Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models

Léopold Simar • Paul W. Wilson
Institut de Statistique and CORE, Université Catholique de Louvain, Voie du Roman Pays, 20, Louvain-la-Neuve, Belgium
Department of Economics, University of Texas at Austin, Austin, Texas 78712

Efficiency scores of production units are generally measured relative to an estimated production frontier. Nonparametric estimators (DEA, FDH, $\cdots$) are based on a finite sample of observed production units. The bootstrap is one easy way to analyze the sensitivity of efficiency scores relative to the sampling variations of the estimated frontier. The main point in order to validate the bootstrap is to define a reasonable data-generating process in this complex framework and to propose a reasonable estimator of it. This paper provides a general methodology of bootstrapping in nonparametric frontier models. Some adapted methods are illustrated in analyzing the bootstrap sampling variations of input efficiency measures of electricity plants.
(*Data Envelopment Analysis; Bootstrap; Resampling Methods; Frontier Efficiency Models*)

1. Introduction

The idea of estimating production efficiency scores in a nonparametric framework dates to the original work of Farrell (1957). The efficiency of production units is typically measured relative to a production frontier, defined as the geometrical locus of optimal production plans. This frontier may be estimated (nonparametrically) from a set of *n* observed production units.

The Data Envelopment Analysis (DEA) approach develops Farrell's ideas and is based on linear programming techniques. When allowing for variable returns to scale, the frontier is taken as the boundary of the convex hull of the set of observations in input/output space. This approach relies on convexity assumptions of the attainable set of production plans (see, e.g., Charnes et al. 1978 or Färe et al. 1985). The Free Disposal Hull (FDH) method developed by Deprins et al. (1984) extends the idea, allowing nonconvex production sets by relying only on disposability assumptions on inputs and outputs.

Although the literature typically refers to DEA and FDH methods as being deterministic, in both cases efficiency is measured relative to an *estimate* of the true (but unobserved) production frontier. Since statistical estimators of the frontier are obtained from finite samples, the corresponding measures of efficiency are sensitive to the sampling variations of the obtained frontier. Korostelev et al. (1995a, 1995b; hereafter KST) have shown the consistency of FDH and DEA estimators under very weak general conditions but the obtained rates of convergence are, as with many nonparametric estimators, very slow.[1]

The bootstrap introduced by Efron (1979) seems an attractive tool to analyze the sensitivity of measured efficiency scores to sampling variation. Bootstrapping is based on the idea of repeatedly simulating the data-generating process (DGP), usually through resampling, and applying the original estimator to each simulated sample so that resulting estimates mimic the sampling distribution of the original estimator. In principle, this can be done for any statistic (estimator) defined on the

[1] Banker (1993) also demonstrates consistency of the DEA frontier estimator, but does not give the rate of convergence.

0025-1909/98/4401/0049$05.00

data, provided the underlying DGP is properly simulated.[2]

The primary difficulty in applying bootstrap methods in complex situations, such as the case of nonparametric frontier estimation, lies in simulating the DGP.[3] In the case of nonparametric frontier estimation, one must first clearly define a model of the DGP. If the DGP is not specified a priori, we cannot know whether the bootstrap mimics the sampling distribution of the estimators of interest, or some other distribution.

This paper proposes a bootstrap strategy motivated through reasonable assumptions regarding the DGP. Section 2 presents the general framework of frontier models and discusses the bootstrap within the context of this framework. We show how the bootstrap analog of the DGP can approximate the sampling variation of the estimated frontier, allowing us to analyze the sensitivity of the efficiency score of a given production unit. Section 3 briefly reviews some of the nonparametric efficiency estimators that have been proposed in the literature (DEA and FDH), along with their statistical properties. The main results of the paper are presented in §4, where we show a reasonable assumption on the DGP allows one to easily produce a bootstrap version. Special attention is devoted to the smooth bootstrap and its implementation in this framework. The results from this section are applied in §5 to the particular case of nonparametric efficiency estimation. Two algorithms are proposed and compared with other possible approaches. Section 6 provides an empirical illustration of the bootstrap algorithms.

[2] For more details on the bootstrap, see Efron (1982), Hall (1992), or Efron and Tibshirani (1993). Simar (1992) introduced bootstrapping in the production frontier framework for parametric, nonparametric, and semiparametric models using panel data, while Hall et al. (1995) have investigated the consistency of bootstrap distributions in the context of parametric frontier estimation. The approach in Simar (1992) requires that efficiency remain constant over time.

[3] DEA represents a complex situation because we are attempting to nonparametrically estimate the boundary of a typically high-dimensional production set. In addition, few, if any, assumptions are typically made on the DGP in DEA models. See Simar (1996) for additional discussion of this point.

