Les Cahiers du GERAD ISSN: 0711–2440

Stochastic Data Envelopment Analysis: A Random Efficiency Perspective

M. Asgharian, M. Davtalab-Olyaie,
V. Partovi Nia

G–2013–96

December 2013

Stochastic Data Envelopment Analysis: A Random Efficiency Perspective

Masoud Asgharian

Department of Mathematics and Statistics
McGill University
Montréal (Québec) Canada

masoud@math.mcgill.ca

Mostafa Davtalab-Olyaie

Department of Mathematics and Computer Sciences
Kharazmi University
Tehran, Iran

m.davtalab-olyaie@tmu.ac.ir

Vahid Partovi Nia

GERAD & Department of Mathematics and Industrial Engineering
Polytechnique Montréal
Montréal (Québec) Canada

vahid.partovinia@polymtl.ca

December 2013

Les Cahiers du GERAD

G–2013–96

Abstract: Since its inception, stochastic Data Envelopment Analysis (DEA) has found many applications. The approach commonly taken in stochastic DEA is via chance constraint models. This approach cannot, however, capture the inherent random fluctuations of the efficiency score caused by the random nature of the input and output variables. Having taken a random efficiency perspective, one can introduce an alternative approach that provides ground for capturing these fluctuations. One aspect of this alternative approach, which seems to have been neglected, concerns the distribution of the random efficiency score. We show that the efficiency score does not have a continuous distribution even if the random input and output variables distributions are continuous. The efficiency score distribution has, in fact, a point mass decomposition at 1. This observation renders the non-parametric bootstrap of efficiency score impossible. We introduce several criteria for the ranking and classification of Decision Making Units (DMUs) using a random efficiency perspective, including an interactive ranking method that incorporates managers' knowledge and preferences. We then apply the point mass decomposition of efficiency score distributions of DMUs and show how these criteria can be implemented. We also discuss how one may estimate the efficiency score distributions of DMUs using both Bayesian and frequentist approaches. Our proposed methodology is illustrated using a real data set.

Key Words: Admissibility; Efficiency Distribution; Empirical Bayes Approach; Interactive Ordering; Stochastic Data Envelopment Analysis; Stochastic Ordering.

1 Introduction

Data Envelopment Analysis (DEA), developed by Charnes et al. (1978, 1979) and extended by Banker et al. (1984), is a non-parametric approach that offers an efficiency analysis for a set of multi-input multi-output decision making units (DMUs). In many applications, uncertainty is associated with some of the input and/or output data. One way to model such uncertainty is by imposing a probability distribution on the input/output data. A common approach to handle the uncertainty is via chance constraint models where the random Production Possibility Set (PPS) is replaced by an average PPS where average is in the sense of Vorob'ev (1984). There has been a surge of articles on the stochastic DEA over the past two decades. Some of the early work on this subject was carried out by Land et al. (1993), Olesen and Petersen (1995), and Cooper et al. (1998), among others. A recent review of the 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 constraint models. An attempt towards this goal was taken by Cooper et al. (1999), Cooper et al. (2001), Despotis and Smirlis (2002), and Kao (2006) using interval data. As discussed by Kao and Liu (2009), this approach is not effective when the intervals are wide since the efficiency interval becomes too wide to permit reasonable conclusions.

Having taken into account the stochastic nature of the efficiency score, Kao and Liu (2009) use a truncated Beta distribution and Monte Carlo estimation approach to study fluctuation of the efficiency score. Lamb and Kai-Hong (2012) use Efron's non-parametric bootstrap of efficiency score to present a more objective analysis and suggest a stochastic ranking based on bootstrap confidence intervals for efficiency scores.

A fundamental aspect of these stochastic approaches, which seems to have been neglected, concerns the distribution of the random efficiency measure. As we show in Theorem 1 the efficiency score distribution does not have a continuous distribution even if both the random input and output variables are continuous. The efficiency score has, in fact, a point mass decomposition at 1, except for DMUs which are almost surely inefficient. The ramifications of Theorem 1 abound. One immediate consequence of this result is that the non-parametric bootstrap of efficiency is impossible when the probability mass at 1 is unknown, which is the case in most, if not all, practical applications.

This particular structure of the efficiency score distribution calls for a more careful treatment of the performance assessment of DMUs. We introduce several measures for ranking DMUs. We start with a partial ordering that leads to the introduction of a minimal requirement, which we call *admissibility*, and which we can use to categorize DMUs into two categories, namely admissible and inadmissible DMUs. Using the point mass decomposition of the efficiency distribution (Theorem 1), we then provide a sufficient condition for admissibility. We further suggest several complete (linear) orderings using the efficiency score distribution, including an interactive ordering that can incorporate preferences of the production manager.

Implementation of these ranking methods require estimation of the efficiency score distributions of DMUs. We simulate the input and output data for each DMU and measure the efficiency score of each DMU using a conventional DEA model, for instance the CCR model, for each set of simulated data. This approach produces a sample from the efficiency score distribution and then using standard statistical methods, we estimate the efficiency score distributions of DMUs. To this end, we discuss both the frequentist and Bayesian approaches. We use the bootstrap method for the former approach (Efron and Tibshirani, 1993), while the latter is achieved using the Markov chain Monte Carlo (MCMC) method (Robert and Casella, 2004).

The frequentist approach has recently been discussed by Kao and Liu (2009) and Lamb and Kai-Hong (2012), who used bootstrap for resampling efficiency score, among others. 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. In contrast, we take a more objective 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. Section 2 includes some preliminaries needed in the next sections. In Section 3 we discuss the point-mass structure of the efficiency distribution. We define several ranking methods for stochastic DMUs using efficiency distribution and explore their relationships in Section 4. Estimation of the efficiency distribution using both Bayesian and Frequentist approach is studied in Section 5. We illustrate our methods using a set of real data in Section 6. The last section, Section 7, includes some possible further extensions of the methods presented earlier.

