tilde{\lambda}_1, ..., \tilde{\lambda}_n\right) \geq 0$ such that

$$S_{\tilde{\Theta}_{\tilde{\lambda}}}(\theta) \geq S_{\tilde{\Theta}_o}(\theta), \text{ for all } \theta \in [0, 1], \tag{15}$$

where $S_{\tilde{\Theta}_{\tilde{\lambda}}}(\cdot)$ is the survival function of the efficiency of the virtual stochastic DMU that use the input $\sum_{j=1}^n \tilde{\lambda}_j X_j$ to produce the output $\sum_{j=1}^n \tilde{\lambda}_j Y_j$. For any $\lambda \geq 0$ define $\Omega_\lambda = \left\{\omega \in \Omega \mid \tilde{\Theta}_\lambda(\omega) > \tilde{\Theta}_o(\omega)\right\}$. We have

$$\begin{aligned} P(\Omega_{\tilde{\lambda}}) &= P(\tilde{\Theta}_{\tilde{\lambda}} \geq \tilde{\Theta}_o) \\ &= P(\tilde{\Theta}_{\tilde{\lambda}} \geq \tilde{\Theta}_o \mid \tilde{\Theta}_o = 1) P(\tilde{\Theta}_o = 1) + P(\tilde{\Theta}_{\tilde{\lambda}} \geq \tilde{\Theta}_o \mid \tilde{\Theta}_o < 1) P(\tilde{\Theta}_o < 1) \\ &= P(\tilde{\Theta}_{\tilde{\lambda}} \geq \tilde{\Theta}_o \mid \tilde{\Theta}_o = 1) p_o + P(\tilde{\Theta}_{\tilde{\lambda}} \geq \tilde{\Theta}_o \mid \tilde{\Theta}_o < 1)(1 - p_o), \end{aligned}$$

where

$$\begin{aligned} P(\tilde{\Theta}_{\tilde{\lambda}} \geq \tilde{\Theta}_o \mid \tilde{\Theta}_o < 1) &= \int_0^1 P(\tilde{\Theta}_{\tilde{\lambda}} \geq \tilde{\Theta}_o \mid \tilde{\Theta}_o = \theta, \tilde{\Theta}_o < 1) \mathrm{d}F(\theta \mid \tilde{\Theta}_o < 1) \\ &= \int_0^1 P(\tilde{\Theta}_{\tilde{\lambda}} \geq \tilde{\Theta}_o \mid \tilde{\Theta}_o = \theta, \tilde{\Theta}_o < 1) \frac{(1 - p_o)}{P(\tilde{\Theta}_o < 1)} \mathrm{d}F^{<}_{\tilde{\Theta}_o}(\theta) \\ &= \int_0^1 P(\tilde{\Theta}_{\tilde{\lambda}} \geq \tilde{\Theta}_o \mid \tilde{\Theta}_o = \theta, \tilde{\Theta}_o < 1) \mathrm{d}F^{<}_{\tilde{\Theta}_o}(\theta). \end{aligned}$$

We note that DMU$_o$ is not on the frontier for any $\omega \in \bigcup_{\lambda} \Omega_\lambda$. Thus given $\omega \in \bigcup_{\lambda} \Omega_\lambda$, the efficiency of DMU$_o$ cannot affect the efficiency of other DMUs. Thus $\tilde{\Theta}_{\tilde{\lambda}}$ is independent of $\tilde{\Theta}_o$ given $\omega \in \bigcup_\lambda \Omega_\lambda$. We therefore have

$P(\tilde{\Theta}_{\tilde{\lambda}} \geq \tilde{\Theta}_o \mid \tilde{\Theta}_o < 1) = \int_0^1 P(\tilde{\Theta}_{\tilde{\lambda}} \geq \theta) \mathrm{d}F^{<}_{\tilde{\Theta}_o}(\theta) = \int_0^1 S_{\tilde{\Theta}_{\tilde{\lambda}}}(\theta) \mathrm{d}F^{<}_{\tilde{\Theta}_o}(\theta).$

Using (15),

$P(\tilde{\Theta}_{\tilde{\lambda}} \geq \tilde{\Theta}_o \mid \tilde{\Theta}_o < 1) \geq \int_0^1 S_{\tilde{\Theta}_o}(\theta) \mathrm{d}F^{<}_{\tilde{\Theta}_o}(\theta).$