2. Data Generating Process and Bootstrap: The General Setup

Given a list of p inputs ($x \in \mathbb{R}_+^p$) and of q outputs ($y \in \mathbb{R}_+^q$), it is common practice in economic analysis to describe the activity of a productive organization by means of the production set Ψ of physically attainable points (x, y):

$$\Psi = \{(x, y) \in \mathbb{R}_+^{p+q} \mid x \text{ can produce } y\}. \quad (2.1)$$

This set can be described by its sections, either an input requirement set defined $\forall\, y \in \Psi$:

$$X(y) = \{x \in \mathbb{R}_+^p \mid (x, y) \in \Psi\}, \quad (2.2)$$

or an output correspondence set defined $\forall\, x \in \Psi$:

$$Y(x) = \{y \in \mathbb{R}_+^q \mid (x, y) \in \Psi\}. \quad (2.3)$$

The relationships between the two sets, along with standard assumptions one may reasonably make on them, are discussed in §9.1 of Sheppard (1970). Convexity of $X(y)$ for all y (and of $Y(x)$ for all x) and disposability of inputs and outputs are the most usual. The Farrell efficiency boundaries are subsets of $X(y)$ (and $Y(x)$ respectively) denoted by $\partial X(y)$ ($\partial Y(x)$ respectively):

$$\partial X(y) = \{x \mid x \in X(y),\ \theta x \notin X(y) \quad \forall\, 0 < \theta < 1\}, \quad (2.4)$$

$$\partial Y(x) = \{y \mid y \in Y(x),\ \beta y \notin X(y) \quad \forall\, \beta > 1\}. \quad (2.5)$$

These may be used to define the Farrell input and output measures of efficiency (respectively) for a given point (x_k, y_k):

$$\theta_k = \min\{\theta \mid \theta x_k \in X(y_k)\}, \quad (2.6)$$

$$\beta_k = \max\{\beta \mid \beta y_k \in Y(x_k)\}. \quad (2.7)$$

If $\theta_k = 1$ ($\beta_k = 1$), the unit (x_k, y_k) is considered as being "input-efficient" ("output-efficient"). The input efficiency score $\theta_k \le 1$ represents the feasible proportionate reduction of inputs the production unit could realize if y_k were produced efficiently. It will be useful for later development to denote the efficient level of input corresponding to the output level y_k as

$$x^{\partial}(x_k \mid y_k) = \theta_k x_k. \quad (2.8)$$

Note that $x^{\partial}(x_k \mid y_k)$ is the intersection of $\partial X(y_k)$ and the ray θx_k.

Similarly, $\beta_k \ge 1$ gives the feasible proportionate increase in outputs the production unit could realize if the given inputs x_k were used efficiently. Note that both are *radial measures* of the distances between (y_k, x_k) and the corresponding frontier ($\partial X(y_k)$ or $\partial Y(x_k)$). Deprins and Simar (1983) analyze the relations between both measures.

We discuss the bootstrap in terms of input efficiency measures to conserve space. Bootstrapping in the output efficiency case largely involves a straightforward translation of the notation in the following discussion.

Typically, Ψ, $X(y)$, and $\partial X(y)$ are unknown; hence, for a given unit (x_k, y_k), θ_k is also unknown. Suppose that some DGP, $\mathcal{P}$, generates a random sample $\mathcal{X} = \{(x_i, y_i) \mid i = 1, \ldots, n\}$. This sample defines, by some method $\mathcal{M}$, the estimators $\hat{\Psi}$, $\hat{X}(y)$, and $\partial \widehat{X(y)}$. Thus, for a given unit (x_k, y_k), we can estimate its efficiency by

$$\hat{\theta}_k = \min\{\theta \mid \theta x_k \in \widehat{X(y_k)}\}. \tag{2.9}$$

Note that the sampling properties of $\hat{\Psi}$, $\widehat{X(y)}$, $\partial \widehat{X(y)}$, and consequently of $\hat{\theta}_k$ depend on $\mathcal{P}$, which is unknown. Further, they are difficult to determine when $\mathcal{M}$ is complex (as in nonparametric methods).