2 Preliminaries

In this section we recall some basic concepts of deterministic DEA. Consider a set of n DMUs, each using m inputs to produce s outputs. The ith input variable ($i = 1, \ldots, m$) and rth output variable ($r = 1, \ldots, s$) of jth DMU ($j = 1 \ldots, n$) are denoted by x_{ij}, and y_{rj}, respectively. The inputs and outputs, x_{ij} and y_{rj}, are all assumed to be non-negative for $i = 1, \ldots, m$, $r = 1, \ldots, s$, and $j = 1, \ldots, n$. The Production Possibility Set (PPS), denoted by T, is the set of all feasible activities,

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

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

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

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 T_{CCR} (Charnes et al., 1978); while setting $L = U = 1$, variable return to scale assumption, gives T_{BCC} (Banker et al., 1984). Should we take T_{CCR}, for instance, we can evaluate the relative efficiency by solving the CCR model

$$\begin{aligned} \theta_o = \min \quad & \theta \\ \text{s.t.} \quad & \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{io}, i = 1, \ldots, m \\ & \sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro}, r = 1, \ldots, s \\ & \lambda_j \geq 0, j = 1, \ldots, n. \end{aligned} \tag{2}$$

If $\theta_o = 1$, then DMU$_o$ is CCR-efficient. We confine our attention to the CCR model to simplify our disscution in the sequal, though our approach is equally applicable to other DEA models.

3 Efficiency Distribution

The efficiency evaluation discussed in Section 2 is suitable for deterministic data. However, it often happens that we need to take into account some uncertainly associated with the inputs and/or outputs of DMUs. As discussed in Section 1, one may consider inputs and outputs as random variables distributed according to some distribution and develop stochastic versions of the aforementioned models. To distinguish between a random variable and its realized values, in the sequel we use capital letters to denote random variables and retain small letters for their realized values. We therefore consider the input and output vectors, X_j and Y_j associated with DMU$_j$, to be random vectors of dimension m and s, respectively. All variables are defined on the probability space $(\Omega, \Im, P)$, where $\Im$ is a σ-algebra of the subsets of Ω and P is a probability measure on $\Im$. The stochastic counterpart of T_{CCR} is

$$\mathcal{T} = \left\{ Z = (X, Y) \mid X_i \geq \sum_{j=1}^{n} \lambda_j X_{ij}, \forall i; Y_r \leq \sum_{j=1}^{n} \lambda_j Y_{rj}, \forall r; \lambda_j \geq 0, j = 1, \ldots, n \right\}. \tag{3}$$

Given that the inputs and outputs are all random vectors, the efficiency Θ, being a function of these quantities, is also a random variable. The following result shows that there is at least one DMU whose efficiency score distribution has a point mass at 1 irrespective of the type input and output variables.

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

See Appendix I for the proof.

In any realistic situation, there should be more than one DMU whose efficiency score carries some positive mass at 1. Otherwise, the only one with positive mass should be almost surely efficient. Of course, there may exist some DMUs which are almost surely inefficient and hence their efficiency score distribution does not have any positive mass at 1. This case does not, however, seem likely in practical applications of stochastic DEA. Given the stochastic nature of the outputs and inputs, even a DMU with mostly weak performance may have a small chance to perform strongly, and vice versa. Let F_{Θ_j} be the cumulative distribution function (cdf) of Θ_j for $j = 1, \ldots, n$, using Theorem 1 we have the following decomposition for the efficiency score of each DMU_j, $j = 1, \ldots, n$,

$$\bar{F}_{\Theta_j}(\theta) = p_j + (1 - p_j)\bar{G}_{\Theta_j}(\theta), \tag{4}$$

where $p_j = P(\Theta_j = 1)$, and $\bar{F}_{\Theta_j}(\theta) = 1 - F_{\Theta_j}(\theta)$. Similarly we define $\bar{G}_{\Theta_j}(\theta) = 1 - G_{\Theta_j}(\theta)$, where G_{Θ_j} represents the cdf of the inefficiency component of the efficiency score distribution. For the sake of simplicity, we use F_j and G_j to denote F_{Θ_j} and G_{Θ_j}, for $j = 1, \ldots, n$ hereafter.

From an analytic perspective, Theorem 1 tells us that the probability measure of Θ_j is not dominated by the Lebesgue measure even if the probability measure of inputs and outputs, X_j and Y_j, are both dominated by the Lebesgue measure.

4 Ranking Methods for Stochastic DMUs

Suppose Θ_j is the efficiency of DMU_j. Having obtained the distribution of Θ_j, for $j = 1, \ldots, n$, ranking DMUs is possible using various methods. The simplest ranking method is through ranking p_j, which we may call p-ranking. Alternatively, one may use different measures of central tendency, such as mean, median or quantiles of $F_j(.)$, for $j = 1, \ldots, n$ to rank DMUs. These ranking methods may be called, mean, median and quantile ranking, respectively. While the aforementioned ranking methods are all based on a summary of $F_j(\cdot)$, 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., $F_j(\cdot)$, that encompasses all the information about Θ_j.

4.1 Ranking Using Stochastic Ordering

Let Ψ be a subset of [0,1].

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

$$\overline{F}_j(\theta) \geq \overline{F}_{j'}(\theta), \quad \forall\theta \in \Psi.$$

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

Figure 1 illustrates the stochastic ordering notion. It depicts the probability density functions (pdf), $f(\theta)$, and the survival functions, $\overline{F}(\theta)$, 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. This shows that the survival function of DMU_1 is uniformly dominated by the survival function of DMU_2. That is, the performance of DMU_2 is always superior to that of DMU_1 probabilistically. 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.

Next we investigate the relationship between this general ranking method with the simple and summary based ranking techniques mentioned above. Let the mean of the random variable Θ_j be denoted by $\mathbb{E}(\Theta_j)$