On the other hand, we know

$S_{\tilde{\Theta}_o}(\theta) = p_o + (1-p_o) S^{<}_{\tilde{\Theta}_0}(\theta), \forall \theta \in [0,1]$; and hence

$\int_0^1 S_{\tilde{\Theta}_o}(\theta) \mathrm{d}F^{<}_{\tilde{\Theta}_o}(\theta) = p_o + (1-p_o) \int_0^1 S^{<}_{\tilde{\Theta}_o}(\theta) \mathrm{d}F^{<}_{\tilde{\Theta}_o}(\theta) = p_o + \frac{(1-p_o)}{2} = \frac{1+p_o}{2}$. Then

$$
\begin{aligned}
P(\Omega_{\tilde{\lambda}}) &\geq P(\tilde{\Theta}_{\tilde{\lambda}} \geq \tilde{\Theta}_o \mid \tilde{\Theta}_o = 1) P(\tilde{\Theta}_o = 1) + \frac{(1-p_o^2)}{2} \\
&\geq P(\tilde{\Theta}_{\tilde{\lambda}} = 1, \tilde{\Theta}_o = 1) + \frac{(1-p_o^2)}{2} \\
&= p_{\tilde{\lambda}} + p_o - P(\tilde{\Theta}_{\tilde{\lambda}} = 1 \text{ or } \tilde{\Theta}_o = 1) + \frac{(1-p_o^2)}{2} \\
&\geq 2p_o - P(\tilde{\Theta}_{\tilde{\lambda}} = 1 \text{ or } \tilde{\Theta}_o = 1) + \frac{(1-p_o^2)}{2} \\
&\geq 2p_o - \frac{p_o^2+1}{2}. \blacksquare
\end{aligned}
$$

Proof of Theorem 4: Suppose DMU$_o$ is inadmissible, then using Lemma 1, there exists $\tilde{\lambda} = \left(\tilde{\lambda}_1, ..., \tilde{\lambda}_n\right)$ such that $P(\Omega_{\tilde{\lambda}}) = P(\tilde{\Theta}_{\tilde{\lambda}} \geq \tilde{\Theta}_o) \geq 2p_o - \frac{p_o^2+1}{2}$.

On the other hand, $\{\omega \in \Omega \mid \tilde{\Theta}_o(\omega) = 1\} = \Omega - \bigcup_\lambda \Omega_\lambda$. Thus

$$
\begin{aligned}
p_o = P(\tilde{\Theta}_o = 1) &= 1 - P(\bigcup_{\lambda} \Omega_\lambda) \\
&\leq 1 - P(\Omega_{\tilde{\lambda}}) \\
&\leq 1 - \left(2p_o - \frac{p_o^2+1}{2}\right) = -2p_o + \frac{p_o^2+3}{2}.
\end{aligned}
$$

Hence, if DMU$_o$ is inadmissible, then $-3p_o + \frac{p_o^2+3}{2} \geq 0$. This inequality is fulfilled if

$p_o \in [0, 3-\sqrt{6}]$. This is a contradiction. ■

Appendix II

Model calculations

posterior predictive

Considering a diagonal variance-covariance matrix $\mathbf{\Sigma}_j$ and $\mathbf{\Psi} = \frac{b}{2}\mathbf{I}$, the hierarchical model (13) simplifies to

$$\begin{aligned}
z_{ijt} \mid \mu_{ij}, \sigma^2_{ij} &\overset{\text{iid}}{\sim} \mathcal{N}(\mu_{ij}, \sigma^2_{ij}), \\
\mu_{ij} \mid \tau, \kappa &\overset{\text{iid}}{\sim} \mathcal{N}(\tau, \kappa\sigma^2_{ij}), \\
\sigma^2_{ij} &\overset{\text{iid}}{\sim} \Gamma^{-1}(\frac{a}{2}, \frac{b}{2}),
\end{aligned} \tag{16}$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes the univariate normal distribution with mean μ and variance σ^2, and $\Gamma^{-1}(a, b)$ denotes the inverse gamma distribution with the shape and scale parameters a and b.