The bootstrap is perhaps most useful in situations such as ours where the sampling properties of estimators are either difficult or impossible to obtain analytically. Suppose that due to our knowledge of $\mathcal{P}$, we can produce a reasonable estimator $\hat{\mathcal{P}}$ of $\mathcal{P}$ from the data $\mathcal{X}$. Consider now a data set $\mathcal{X}^* = \{(x_i^*, y_i^*), i = 1, \ldots, n\}$ generated by $\hat{\mathcal{P}}$. This pseudo-sample defines, by the same method $\mathcal{M}$, the corresponding quantities $\hat{\Psi}^*$, $\widehat{X^*(y)}$, $\partial \widehat{X^*(y)}$. In particular, for the given unit (x_k, y_k) its measure of efficiency $\hat{\theta}_k^*$ is given by:

$$\hat{\theta}_k^* = \min\{\theta \mid \theta x_k \in \widehat{X^*(y_k)}\}. \tag{2.10}$$

Note that conditionally on $\mathcal{X}$ the sampling distributions of the estimators $\hat{\Psi}^*$, $\hat{X}^*(y)$, and $\partial \widehat{X^*(y)}$ are (in principle) completely known since $\hat{\mathcal{P}}$ is known, although they may be difficult to compute analytically. However, the sampling distributions are easily approximated by Monte Carlo methods. Using $\hat{\mathcal{P}}$ to generate B samples $\mathcal{X}_b^*$, $b = 1, \ldots, B$, and applying $\mathcal{M}$ to each of these pseudo samples yields sets of pseudo estimates $\hat{\Psi}_b^*$, $\hat{X}_b^*(y)$, and $\partial \widehat{X_b^*(y)}$, $b = 1, \ldots, B$. In particular, for a given unit (x_k, y_k), we have $\{\hat{\theta}_{kb}^*\}_{b=1}^B$; the empirical density function of $\{\hat{\theta}_{kb}^*\}_{b=1}^B$ is the Monte Carlo approximation of the distribution of $\hat{\theta}_k^*$ conditional on $\hat{\mathcal{P}}$.

The bootstrap method is based on the idea that if $\hat{\mathcal{P}}$ is a reasonable estimator of $\mathcal{P}$, the known bootstrap distributions will mimic the original unknown sampling distributions of the estimators of interest. More specifically, for the efficiency measure θ_k of a given fixed unit (x_k, y_k) we have:

$$(\hat{\theta}_k^* - \hat{\theta}_k) \mid \hat{\mathcal{P}} \sim (\hat{\theta}_k - \theta_k) \mid \mathcal{P}, \tag{2.11}$$

where θ_k, $\hat{\theta}_k$, and $\hat{\theta}_k^*$ by (2.6), (2.9), and (2.10). To be more explicit, the analogy defined by (2.11) is valid provided $\hat{\mathcal{P}}$ is a consistent estimator of $\mathcal{P}$.[4]

The key expression (2.11) allows us to estimate the bias of $\hat{\theta}_k$, the original estimator of θ_k:

$$\text{bias}_{\mathcal{P},k} = E_{\mathcal{P}}(\hat{\theta}_k) - \theta_k, \tag{2.12}$$

by its bootstrap estimate:

$$\text{bias}_{\hat{\mathcal{P}},k} = E_{\hat{\mathcal{P}}}(\hat{\theta}_k^*) - \hat{\theta}_k. \tag{2.13}$$

The latter quantity may be approximated through the Monte-Carlo realizations $\hat{\theta}_{k,b}^*$:

$$\widehat{\text{bias}}_k = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_{k,b}^* - \hat{\theta}_k = \bar{\theta}_k^* - \hat{\theta}_k. \tag{2.14}$$

Therefore, a bias-corrected estimator of θ_k is[5]:

$$\tilde{\theta}_k = \hat{\theta}_k - \widehat{\text{bias}}_k = 2\hat{\theta}_k - \bar{\theta}_k^*. \tag{2.15}$$

The standard error of $\hat{\theta}_k$ may be estimated by:

$$\widehat{\text{se}} = \left\{ \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}_{k,b}^* - \bar{\theta}_k^*)^2 \right\}^{1/2}. \tag{2.16}$$

Finally, the empirical distribution of $\hat{\theta}_{k,b}^*$, $b = 1, \ldots, B$ provides, after correction for bias, confidence intervals for θ_k. The correction for bias is obtained as follows: we want the corrected empirical density function to be centered on $\tilde{\theta}_k$, the bias corrected estimator of θ_k. Therefore, the empirical density function of $\hat{\theta}_{k,b}^*$ has to be shifted by 2 $\widehat{\text{bias}}_k$ to the left since a correction of 1 $\widehat{\text{bias}}_k$ would center on the biased $\hat{\theta}_k$ rather than $\tilde{\theta}_k$. Hence we now

[4] See Hall (1992).