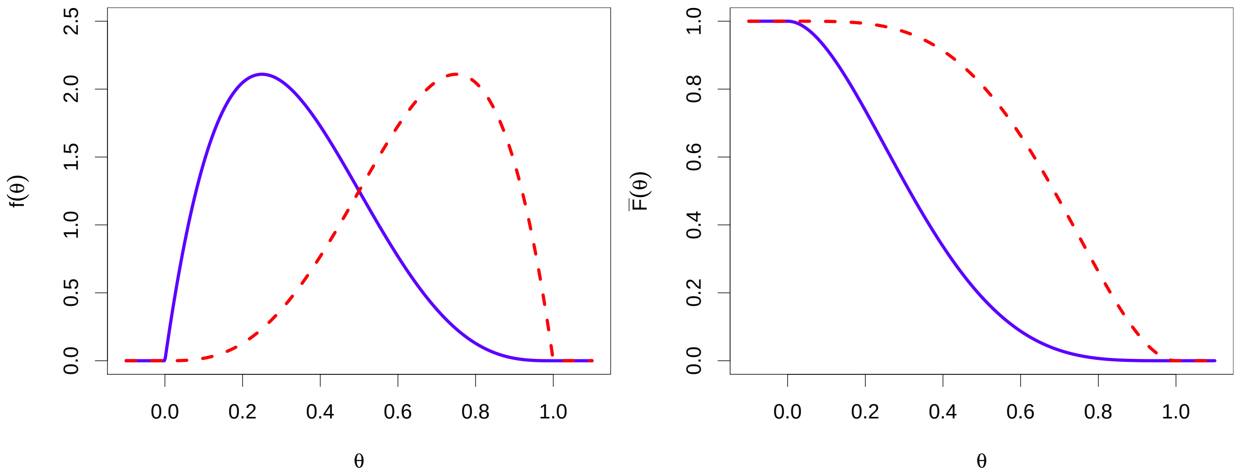

Figure 1: Stochastic ordering of efficiency distributions of DMU_1 (solid curve) versus DMU_2 (dashed curve). Comparison densities, left panel, and survival function, right panel (see Definition 1).

and its β-quantile by $\tilde{\mathbb{E}}_\beta(\Theta_j)$, where $0 < \beta < 1$. Each of these quantities can be used for a complete (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 2 *If $\Theta_j \succ \Theta_{j'}$ then $p_j > p_{j'}, \mathbb{E}(\Theta_j) > \mathbb{E}(\Theta_{j'})$, 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 p-ranking method is perhaps the simplest method of ranking among the methods suggested above. Theorem 2 shows that ranking DMUs using stochastic ordering implies p-ranking. The following result establishes a partial reverse.

Theorem 3 *Let $\Psi = [\psi, 1]$ and $p_j \geq p_{j'}$. Then $\Theta_j \succ_\Psi \Theta_{j'}$, if one of the following three conditions holds*

1. $\inf_{\theta\in\Psi} \left\{\overline{G}_{j'}(\theta) - \overline{G}_j(\theta)\right\} > 0$,
2. $\inf_{\theta\in\{\theta|\overline{G}_j(\theta)>0\}\cap\Psi} \left\{\frac{\overline{G}_j(\theta)-\overline{G}_{j'}(\theta)}{\overline{G}_j(\theta)}\right\} > \frac{p_j-p_{j'}}{1-p_j}$,
3. $\inf_{\theta\in\{\theta|\overline{G}_{j'}(\theta)>0\}\cap\Psi} \left\{\frac{\overline{G}_j(\theta)-\overline{G}_{j'}(\theta)}{\overline{G}_{j'}(\theta)}\right\} > \frac{p_j-p_{j'}}{1-p_{j'}}$.

See Appendix I for the proof.

The following corollary follows immediately.

Corollary 1 *If $\overline{G}_j(\theta) = \overline{G}_{j'}(\theta)$ for all $\theta \in \Psi$, then $\Theta_j \succ_\Psi \Theta_{j'}$ if and only if $p_j > p_{j'}$.*

The notion of inadmissibility was introduced in Definition 1. To further investigate and distinguish inadmissible DMUs from the admissible ones, i.e. those not inadmissible, in $\mathcal{T}$, we need the following definition. Let $\Gamma = \{\Theta_Z : Z \in \mathcal{T}\}$ where Θ_Z is the efficiency variable of Z and $\mathcal{F} = \{\overline{F}_\Theta : \Theta \in \Gamma\}$.

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

Direct verification of admissibility using Definition 2 is cumbersome. Using the mass point decomposition of the efficiency distribution, equation 4, 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.

The notion of admissibility by means of the stochastic ordering provides a partial ranking of DMUs. In Section 4.2 we propose another ranking method which provides a complete (linear) ordering of DMUs by taking the production manager preference into account.

4.2 Ranking Using Interactive Ordering

Ranking using stochastic ordering when the efficiency level is restricted to Ψ is motivated by possible preferences a manager may attribute to different levels of efficiency. If the preferences can be modeled in terms of relative weights, a finer notion of stochastic efficiency which is interactively defined, can be introduced. Suppose π is a probability measure which represents these relative weights on [0,1] modeled with preferential information given by the manager. Let $I_\pi(\overline{F}_j) = \int_\theta \overline{F}_j(\theta)\pi(d\theta)$. Then $I_\pi(\overline{F}_j)$ can essentially be interpreted as an interactive efficiency score according to π for each DMU_j, $j = 1, \ldots, n$.

Definition 3 *We say DMU_j is interactively more efficient than $DMU_{j'}$, if $I_\pi(\overline{F}_j) \geq I_\pi(\overline{F}_{j'})$.*

While Definition 1 only allows a partial ordering of DMUs, a complete (linear) ordering is possible using Definition 3. Although a complete (linear) ordering of DMUs can also be achieved using measures of central tendencies such as mean and median of the efficiency score distribution, the ranking using interactive ordering offers an adaptive approach to a manager's preferences and better interpretation. A manager can specify different weights over different regions in $[0, 1]$ through the probability measure π. The following theorem establishes a close tie between admissibility and ranking using interactive ordering.

Theorem 5 *$\overline{F}_j \in \mathcal{G}$ is admissible with respect to $\mathcal{G}$, if there exists a π such that $I_\pi(\overline{F}_j) > I_\pi(\overline{F})$ for all $\overline{F} \in \mathcal{G}$.*

See Appendix I for the proof.

5 Efficiency Estimation

Implementing the above ranking methods requires estimation of the efficiency score distribution. Suppose $Z_j \sim f_{Z_j}(. \mid \nu_j)$ where the pdf f_{Z_j} is known up to finitely many unknown parameters $\nu_j, j = 1, \ldots, n$. To generate a sample from the efficiency score distribution of DMU_j, $j = 1, \ldots, n$, one needs a sample from each DMU_j, $j = 1, \ldots, n$, i. e., a sample from the PPS. This can be achieved using either the Bayesian or frequentist perspective. We discuss both approaches in the following sections.

Motivated by our example in Section 6, 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 $t = 1, \ldots, T$. A stochastic approach seems more reasonable for such a framework since the values of inputs and outputs can vary through time. Denote stochastic DMU_j at time t by $Z_{jt} = (X_{jt}, Y_{jt})'$ for $j = 1, \ldots, n$ 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 and " $'$ " represents transpose. Therefore we essentially have a PPS, T_t 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 $\mathbf{z}_{ij} = [z_{ijt}], t = 1, \ldots, T$. We assume that DMU_1, ..., DMU_n are independent, but may not be identically distributed.