Given the independence between the DMUs and within the components of each DMU, we have

$$f(\mathbf{z} \mid \varphi) = \prod_{j=1}^{n} \prod_{i=1}^{m+s} f(\mathbf{z}_{ij} \mid \varphi),$$

where each $\mathbf{z}_{ij}$ is a vector of length T and

$$f(\mathbf{z}_{ij} \mid \varphi) = \int_0^\infty f(\mathbf{z}_{ij} \mid \sigma^2_{ij}) f(\sigma^2_{ij}) \mathrm{d}\sigma^2_{ij}.$$

We first calculate

$$\begin{aligned} f(\mathbf{z}_{ij} \mid \sigma_{ij}^2) &= \int_{-\infty}^{\infty} f(\mathbf{z}_{ij} \mid \mu_{ij}, \sigma_{ij}^2) \mathrm{d}\mu_{ij} \\ &= \int_{-\infty}^{\infty} \prod_{t=1}^{T} f(z_{ijt} \mid \mu_{ij}, \sigma_{ij}^2) f(\mu_{ij} \mid \sigma_{ij}^2) \mathrm{d}\mu_{ij} \\ &= (2\pi\sigma_{ij}^2)^{-\frac{T}{2}} (2\pi\sigma_{ij}^2)^{-\frac{1}{2}} \\ &\quad \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2\sigma_{ij}^2} \sum_{t=1}^{T} (z_{ijt} - \mu_{ij})^2 - \frac{1}{2\kappa\sigma_{ij}^2} (\mu_{ij} - \tau)^2 \right\} \mathrm{d}\mu_{ij}. \end{aligned}$$

After some simple algebra

$$f(\mathbf{z}_{ij} \mid \sigma_{ij}^2) = (2\pi)^{-\frac{T}{2}} |\sigma_{ij}^2 \mathbf{V}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{z}_{ij} - \tau\mathbf{1})' (\sigma_{ij}^2 \mathbf{V})^{-1} (\mathbf{z}_{ij} - \tau\mathbf{1}) \right\}, \qquad (17)$$

where $\mathbf{1}$ is a vector of length T whose components are all equal 1, $\mathbf{V}$ is a $T \times T$ symmetric matrix with diagonal elements $1 + \kappa$ and equal off-diagonals κ, and $|\mathbf{V}|$ denotes the determinant of $\mathbf{V}$. Next, we integrate (17) with respect to the inverse gamma density $f(\sigma_{ij}^2)$,

$$\begin{aligned} f(\mathbf{z}_{ij} \mid \varphi) &= (2\pi)^{-\frac{T}{2}} |\mathbf{V}|^{-\frac{1}{2}} b^{\frac{a}{2}} 2^{-\frac{a}{2}} \Gamma(\frac{a}{2})^{-1} \times \\ &\quad \int_{0}^{\infty} (\sigma_{ij}^2)^{1-\frac{T}{2}-\frac{a}{2}} \exp\left\{ -\frac{(\mathbf{z}_{ij} - \tau\mathbf{1})' \mathbf{V}^{-1} (\mathbf{z}_{ij} - \tau\mathbf{1})}{2\sigma_{ij}^2} - \frac{b}{2\sigma_{ij}^2} \right\} \mathrm{d}\sigma_{ij}^2. \end{aligned}$$

After changing the variable of integration using the substitution $\gamma = \sigma_{ij}^{-2}$

$$\begin{aligned} f(\mathbf{z}_{ij} \mid \varphi) &= (2\pi)^{-\frac{T}{2}} |\mathbf{V}|^{-\frac{1}{2}} b^{\frac{a}{2}} 2^{-\frac{a}{2}} \Gamma(\frac{a}{2})^{-1} \times \\ &\quad \int_{0}^{\infty} \gamma^{\frac{a+T-2}{2}} \exp\left\{ -\frac{b + (\mathbf{z}_{ij} - \tau\mathbf{1})' \mathbf{V}^{-1} (\mathbf{z}_{ij} - \tau\mathbf{1})}{2} \gamma \right\} \mathrm{d}\gamma. \end{aligned}$$

The last integral is the gamma integral and therefore,

$$f(\mathbf{z}_{ij} \mid \varphi) = \frac{b^{\frac{a}{2}}\Gamma\left(\frac{a+T}{2}\right)}{\pi^{\frac{T}{2}}|\mathbf{V}|^{\frac{1}{2}}\Gamma\left(\frac{a}{2}\right)\left\{b + (\mathbf{z}_{ij} - \tau\mathbf{1})'\mathbf{V}^{-1}(\mathbf{z}_{ij} - \tau\mathbf{1})\right\}^{\frac{a+T}{2}}}. \tag{18}$$