[5] We say that $\tilde{\theta}_k$ is a bias-corrected estimator, rather than an unbiased estimator, since the adjustment in (2.15) only makes a first-order correction. Our bootstrap procedure is consistent; however, additional work is required to ascertain finite-sample properties of DEA estimators.

consider the empirical density function of $\tilde{\theta}^*_{k,b}$, $b = 1, \ldots, B$, where:

$$\tilde{\theta}^*_{k,b} = \hat{\theta}^*_{k,b} - 2\,\widehat{\text{bias}}_k. \tag{2.17}$$

Then the usual percentile confidence interval for θ_k with intended coverage $(1 - 2\alpha)$ is given by:

$$(\hat{\theta}_{k,low}, \hat{\theta}_{k,up}) = (\tilde{\theta}^{*(\alpha)}_k, \tilde{\theta}^{*(1-\alpha)}_k), \tag{2.18}$$

where $\tilde{\theta}^{*(\alpha)}_k$ indicates the $100\cdot\alpha$th percentile of the empirical density function of $\tilde{\theta}^*_{k,b}$, $b = 1, \ldots, B$.

If the empirical density function of $\tilde{\theta}^*_{k,b}$ is skewed, it may be preferable to center the median of the distribution on $\hat{\theta}_k$. This is achieved through the following median-bias corrected confidence intervals (see Efron 1982):

$$(\hat{\theta}_{k,low}, \hat{\theta}_{k,up}) = (\tilde{\theta}^{*(\alpha_1)}_k, \tilde{\theta}^{*(\alpha_2)}_k), \tag{2.19}$$

where $\alpha_1 = \Phi(2\hat{z}_0 + z^{(\alpha)})$, $\alpha_2 = \Phi(2\hat{z}_0 + z^{(1-\alpha)})$, $\hat{z}_0 = \Phi^{-1}(\#\{\tilde{\theta}^*_{k,b} < \hat{\theta}_k\}/B)$, and Φ is the standard normal cumulative density function and $z^{(\alpha)}$ is the $100\cdot\alpha$th percentile of the standard normal distribution so that $\Phi(z^{(\alpha)}) = \alpha$.

Roughly speaking, $\hat{z}_0$ measures the discrepancy between the median of $\hat{\theta}^*$ and $\hat{\theta}$ in normal units. If there is no bias, $\hat{z}_0 = 0$, then $\alpha_1 = \alpha$ and $\alpha_2 = 1 - \alpha$.

The main question remains: How should $\hat{\mathcal{P}}$ be chosen? Since the answer depends on the estimation method $\mathcal{M}$, the next section briefly discusses typical nonparametric estimators that have appeared in the literature, while §4 addresses the basic issue.

3. Nonparametric Frontier Estimation

The DEA approach is based on Farrell's (1957) ideas, and typically involves measurement of efficiency for a given unit (x_k, y_k) relative to the boundary of the convex hull of $\mathcal{X} = \{(x_i, y_i), i = 1, \ldots, n\}$. More precisely, we have[6]:

$$\hat{\Psi}_{\text{DEA}} = \left\{(x, y) \in \mathbb{R}^{p+q} \,|\, y \le \sum_{i=1}^{n} \gamma_i y_i;\right.$$
$$\left. x \ge \sum_{i=1}^{n} \gamma_i x_i;\ \sum_{i=1}^{n} \gamma_i = 1;\ \gamma_i \ge 0,\ i = 1, \ldots, n\right\}. \tag{3.1}$$

Then the estimated input-requirement set for output level y is

$$\widehat{X(y)} = \{x \in \mathbb{R}^p_+ \,|\, (x, y) \in \hat{\Psi}_{\text{DEA}}\}. \tag{3.2}$$

The corresponding input-efficient boundary for output level y, denoted $\partial\widehat{X(y)}$, consists of the intersection of $\widehat{X(y)}$ and the closure of the compliment of $\hat{\Psi}_{\text{DEA}}$.[7]

Finally, for any given point (x_k, y_k), $\hat{\theta}_k$ is obtained from (2.9) by solving the following linear program:

$$\hat{\theta}_k = \min\left\{\theta \,|\, y_k \le \sum_{i=1}^{n} \gamma_i y_i;\ \theta x_k \ge \sum_{i=1}^{n} \gamma_i x_i;\ \theta > 0;\right.$$
$$\left. \sum_{i=1}^{n} \gamma_i = 1;\ \gamma_i \ge 0,\ i = 1, \ldots, n\right\}. \tag{3.3}$$