5.1 Bayesian Perspective

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 commonly assume a proper prior in Bayesian analysis, i.e., we consider a class of prior distributions such that $\int_{-\infty}^{+\infty} f(\vartheta_o \mid \varphi) d\vartheta_o = 1$. We take an empirical Bayes approach, and devise a numerical approximation using sampling from the posterior predictive distribution of data. In order to sample from the posterior predictive distribution, parameters of the prior distribution, the so-called hyper-parameters, must either be known or estimated using the 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 expand this road map and explain how this approach can be implemented.

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

$$f(\mathbf{z} \mid \varphi) = \prod_{j=1}^{n} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(\mathbf{z}_{j1},\ldots,\mathbf{z}_{jt},\ldots,\mathbf{z}_{jT} \mid \vartheta_j) f(\vartheta_j \mid \varphi) d\vartheta_j. \tag{5}$$

The Empirical Bayes estimate of φ is

$$\varphi_{\max} = \text{argmax}_{\varphi} \log f(\mathbf{z} \mid \varphi). \tag{6}$$

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

$$f(\mathbf{z}_o^* \mid \mathbf{z}_{o1},\ldots \mathbf{z}_{oT},\varphi_{\max}) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(\mathbf{z}_o^* \mid \vartheta_o) f(\vartheta_o \mid \mathbf{z}_{o1},\ldots,\mathbf{z}_{oT},\varphi_{\max}) d\vartheta_o, \tag{7}$$

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 T^* which itself requires a predictive sample of all DMUs. Having produced T^*, a sample from the efficiency distribution of each DMU (including DMU_o) can be obtained by solving the CCR model on T^*. We repeat this procedure B times, for B large enough, to find B samples from the efficiency distribution of each DMU. The pseudo-code for this procedure is provided below:

1. Estimate hyperparameters by $\phi_{\max} = \text{argmax} \log f(\mathbf{z} \mid \phi)$.
2. For all $j = 1,\ldots,n$, sample from posterior of data parameters

$$\vartheta_j^{\text{post}} \sim f(\vartheta_j \mid \mathbf{z}_{j1},\ldots,\mathbf{z}_{jT},\varphi_{\max}).$$

3. For all $j = 1,\ldots,n$, sample a new data point of DMU_j by sampling from

$$z_j^* \sim f(\mathbf{z}_j \mid \vartheta_j^{\text{post}}).$$

4. For all $j = 1,\ldots,n$, compute θ_j^* by solving model (2).
5. Repeat steps 1 to 4, B (large enough) times to produce B samples $(\theta_{1b}^*,\ldots,\theta_{nb}^*)$, for $b = 1,\ldots,B$.

We note that our algorithm generates B samples from the multivariate density $f(\theta_1, \ldots, \theta_n \mid \phi_{\max})$. We then use these B samples to estimate the unknown parameters of the efficiency distribution for each DMU, namely p_o and $\overline{G}_o(.)$.

The estimate of p_o, say $\hat{p}_o$, is the proportion of θ_o^* being one out of B samples. The non-parametric maximum likelihood estimate of $\overline{G}_o(.)$ based on a sample of size B, $\theta_{o1}^*, \ldots, \theta_{oB}^*$, is given by

$$\hat{\overline{G}}_{oB}(t) = \frac{1}{B} \sum_{i=1}^{B} \varepsilon_{\{t \mid \theta_{oi} > t\}}(t),$$

where $\varepsilon_A(x) = 1$ if $x \in A$, and equal to zero otherwise. One can show that

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

When the input and output variables are continuous, $G_o(\theta) = 1 - \overline{G}_o(\theta)$ has a density with respect to the Lebesgue measure which can be estimated using the kernel, or other, density estimation method if visualization of the density is required.

5.2 Frequentist Perspective

In the frequentist approach, the parameter ν is treated as an unknown, but fixed. Given that finding a closed form for $\overline{G}_\Theta$ is cumbersome, one usually resorts to computational approaches to approximate p and $\overline{G}_\Theta$, and hence $\overline{F}_\Theta$. The main computational tool for such purposes in the frequentist approach is bootstrapping (Efron and Tibshirani, 1993). As we briefly discussed in the introduction, bootstrapping a mixture distribution such as F_Θ, when the mixing proportion p is unknown is not directly possible. One can, however, use the bootstrap method to generate Z_j, $j = 1, \ldots, n$ and repeatedly solve a conventional DEA model, such as CCR or BCC model to generate samples from F_Θ. If ν_j is known, this approach can be readily implemented. Otherwise, one needs several replications on each Z_j, $j = 1, \ldots, n$. In longitudinal studies on the performance of a DMU, we often collect repeated measurements on the input and output variables. If such repeated measurements can be assumed independent, then bootstrap can readily be implemented. Otherwise, the dependence of such repeated measurements should either be estimated or modeled. Short follow-up studies render dependence estimation infeasible. Long follow-up studies, in contrast, allow estimation of a possible dependence structure using block bootstrap, see Kunsch (1989), Carlstein et al. (1998), and Inoue and Kilian (2002).

6 Data Analysis

We illustrate the methodologies developed in the previous sections using the airline data of Greene (2011).[1] We 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. To simplify the computational aspects of our illustration, in the sequel we assume that the data are independent through time. Let

$$\begin{aligned} \mathbf{z}_{jt} \mid \mu_j, \mathbf{\Sigma}_j &\overset{\text{iid}}{\sim} \mathcal{N}_{(m+s)}(\mu_j, \mathbf{\Sigma}_j), \\ \mu_j \mid \tau, \kappa, \mathbf{\Sigma}_j &\overset{\text{iid}}{\sim} \mathcal{N}_{(m+s)}(\nu, \kappa\mathbf{\Sigma}_j), \\ \mathbf{\Sigma}_j &\overset{\text{iid}}{\sim} \mathcal{W}^{-1}(a, \mathbf{\Psi}), \end{aligned} \tag{8}$$

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

$$f(\mathbf{z} \mid \varphi) = \prod_{j=1}^{n} \prod_{i=1}^{m+s} \frac{b^{\frac{a}{2}} \Gamma\left(\frac{a+T}{2}\right)}{\pi^{\frac{T}{2}} |\mathbf{V}|^{\frac{1}{2}} \Gamma\left(\frac{a}{2}\right) \left\{b + (\mathbf{z}_{ij} - \tau\mathbf{1})' \mathbf{V}^{-1} (\mathbf{z}_{ij} - \tau\mathbf{1})\right\}^{\frac{a+T}{2}}}, \tag{9}$$

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

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

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. Given that the input and output variables are all continuous variables, the inefficiency component in (4), $\overline{G}$, has the following integral representation $\overline{G}(\theta) = \int_\theta^1 g(u)du$ where $g(\theta)$ is the pdf of the inefficiency component in (4).