Posterior density

Given the independence assumption, the full posterior is the product of individual posteriors

$$f(\mu_{ij}, \sigma_{ij}^2 \mid \mathbf{z}_{ij}) \propto f(\mathbf{z}_{ij} \mid \mu_{ij}, \sigma_{ij}^2) f(\mu_{ij}, \sigma_{ij}^2),$$

where $f(\mathbf{z}_{ij} \mid \mu_{ij}, \sigma_{ij})$ is a normal distribution and $f(\mu_{ij}, \sigma_{ij}^2) = f(\mu_{ij} \mid \sigma_{ij}^2) f(\sigma_{ij}^2)$ are normal-inverse-gamma distributions where the normal-inverse-gamma with parameters τ, κ, a, b is

$$\begin{aligned} f(\mu, \sigma^2) &= \frac{b^a (\sigma^2)^{-a-\frac{3}{2}}}{\Gamma(a)(2\pi\kappa)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2\kappa\sigma^2}(\mu - \tau)^2 - \frac{b}{\sigma^2}\right\} \\ & \mu, \tau \in \mathcal{R}, \quad \sigma^2, a, b, \kappa > 0. \end{aligned}$$

Therefore, the normal-inverse-gamma has the kernel

$$f(\mu, \sigma^2) \propto (\sigma^2)^{-(a+\frac{3}{2})} \exp\left\{-\frac{1}{2\kappa\sigma^2}(\mu - \tau)^2 - \frac{b}{\sigma^2}\right\}. \tag{19}$$

As normal and normal-inverse-gamma are conjugate forms, the posterior is also normal-inverse-gamma. More precisely,

$$\begin{aligned} f(\mu_{ij}, \sigma_{ij}^2 \mid \mathbf{z}_{ij}) &\propto \left\{\prod_{t=1}^{T} f(z_{ijt} \mid \mu_{ij}, \sigma_{ij}^2)\right\} f(\mu_{ij}, \sigma_{ij}^2) \\ &\propto \frac{1}{(2\pi\sigma_{ij}^2)^{\frac{T}{2}}} \exp\left\{-\frac{1}{2\sigma_{ij}^2}\sum_{t=1}^{T}(z_{ijt} - \mu_{ij})^2\right\} \times \\ &\quad \frac{b^{\frac{a}{2}}(\sigma_{ij}^2)^{-\frac{a+3}{2}}}{2^{\frac{a}{2}}\Gamma(\frac{a}{2})(2\pi\kappa)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2\kappa\sigma_{ij}^2}(\mu_{ij} - \tau)^2 - \frac{b}{2\sigma_{ij}^2}\right\}. \end{aligned}$$

After some algebraic simplifications

$$f(\mu_{ij}, \sigma_{ij}^2 \mid \mathbf{z}_{ij}) \propto (\sigma_{ij}^2)^{-(\frac{a+T}{2}+\frac{3}{2})} \times \exp\left\{-\frac{1+\kappa T}{2\kappa\sigma_{ij}^2}\left(\mu - \frac{\tau+\kappa\sum_{t=1}^{T} z_{ijt}}{1+\kappa T}\right)^2 - \frac{1}{2\kappa\sigma_{ij}^2}(\tau^2+\kappa b+\kappa\sum_{t=1}^{T} z_{ijt}^2)\right\}. \tag{20}$$

Comparing (20) with (19) we see that the posterior is in the normal-inverse-gamma form with parameters

$$\tau^* = \frac{\tau+\kappa\sum_{t=1}^{T} z_{ijt}}{1+\kappa T}, \qquad \kappa^* = \frac{\kappa}{1+\kappa T},$$
$$a^* = \frac{a+T}{2}, \qquad b^* = \frac{1}{2\kappa}\left(\tau^2+\kappa b+\kappa\sum_{t=1}^{T} z_{ijt}^2\right).$$

References

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

Banker, R. (1993) Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. *Management Science* **39**, 1265–1273.

Banker, R., Charnes, A. and Cooper, W. (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* **30**, 1078–1092.