In fact, $\hat{\theta}_k$ measures the *radial distance* between the point of interest (x_k, y_k) and $(\hat{x}^{\partial}(x_k \,|\, y_k), y_k)$, where $\hat{x}^{\partial}(x_k \,|\, y_k)$ is the level of the inputs the unit should reach in order to be on the efficient boundary of $\hat{\Psi}_{\text{DEA}}$, with the same level of output y_k and the same proportion of inputs (i.e., moving from x_k to $\hat{x}^{\partial}(x_k \,|\, y_k)$ along the ray θx_k); i.e.,

$$\hat{x}^{\partial}(x_k \,|\, y_k) = \hat{\theta}_k x_k. \tag{3.4}$$

Note that $\hat{\Psi}_{\text{DEA}} \subseteq \Psi$, and so $\partial\widehat{X(y)}$ is an inward-biased estimator of $\partial X(y)$. For the kth observed production unit $(x_k, y_k) \in \hat{\Psi}_{\text{DEA}}$, $\hat{\theta}_k \le 1$ is an upward-biased estimator of θ_k. If we choose an (unobserved) point $(x', y') \in \Psi$ but $\notin \hat{\Psi}_{\text{DEA}}$ (see footnote 7), then $\hat{\theta}' > 1$, confirming the inward bias of $\partial\widehat{X(y)}$. In this case $\hat{\theta}'$ would be interpreted as the proportionate increase in inputs required to move the point (x', y') onto the boundary of $\hat{\Psi}_{\text{DEA}}$, computed from $\mathcal{X}$.

[6] This is the definition of $\hat{\Psi}_{\text{DEA}}$ with varying returns to scale. If the equality constraint in (3.1) is replaced by the inequality $\sum_{i=1}^{n} \gamma_i \le 1$, this adds the origin in the feasible set and implies nondecreasing returns to scale. For constant returns to scale this constraint is suppressed. See Grosskopf (1986).

[7] Note that $\partial\widehat{X(y)}$ is only defined for y such that: $y \le \sum_{i=1}^{n} \gamma_i y_i$; $\sum_{i=1}^{n} \gamma_i = 1$; $\gamma_i \ge 0$, $i = 1, \ldots, n$. In particular, if $q = 1$, it is not defined if $y > \max(y_1, \ldots, y_n)$.

The consistency of the DEA estimator of Ψ has been investigated by KST (1995b) for the case $q = 1$, who show that under very weak general conditions ((x, y) have a strictly positive density on $\partial X(y)$) $\hat{\Psi}_{\text{DEA}}$ is, among the convex sets with monotone boundaries, the maximum likelihood estimator of Ψ.

The convergence rate of an estimator $\hat{\Psi}$ to Ψ depends on the criterion chosen to measure the discrepancy between the two sets. In this general setup, the Lebesgue measure (volume) of the symmetric difference is often chosen:

$$d_{\Delta}(\Psi, \hat{\Psi}) = \text{mes}(\Psi \Delta \hat{\Psi}). \tag{3.5}$$

KST (1995b) prove $E_{\mathcal{P}}(n^{2/p+2} d_{\Delta}(\Psi, \hat{\Psi}_{\text{DEA}}))$ is asymptotically bounded. This means that for large values of n, the discrepancy between Ψ and $\hat{\Psi}_{\text{DEA}}$ is $\mathcal{O}_p(n^{-2/p+2})$. It is further proved that no other estimator in the class of convex sets with monotone boundaries can converge at a faster rate. Although optimal, the achieved convergence rate is very low if p increases.

The FDH estimator proposed by Deprins et al. (1984) provides an alternative nonparametric estimate of Ψ. The FDH estimator relaxes the assumption of convexity of Ψ and may be defined as:

$$\hat{\Psi}_{\text{FDH}} = \{(x, y) \in \mathbb{R}_+^{p+q} \mid y \leq y_i, x \geq x_i, (x_i, y_i) \in \mathcal{X}\}. \tag{3.6}$$

$\hat{\Psi}_{\text{FDH}}$ is the union of all positive orthants in the inputs and of the negative orthants in the outputs whose origin coincides with the observed points $(x_i, y_i) \in \mathcal{X}$. To stress the analogy with the DEA estimator $\hat{\Psi}_{\text{DEA}}$, note that (3.6) may be rewritten as:

$$\hat{\Psi}_{\text{FDH}} = \left\{(x, y) \in \mathbb{R}_+^{p+q} \mid y \leq \sum_{i=1}^{n} \gamma_i y_i;\ x \geq \sum_{i=1}^{n} \gamma_i x_i; \right.$$
$$\left. \sum_{i=1}^{n} \gamma_i = 1;\ \gamma_i \in \{0, 1\}, i = 1, \ldots, n \right\}. \tag{3.7}$$

Therefore, definitions of $\widehat{\partial X(y)}$ and $\hat{\theta}_k$ corresponding to $\hat{\Psi}_{\text{FDH}}$ follow from (3.2) and (3.3) after replacing the constraint $\gamma_i \geq 0$ with $\gamma_i \in \{0, 1\}$.[8]

[8] The definition (3.7) is convenient for stressing the analogy with DEA but is typically not used for computational purposes. Rather the efficiency of a given unit (x_k, y_k) is easily computed by the program: $\hat{\theta}_k = \min\{\theta \mid y_k \leq y_i;\ \theta x_k \geq x_i, (x_i, y_i) \in \mathcal{X}\}$.