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 R 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 and the estimation of the inefficiency density, $g(.)$, is reported in Figure 3. In the upper panel of Figure 2 the continuous parts are almost uniform. In the bottom 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 bottom 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 result 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 $\Theta_k \succ \Theta_j$, and there is no i such that $\Theta_k \succ \Theta_i$ and $\Theta_i \succ \Theta_j$, see Rutherford (1965). Given that ranking using stochastic ordering provides the most comprehensive ranking method, the arrows in the Hess diagram in Figure 2 show which dominance in DMUs cannot be changed through different specifications of the probability measure Π in ranking using interactive ordering or 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 using different weighting measures π or 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 $\mathcal{T}$.

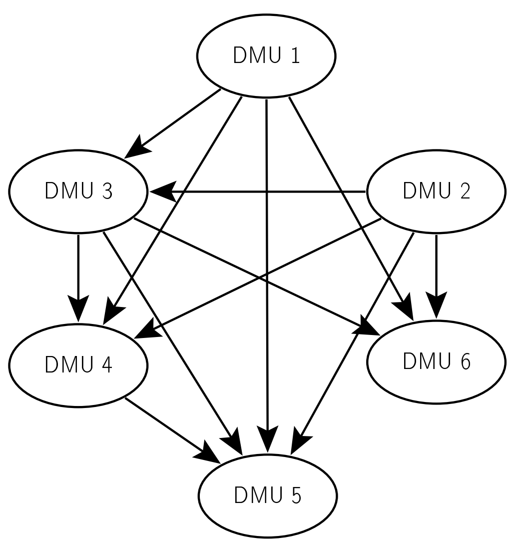

Figure 2: Hasse diagram of DMU domination, visualizing the result of ranking using stochastic ordering in Table 1.

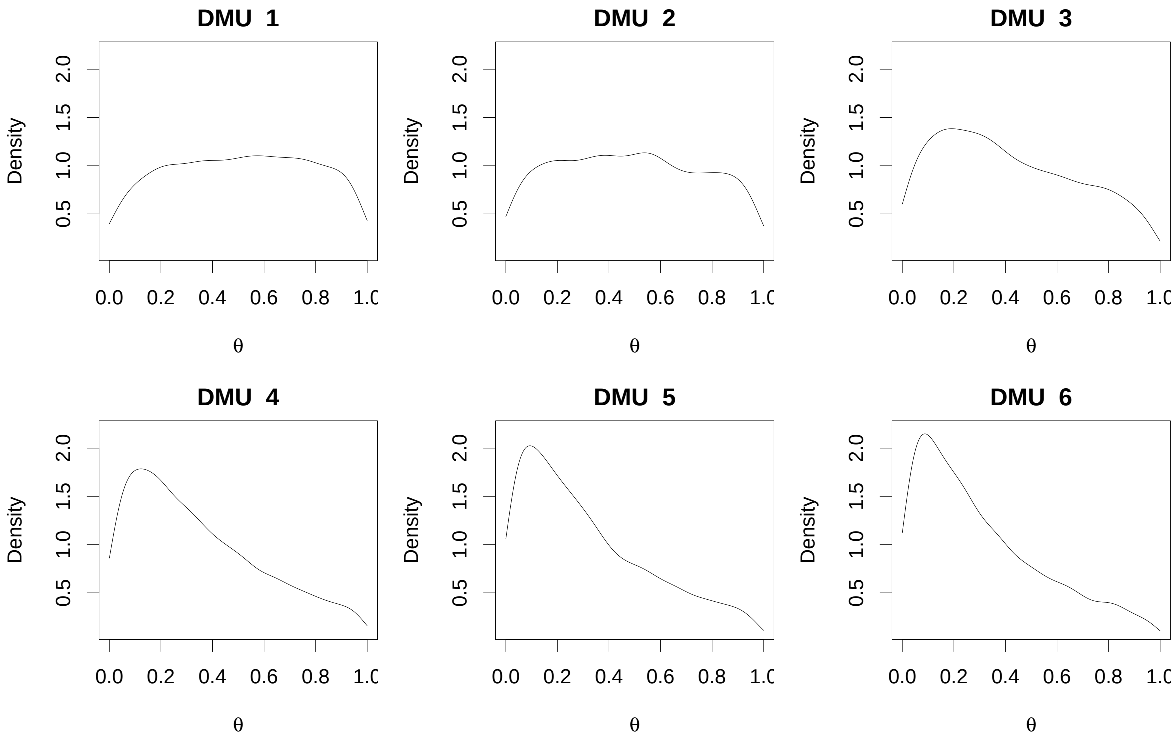

Figure 3: Estimation of the continuous part of DMUs $g_j(.)$ using kernel density estimation.

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

	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	$\{3,4,5,6\}$	$\{3,4,5,6\}$	$\{4,5,6\}$	$\{5\}$	$\{\}$	$\{\}$
interactive efficiency index $\hat{I}_\Pi(\overline{F})$	0.83	0.74	0.58	0.40	0.36	0.34

Several DMUs remain unordered according to the stochastic ordering method. The preference of the manager can be used to achieve a complete ordering of DMUs. An example of the preference of the manager is

$$\begin{cases} 6 & \text{if } \theta \in [0.9, 1] \\ 3 & \text{if } \theta \in [0.5, 0.9) \\ 1 & \text{if } \theta \in [0, 0.5). \end{cases} \tag{10}$$

This preference can be translated into the probability measure $\pi(\theta)$ by re-normalizing this function to satisfy $\int_0^1 \pi(\theta) d\theta = 1$.

$$\pi(\theta) = \begin{cases} 2.609 & \text{if } \theta \in [0.9, 1] \\ 1.304 & \text{if } \theta \in [0.5, 0.9) \\ 0.435 & \text{if } \theta \in [0, 0.5). \end{cases} \tag{11}$$

After assigning $\pi(\theta)$, the interactive efficiency index, $I_\pi(\overline{F})$, can be calculated for all DMUs for a given $\overline{F}$. The estimation of the interactive efficiency index is achieved by substituting $\overline{F}$ with its empirical estimate $1 - \hat{F}$, in which $\hat{F}$ is the empirical cumulative distribution function,

$$\hat{I}_\pi(\overline{F}) = \int_0^1 \{1 - \hat{F}(\theta)\} \pi(\theta) d\theta.$$

According to the last row of Table 1, interactive ordering using (11) suggests that DMU_1 has the best performance.