Bogetoft, P. and Otto, L. (2010) *Benchmarking with DEA, SFA, and R.* New York: Springer.

Carlin, B. P. and Louis, T. A. (2008) *Bayesian Methods for Data Analysis.* New York: CRC Press.

Charnes, A., Cooper, W. and Rhodes, E. (1978) Measuring the efficiency of decision making units. *European Journal of Operational Research* **2**(6), 429–444.

Charnes, A., Cooper, W. W. and Rhodes, E. (1979) Short communication: Measuring the efficiency of decision-making units. *European Journal of Operational Research* **3**(4), 339.

Cooper, W. W., Huang, Z., Lelas, V., Li, S. X. and Olesen, O. B. (1998) Chance constrained programming formulations for stochastic characterizations of efficiency and dominance in DEA. *Journal of Productivity Analysis* **9**, 53–79.

Cooper, W. W., Huang, Z. and Li, S. X. (2011) Chance-constrained DEA. In *Handbook on Data Envelopment Analysis*. Springer.

Gijbels, I., Mammen, E., Park, B. and Simar, L. (1999) On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association* **94**, 220–228.

Greene (1993) *The econometric approach to efficiency analysis, in H. O. Fried, C. A. K. Lovell, and S. S. Schmidt, eds., The Measurement of Productive Efficiency and Productivity growth*. Oxford: Oxford University Press.

Greene, W. H. (2011) *Econometric Analysis*. 7th edition. New York: Prentice Hall.

Hall, P. and Simar, L. (2002) Estimating a changepoint, boundary or frontier in the presence of observation error. *Journal of the American Statistical Association* **97**, 523–534.

Kao, C. and Liu, S.-T. (2009) Stochastic data envelopment analysis in measuring the efficiency of Taiwan commercial banks. *European Journal of Operational Research* **196**, 312–322.

Kneip, A., Park, B. and Simar, L. (1998) A note on the convergence of nonparametric dea estimators for production efficiency scores. *Econometric Theory* **14**, 783–793.

Kneip, A., Simar, L. and Wilson, P. W. (2008) Asymptotics and consistent bootstraps for dea estimators in non-parametric frontier models. *Econometric Theory* **24**, 1663–1697.

Kneip, A., Simar, L. and Wilson, P. W. (2011) A computationally efficient, consistent bootstrap for inference with non-parametric dea estimators. *Computational Economics* **38**, 483–515.

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

Lamb, J. D. and Tee, K.-H. (2012) Resampling dea estimates of investment fund performance. *European Journal of Operational Research* **223**, 834–841.

Land, K., Lovell, C. and Thore, S. (1993) Chance-constrained data envelopment analysis. *Managerial and Decision Economics* **14**(6), 541–554.

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

Olesen, O. B. and Petersen, N. C. (1995) Chance constrained efficiency evaluation. *Management Science* **41**, 442–457.

R Development Core Team (2005) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Robert, C. P. and Casella, G. (2004) *Monte Carlo Statistical Methods*. Second edition. Springer.

Rutherford, D. (1965) *Introduction to Lattice Theory*. 7th edition. UK: Oliver and Boyd, Edinburgh.

Schmidt, P. (1985) Frontier production functions. *Econometric Reviews* **4**, 289–328.

Simar, L. (2007) How to improve the performances of DEA/FDH estimators in the presence of noise. *Journal of Productivity Analysis* **28**, 183–201.

Simar, L. and Wilson, P. (1998) Sensitivity of efficiency scores: How to bootstrap in non-parametric frontier models. *Management Science* **44**, 49–61.

Simar, L. and Wilson, P. (2000) A general methodology for bootstrapping in nonparametric frontier models. *Journal of Applied Statistics* **27**, 779–802.

Simar, L. and Wilson, P. (2007) Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics* **136**, 31–64.

Simar, L. and Wilson, P. W. (2015) Statistical approaches for non-parametric frontier models: A guided tour. *International Statistical Review* **83**, 77–110.

Tsionas, E. G. and Papadakis, E. N. (2010) A Bayesian approach to statistical inference in stochastic DEA. *Omega* **38**(5), 309–314.

Vorob'ev, O. (1984) *Srednemernoje Modelirovanie (Mean-Measure Modelling)*. Moscow: Nauka.