The consistency of $\hat{\Psi}_{\text{FDH}}$ has been investigated for the case $q = 1$ by KST (1995a, 1995b). Here, $\hat{\Psi}_{\text{FDH}}$ is, under very weak general conditions, among the sets with monotone boundaries, the maximum likelihood estimator of Ψ (the main condition being again that the density of (x, y) is strictly positive on the frontier). The convergence rate for $\hat{\Psi}_{\text{FDH}}$ is analyzed in KST (1995b) w.r.t. the Lebesgue measure of the symmetric difference between an estimator $\hat{\Psi}$ and Ψ given by (3.5). In this case it is proved that $E_{\mathcal{P}}(n^{1/p+2} d_{\Delta}(\Psi, \hat{\Psi}_{\text{FDH}}))$ is asymptotically bounded. For large n, $d_{\Delta}(\Psi, \hat{\Psi}_{\text{FDH}})$ is $\mathcal{O}_p(n^{-1/p+1})$ and again no other estimator, in the class of nonconvex sets with monotone boundaries, can converge with a faster rate. Note that this convergence rate is lower than for DEA estimators, due to the more general framework (no convexity assumptions).

In KST (1995a), the Hausdorff metric is used to measure the discrepancy between Ψ and an estimator $\hat{\Psi}$:

$$d_H(\Psi, \hat{\Psi}) = \max\{\max_{z \in \Psi} d(z, \hat{\Psi}), \max_{z \in \hat{\Psi}} d(z, \Psi)\}, \tag{3.8}$$

where $d(z, A) = \min_{\omega \in A} |z - \omega|$ is the Euclidean distance between a point z and a set A. It is there shown that

$$E_{\mathcal{P}}\left(\left(\frac{n}{\log n}\right)^{1/p+1} d_H(\Psi, \hat{\Psi}_{\text{FDH}})\right)$$

is asymptotically bounded. This again is a very low rate of convergence when p is large.

4. The Bootstrap

From §2, we know that the key to successful implementation of the bootstrap is to find a reasonable estimate $\hat{\mathcal{P}}$ of the DGP, $\mathcal{P}$.

4.1. General Theory

Since $\mathcal{P}$ generates $\mathcal{X} = \{(x_i, y_i), i = 1, \ldots, n\}$, a naive estimator of $\mathcal{P}$ would be the empirical distribution function defined as the discrete distribution that puts probability $1/n$ on each point (x_i, y_i). Then a bootstrap sample $\mathcal{X}^* = \{(x_i^*, y_i^*), i = 1, \ldots, n\}$ would simply be obtained by randomly sampling with replacement from $\mathcal{X}$ along the lines of the "correlation model" proposed by Freedman (1981) in a regression framework. The strategy is certainly attractive since it is robust with respect to assumptions made on the DGP; however, in the frontier framework, this does not appear to provide a reasonable estimate of the DGP.

First, it does not reflect the Farrell measure of input inefficiency where, for a given value of y, interest lies in radial deviations of input vectors x from $x^{\partial}(x\,|\,y)$. Second, for some realizations of $\mathcal{X}^*$ produced from the empirical distribution, $\widehat{\partial X^*(y)}$ cannot be defined, and so the corresponding $\hat{\theta}_k^*$ also cannot be defined (see footnote 4). In the regression framework, it is often preferable to bootstrap on the "residuals" (see Freedman 1981, Wu 1986, Efron and Tibshirani 1993). Here, the "residuals" are characterized by $\hat{\theta}_i$. Basing the bootstrap on the $\hat{\theta}_i$ will account for the fact that the observed inefficiencies are conditional on the observed outputs as well as the observed frontier.