7 Further Extensions

Up to now we have only considered a deterministic combination of DMUs, i.e., using deterministic λ. A possible extension is to allow random combinations of DMUs, i.e., $\zeta DMU_i + (1-\zeta)DMU_j$ where ζ is also a random variable defined on the same probability space as the DMUs. This then amounts to defining PPS path-by-path, i.e., for each $\omega \in \Omega$. Having taken this view, our PPS for each $\omega \in \Omega$ is

$$\tilde{T}(\omega) = \left\{ Z(\omega) = \begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix} \mid X_i(\omega) \geq \sum_{j=1}^{n} \lambda_j(\omega) X_{ij}(\omega), \forall i; Y_r(\omega) \leq \sum_{j=1}^{n} \lambda_j(\omega) Y_{rj}(\omega), \forall r; \lambda_j(\omega) \geq 0, \forall j \right\}. \tag{12}$$

Denote the associated set of efficiency variables and their survival distributions by $\tilde{\Gamma}$ and $\tilde{\mathcal{F}}$, respectively. It is clear that $\Gamma \subseteq \tilde{\Gamma}$ and $\mathcal{F} \subseteq \tilde{\mathcal{F}}$. The following result presents a necessary condition for admissibility in $\tilde{\mathcal{F}}$.

Theorem 6 *Suppose $\overline{F}_j \in \tilde{\mathcal{F}}$ is admissible with respect to $\tilde{\mathcal{F}}$. Then there exists a probability measure π on [0,1] such that $I_\pi(\overline{F}_j) \geq I_\pi(\overline{F}), \forall \overline{F} \in \tilde{\mathcal{F}}$.*

See Appendix I for the proof.

Remark 2 *We should note that $\tilde{\mathcal{F}}$ is typically a much larger set than $\mathcal{F}$. As such, the requirement for admissibility in $\tilde{\mathcal{F}}$ is a more stringent condition to be fulfilled than admissibility in $\mathcal{F}$. Theorem 6 holds true only for a subset of admissible elements of $\mathcal{F}$; those which are not dominated by convex combination of elements of $\mathcal{F}$ either.*

Remark 3 *An inspection of Theorem 4 and 5 shows that these theorems also hold true in this setting.*

Appendix I

Proof of Theorems

Proof of Theorem 1: We first note that Θ is a random variable defined on the probability space $(\Omega, \Im, P)$. We further note that for any $\omega \in \Omega$ we have a PPS. Let $A_i = \{\omega \in \Omega : \Theta_i(\omega) = 1\}$ for $i = 1, \ldots, n$. Since in any PPS there is at least one efficient DMU, we have $\Omega = \bigcup_{i=1}^{n} A_i$. Now suppose that there is no mass point at 1 for any DMU, i.e., $p_i = P(\omega : \Theta_i(\omega) = 1) = P(A_i) = 0$ for $i = 1, \ldots, n$. Then using Boole's inequality, $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i) = 0$. On the other hand, $P(\bigcup_{i=1}^{n} A_i) = P(\Omega) = 1$. This is a contradiction. $\square$

Proof of Theorem 3: Suppose $p_j > p_{j'}$ and Condition 1 is satisfied; i.e., $G_{j'}(\theta) > G_j(\theta)$, $\forall \theta \in \Psi$. Then

$$(1-p_{j'})G_{j'}(\theta) - (1-p_j)G_j(\theta) > 0,$$

which implies

$$(1-p_{j'})\left(1-\bar{G}_{j'}(\theta)\right) - (1-p_j)\left(1-\bar{G}_j(\theta)\right) > 0,$$

and hence

$$p_j + (1-p_j)\bar{G}_j(\theta) > p_{j'} + (1-p_{j'})\bar{G}_{j'}(\theta).$$

Therefore $\overline{F}_j(\theta) > \overline{F}_{j'}(\theta), \forall \theta \in \Psi$; which implies $\Theta_j \succ_\Psi \Theta_{j'}$.

Suppose Condition 2 holds, that is, $\forall \theta \in \{\theta \mid G_j(\theta) > 0\} \cap \Psi$, and

$$\inf_{\theta \in \Psi} \left\{ \frac{G_j(\theta) - G_{j'}(\theta)}{G_j(\theta)} \right\} > \frac{p_j - p_{j'}}{1-p_j},$$

This is equivalent to

$$\frac{\bar{G}_{j'}(\theta) - \bar{G}_j(\theta)}{1 - \bar{G}_{j'}(\theta)} > \frac{p_j - p_{j'}}{1 - p_j}, \forall \theta \in \Psi,$$

hence

$$\left\{1 - \bar{G}_j(\theta)\right\} - \left\{1 - \bar{G}_{j'}(\theta)\right\} - \frac{\left\{1 - \bar{G}_{j'}(\theta)\right\}}{1 - p_j} \left\{(1 - p_{j'}) - (1 - p_j)\right\} > 0,$$

which, in turn, implies

$$(1 - p_j)\left\{1 - \bar{G}_j(\theta)\right\} - (1 - p'_j)\left\{1 - \bar{G}_{j'}(\theta)\right\} > 0.$$

Thus $p_j + (1 - p_j)\bar{G}_j(\theta) > p_{j'} + (1 - p_{j'})\bar{G}_{j'}(\theta)$, and therefore $\overline{F}_j(\theta) > \overline{F}_{j'}(\theta), \forall \theta \in \Psi$; yielding $\Theta_j \succ_\Psi \Theta_{j'}$.

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 $\tilde{\lambda} = \left(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n\right) \geq 0$ such that $P(\Omega_{\tilde{\lambda}}) = P(\Theta_{\tilde{\lambda}} \geq \Theta_o) \geq 2p_o - \frac{p_o^2+1}{2}$.*

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

$$\overline{F}_{\tilde{\lambda}}(\theta) \geq \overline{F}_o(\theta), \forall \theta \in [0, 1]. \tag{13}$$

For any $\lambda \geq 0$ define $\Omega_\lambda = \{\omega \in \Omega \mid \Theta_\lambda(\omega) > \Theta_o(\omega)\}$. We have

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

where

$$\begin{aligned} P(\Theta_{\tilde{\lambda}} \geq \Theta_o \mid \Theta_o < 1) &= \int_0^1 P(\Theta_{\tilde{\lambda}} \geq \Theta_o \mid \Theta_o = \theta, \Theta_o < 1) dF(\theta \mid \Theta_o < 1) \\ &= \int_0^1 P(\Theta_{\tilde{\lambda}} \geq \Theta_o \mid \Theta_o = \theta, \Theta_o < 1) \frac{(1 - p_o)}{P(\Theta_o < 1)} dG_{\Theta_o}(\theta) \\ &= \int_0^1 P(\Theta_{\tilde{\lambda}} \geq \Theta_o \mid \Theta_o = \theta, \Theta_o < 1) dG_{\Theta_o}(\theta). \end{aligned}$$

We note that DMU$_o$ is not on the boundary 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 $\Theta_{\tilde{\lambda}}$ is independent of Θ_o given $\omega \in \bigcup_\lambda \Omega_\lambda$. We therefore have

$$P(\Theta_{\tilde{\lambda}} \geq \Theta_o \mid \Theta_o < 1) = \int_0^1 P(\Theta_{\tilde{\lambda}} \geq \theta) dG_{\Theta_o}(\theta) = \int_0^1 \overline{F}_{\Theta_{\tilde{\lambda}}}(\theta) dG_{\Theta_o}(\theta).$$

Using 13,

$$P(\Theta_{\tilde{\lambda}} \geq \Theta_o \mid \Theta_o < 1) \geq \int_0^1 \overline{F}_{\Theta_o}(\theta) dG_{\Theta_o}(\theta).$$

On the other hand, we know

$$\overline{F}_{\Theta_o}(\theta) = p_o + (1 - p_o)\overline{G}_{\Theta_0}(\theta), \forall \theta \in [0, 1]; \text{ and hence}$$
$$\int_0^1 \overline{F}_{\Theta_o}(\theta) dG_{\Theta_o}(\theta) = p_o + (1 - p_o) \int_0^1 \overline{G}_{\Theta_o}(\theta) dG_{\Theta_o}(\theta) = p_o + \frac{(1 - p_o)}{2} = \frac{1 + p_o}{2}.$$

Then

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

□

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

On the other hand, $\{\omega \in \Omega \mid \Theta_o(\omega) = 1\} = (\bigcup_\lambda \Omega_\lambda)^c$, where ^{c} stands for the complement. Thus

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

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. □

Proof of Theorem 5: Suppose DMU$_o$ is inadmissible. Then there exists an $\overline{F}_* \in \mathcal{G}$ such that $\overline{F}_*(\theta) \geq \overline{F}_o(\theta), \forall\theta$. Let π be a probability measure on [0,1]. Then

$$I_\pi(\overline{F}_*) = \int_\theta \overline{F}_*(\theta)\pi(d\theta) \geq \int_\theta \overline{F}_o(\theta)\pi(d\theta) = I_\pi(\overline{F}_o).$$

This is a contradiction. □

To prove Theorem 6, we first need establish the following lemma.

Lemma 2 *$\tilde{\mathcal{F}}$ is a convex set.*

Proof. Suppose $\overline{F}^{'}$ and $\overline{F}^{''} \in \tilde{\mathcal{F}}$. Then there exist $\Theta_1^{'}$ and $\Theta_2^{''}$ in the set $\tilde{\Gamma}$ associated with $\overline{F}^{'}$ and $\overline{F}^{''}$ respectively, and hence we have DMU$^{'}$ and DMU$^{''}$, possibly virtual DMUs, associated with $\Theta_1^{'}$ and $\Theta_2^{''}$. Consider $\alpha\overline{F}^{'} + (1-\alpha)\overline{F}^{''}$ where $0 \leq \alpha \leq 1$. Note that $\alpha\overline{F}^{'}(t) + (1-\alpha)\overline{F}^{''} = \alpha\mathbb{E}\left[\varepsilon_{\{\omega|\Theta'(\omega)>t\}}(\omega)\right] + (1-\alpha)\mathbb{E}\left[\varepsilon_{\{\omega|\Theta''(\omega)>t\}}(\omega)\right]$ where

$$\varepsilon_A(\omega) = \begin{cases} 1 & \text{if } \omega \in \text{A} \\ 0 & \text{if } \omega \notin \text{A} \end{cases}$$

and $\mathbb{E}$ stands for mathematical expectation. Define binary random variable η such that $P(\eta = 1) = \alpha$. Then

$$\alpha\overline{F}^{'}(t) + (1-\alpha)\overline{F}^{''} = \mathbb{E}\left[\varepsilon_{\{\omega|\eta(\omega)\Theta'(\omega)+(1-\eta(\omega))\Theta''(\omega)>t\}}(\omega)\right].$$

On the other hand $\eta\Theta^{'} + (1-\eta)\Theta^{''}$ is the efficiency variable associated with $\eta DMU^{'} + (1-\eta)DMU^{''}$. Hence $\eta\Theta^{'} + (1-\eta)\Theta^{''} \in \tilde{\Gamma}$ and therefore $\alpha\overline{F}^{'}(t) + (1-\alpha)\overline{F}^{''} \in \tilde{\mathcal{F}}$. □

Proof of Theorem 6: We first note that if DMU_o is admissible, then $\overline{F}_o \in \partial\tilde{\mathcal{F}}$ where $\partial\tilde{\mathcal{F}}$ denotes the boundary of $\tilde{\mathcal{F}}$. We further note from Lemma 2, $\tilde{\mathcal{F}}$ is convex. Define $\Delta(\overline{F}_o) = \{f \mid f \in \mathcal{C}([0,1])$ and $f(\theta) \geq \overline{F}_o(\theta), \forall\theta\}$ where $\mathcal{C}([0,1])$ is the space of continuous functions on [0,1]. Note that $\left(\Delta(\overline{F}_o) \setminus \{\overline{F}_o\}\right) \bigcap \tilde{\mathcal{F}} = \emptyset$. On the other hand, $\Delta(\overline{F}_o)\setminus\{\overline{F}_o\}$ is a convex body, i.e., $\Delta(\overline{F}_o)\setminus\{\overline{F}_o\}$ has an interior point. The rest of the proof goes along the lines of Asgharian and Noorbaloochi (1998). In fact, a version of the separation theorem provides a continuous linear functional. A form of the Riesz representation theorem gives us a signed measure. Finally, the construction of the convex sets shows that the linear functional is positive and this implies that the corresponding measure in the Riesz representation theorem is a non-negative measure. □