In fact, the DGP may be described as follows: For a given value of y (the output vector), we know that $x \in X(y)$. Due to the presence of inefficiency, x may not be on $\partial X(y)$ but is generated along a fixed ray (fixed proportion of inputs) away from $x^{\partial}(x\,|\,y)$. Therefore, a particular unit (x_i, y_i) may be considered as being generated, conditionally on y_i and on the observed proportion of inputs by the random variables $\theta_i \in (0, 1]$ such that $x_i = x^{\partial}(x_i\,|\,y_i)/\theta_i$. Suppose that the process generating inefficiencies θ_i is the following:

$$(\theta_1, \ldots, \theta_n) \sim \text{i.i.d. } F, \tag{4.1}$$

where F is a density function on $(0, 1]$.

Then the DGP, $\mathcal{P}_i$ generating x_i conditionally on the observed output values y_i, and on the observed proportion of inputs is completely characterized by $x^{\partial}(x_i\,|\,y_i)$ and F:

$$\mathcal{P}_i = (x^{\partial}(x_i\,|\,y_i), F), \quad i = 1, \ldots, n, \tag{4.2}$$

and finally the whole DGP is $\mathcal{P} = (\mathcal{P}_1, \ldots, \mathcal{P}_n)$.

If $\partial X(y_i)$ and so $x^{\partial}(x_i\,|\,y_i)$ were known, we could calculate $\theta_i = (x^{\partial}(x_i\,|\,y_i))/x_i$ and estimate F by their empirical distribution function. Unfortunately, however, $\partial X(y_i)$ is not known, but we can use $\widehat{\partial X(y_i)}$ and hence $\hat{x}^{\partial}(x_i\,|\,y_i)$ to calculate approximate efficiency scores $\hat{\theta}_i$ (by the methods explained in §3). An obvious estimate of F is then the empirical distribution of the $\hat{\theta}_i$:

$$\hat{F}(t) = \begin{cases} n^{-1} & \text{if } t = \hat{\theta}_i, i = 1, \ldots, n, \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

Defining

$$\hat{\mathcal{P}}_i = (\hat{x}^{\partial}(x_i\,|\,y_i), \hat{F}), \quad i = 1, \ldots, n, \tag{4.4}$$

we can generate pseudo-samples $\mathcal{X}^* = \{x_i^*, y_i\}$ conditionally on y_i and on the observed proportions of inputs of the unit i, for $i = 1, \ldots, n$. First, select at random with replacement θ_i^* $i = 1, \ldots, n$ from $\hat{\theta}_1, \ldots, \hat{\theta}_n$:

$$\theta_1^*, \ldots, \theta_n^* \sim \text{i.i.d. } \hat{F}. \tag{4.5}$$

Then, for $i = 1, \ldots, n$, the bootstrap inputs are given by:

$$x_i^* = \frac{\hat{x}^{\partial}(x_i\,|\,y_i)}{\partial_i^*} = \frac{\hat{\theta}_i}{\theta_i^*}\, x_i. \tag{4.6}$$

The general principles of §2 can now be applied: From this pseudo-sample $\mathcal{X}^*$ we can compute $\hat{\Psi}^*$ and for any fixed point (x_k, y_k), $\widehat{\partial X^*(y_k)}$, $\hat{x}^*(y_k)$, and $\hat{\theta}_k^*$.

In particular, consider a fixed production unit (x_0, y_0). The estimation of its efficiency score depends on the chosen estimator $\hat{\Psi}$ based on the sample $\mathcal{X}$. For instance, if DEA is used we have by (3.3):

$$\hat{\theta}_0 = \min\left\{\theta\,|\,y_0 \le \sum_{i=1}^{n} \gamma_i y_i;\ \theta x_0 \ge \sum_{i=1}^{n} \gamma_i x_i;\ \theta > 0; \right.$$
$$\left. \sum_{i=1}^{n} \gamma_i = 1;\ \gamma_i \ge 0, i = 1, \ldots, n\right\}. \tag{4.7}$$

In order to compute its efficiency score w.r.t. to $\mathcal{X}^*$, we have to define the corresponding estimator $\hat{\Psi}_{\text{DEA}}^*$:

$$\hat{\Psi}_{\text{DEA}}^* = \left\{(x, y) \in \mathbb{R}_+^{p+q}\,|\,y \le \sum_{i=1}^{n} \gamma_i y_i;\ x \ge \sum_{i=1}^{n} \gamma_i x_i^*; \right.$$
$$\left. \sum_{i=1}^{n} \gamma_i = 1;\ \gamma_i \ge 0, i = 1, \ldots, n\right\}. \tag{4.8}$$