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 (8) simplifies to

$$\begin{aligned} z_{ijt} \mid \mu_{ij}, \sigma^2_{ij} &\overset{\text{iid}}{\sim} \mathcal{N}(\mu_{ij}, \sigma^2_{ij}), \\ \mu_{ij} \mid \tau, \kappa &\overset{\text{iid}}{\sim} \mathcal{N}(\tau, \kappa\sigma^2_{ij}), \\ \sigma^2_{ij} &\overset{\text{iid}}{\sim} \Gamma^{-1}(\frac{a}{2}, \frac{b}{2}), \end{aligned} \tag{14}$$

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 respectively.

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

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

We first calculate

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

After some simple algebra

$$f(\mathbf{z}_{ij} \mid \sigma^2_{ij}) = (2\pi)^{-\frac{T}{2}} |\sigma^2_{ij}\mathbf{V}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(\mathbf{z}_{ij} - \tau\mathbf{1})'(\sigma^2_{ij}\mathbf{V})^{-1}(\mathbf{z}_{ij} - \tau\mathbf{1}) \right\}, \tag{15}$$

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 (15) with respect to the inverse gamma density $f(\sigma^2_{ij})$

$$
\begin{aligned}
f(\mathbf{z}_{ij} \mid \varphi) &= (2\pi)^{-\frac{T}{2}}|\mathbf{V}|^{-\frac{1}{2}}b^{\frac{a}{2}}2^{-\frac{a}{2}}\Gamma(\frac{a}{2})^{-1} \times \\
&\int_0^{+\infty}(\sigma_{ij}^2)^{1-\frac{T}{2}-\frac{a}{2}} \exp\left\{-\frac{(\mathbf{z}_{ij}-\tau\mathbf{1})'\mathbf{V}^{-1}(\mathbf{z}_{ij}-\tau\mathbf{1})}{2\sigma_{ij}^2} - \frac{b}{2\sigma_{ij}^2}\right\} d\sigma_{ij}^2.
\end{aligned}
$$

After changing the variable of integration with respect to $\gamma = \sigma_{ij}^{-2}$

$$
\begin{aligned}
f(\mathbf{z}_{ij} \mid \varphi) &= (2\pi)^{-\frac{T}{2}}|\mathbf{V}|^{-\frac{1}{2}}b^{\frac{a}{2}}2^{-\frac{a}{2}}\Gamma(\frac{a}{2})^{-1} \times \\
&\int_0^{+\infty}\gamma^{\frac{a+T-2}{2}} \exp\left\{-\frac{b+(\mathbf{z}_{ij}-\tau\mathbf{1})'\mathbf{V}^{-1}(\mathbf{z}_{ij}-\tau\mathbf{1})}{2}\gamma\right\} d\gamma.
\end{aligned}
$$

The last integral is the gamma integral and therefore,

$$
f(\mathbf{z}_{ij} \mid \varphi) = \frac{b^{\frac{a}{2}}\Gamma\left(\frac{a+T}{2}\right)}{\pi^{\frac{T}{2}}|\mathbf{V}|^{\frac{1}{2}}\Gamma\left(\frac{a}{2}\right)\left\{b+(\mathbf{z}_{ij}-\tau\mathbf{1})'\mathbf{V}^{-1}(\mathbf{z}_{ij}-\tau\mathbf{1})\right\}^{\frac{a+T}{2}}}. \tag{16}
$$

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

$$
f(\mu, \sigma^2) \propto (\sigma^2)^{-(a+\frac{3}{2})} \exp\left\{-\frac{1}{2\kappa\sigma^2}(\mu-\tau)^2 - \frac{b}{\sigma^2}\right\}. \tag{17}
$$

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

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

After some algebraic simplifications

$$
\begin{aligned}
f(\mu_{ij}, \sigma_{ij}^2 \mid \mathbf{z}_{ij}) &\propto (\sigma_{ij}^2)^{-(\frac{a+T}{2}+\frac{3}{2})} \times \\
&\exp\left\{-\frac{1+\kappa T}{2\kappa\sigma_{ij}^2}\left(\mu - \frac{\tau+\kappa\sum_{t=1}^{T} z_{ijt}}{1+\kappa T}\right)^2 - \frac{1}{2\kappa\sigma_{ij}^2}(\tau^2+\kappa b+\kappa\sum_{t=1}^{T} z_{ijt}^2)\right\}.
\end{aligned} \tag{18}
$$

Comparing (18) with (17), shows that the posterior is in the normal-inverse-gamma form with parameters τ^*, κ^*, a^*, and b^* being

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

References

Asgharian, M. and Noorbaloochi, S. (1998) Note on a fundamental relationship between admissible and Bayesian decision rules. *Statistics* **31**, 21–34.

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.

Carlstein, E., Do, K.-A., Hall, P., Hesterberg, T. and Künsch, H. (1998) Matched-block bootstrap for dependent data. *Bernoulli* **4**, 305–328.

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., Park, K. and Yu, G. (1999) IDEA and AR-IDEA: Models for dealing with imprecise data in DEA. *Management Science* **45**, 597–607.

Cooper, W., Park, K. and Yu, G. (2001) An illustrative application of IDEA (imprecise data envelopment analysis) to a Korean mobile telecommunication company. *Operations Research* **49**, 807–820.

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.

Despotis, D. and Smirlis, Y. (2002) Data envelopment analysis with imprecise data. *European Journal of Operational Research* **140**, 24–36.

Efron, B. and Tibshirani, R. (1993) *An introduction to the bootstrap.* Volume 57. CRC press.

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

Inoue, A. and Kilian, L. (2002) Bootstrapping smooth functions of slope parameters and innovation variances in var(∞) models. *International Economic Review* **43**, 309–331.

Kao, C. (2006) Interval efficiency measures in data envelopment analysis with imprecise data. *European Journal of Operational Research* **174**, 1087–1099.

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.

Kunsch, H.R. (1989) The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics* **17**(3), 1217–1241.

Lamb, J.D. and Kai-Hong, T. (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.

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.

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.