Note that $\hat{\Psi}_{\text{DEA}}^* \subseteq \hat{\Psi}_{\text{DEA}}$, which in the bootstrap world mimics the original fact $\hat{\Psi}_{\text{DEA}} \subseteq \Psi$. Then $\hat{\theta}_0^*$ is obtained by solving the linear program

$$\hat{\theta}_0^* = \min\left\{\theta\,|\,y_0 \le \sum_{i=1}^{n} \gamma_i y_i;\ \theta x_0 \ge \sum_{i=1}^{n} \gamma_i x_i^*; \right.$$
$$\left. \theta > 0 \sum_{i=1}^{n} \gamma_i = 1;\ \gamma_i \ge 0, i = 1, \ldots, n\right\}. \tag{4.9}$$

In the bootstrap world (given $\hat{\mathcal{P}}$), $\hat{\theta}_0^*$ may be viewed as an estimator of $\hat{\theta}_0$, in the same way as in the original world (given $\mathcal{P}$), $\hat{\theta}_0$ is an estimator of θ_0. Formally, by (2.11):

$$(\hat{\theta}_0^* - \hat{\theta}_0) \,|\, \hat{\mathcal{P}} \sim (\hat{\theta}_0 - \theta_0) \,|\, \mathcal{P}. \tag{4.10}$$

Hence, the sensitivity analysis of the efficiency $\hat{\theta}_0$ of the production unit (x_0, y_0) can be achieved along the lines of §2 (correction for bias, percentile confidence interval, etc.).

In practical problems, one is typically interested in analyzing the sensitivity of the efficiency scores $\hat{\theta}_i$ of the original units (x_i, y_i), $i = 1, \ldots, n$. This is discussed in §5.

REMARK 1. It should be noted that $\hat{\theta}_0$, and consequently $\hat{\theta}_0^*$, are only defined if $y_0 \le \sum_{i=1}^n \gamma_i y_i$; $\sum_{i=1}^n \gamma_i = 1$, $\gamma_i \ge 0$, $i = 1, \ldots, n$ (see footnote 4). Note also that $\hat{\theta}_0$ (as $\hat{\theta}_0^*$) may be less than, equal to, or greater than 1 (see §3), but if $\hat{\theta}_0 \ge 1$, then $\hat{\theta}_0^* \ge 1$ with probability 1; specifically, we have $\hat{\theta}_0^* \ge \hat{\theta}_0 \ge \theta_0$.

REMARK 2. The DGP $\mathcal{P}: (\mathcal{P}_1, \ldots, \mathcal{P}_n)$ where $\mathcal{P}_i$ is given by (4.2) relies on a very restrictive Hypothesis (4.1). This hypothesis validates the choice (4.5) of generating $\theta_1^*, \ldots, \theta_n^*$. A less restrictive hypothesis would be

$$\theta_i \sim \text{independent } F_i, \,. \tag{4.11}$$

allowing the inefficiency levels to be related to x_i. Unfortunately, F_i cannot be estimated from only one observation of $\hat{\theta}_i$. In this framework the *wild bootstrap* proposed by Härdle and Mammen (1991) does not apply. Of course, if a panel of data $\mathcal{X} = \{(x_{it}, y_{it}); i = 1, \ldots, n; t = 1, \ldots, T\}$ were available, the hypothesis (4.1) could then be replaced by

$$\theta_{i1}, \ldots, \theta_{iT} \sim \text{i.i.d. } F_i. \tag{4.12}$$

Then $\hat{F}_i$ could be obtained from $\hat{\theta}_{i1}, \ldots, \hat{\theta}_{iT}$ and θ_{it}^*, $t = 1, \ldots, T$ could be generated according to:

$$\theta_{i1}^*, \ldots, \theta_{iT}^* \sim \text{i.i.d. } \hat{F}_i. \tag{4.13}$$

This approach would be similar to that described by Simar (1992), but requires the assumption that the level of inefficiency does not vary over time.

REMARK 3. The DEA (or FDH) estimator may produce a large number of ostensibly efficient units with $\hat{\theta}_i = 1$ (the number of such units is likely to increase with p, the number of inputs). Consequently, $\hat{F}$ will provide a poor estimate of F near the upper bound (1) of its support (indeed, it can be shown that near the upper bound, the empirical distribution function is not a consistent estimator of F; see Efron and Tibshirani 1993 for an example). The problem is that F is (typically) by definition continuous on $(0, 1]$, whereas with probability 1, $\hat{F}$ puts a positive mass at $\theta = 1$. Furthermore, it is well known that it is difficult to estimate F from the empirical distribution $\hat{F}$ in the extreme tails when, as is the case here, the support of F is bounded. Note that in the context of frontier efficiency estimation, only the upper bound for θ (namely $\theta = 1$) raises a problem. In particular, bootstrap estimates may be inconsistent if this issue is not addressed.

4.2. The Smoothed Bootstrap

One way to improve the estimation of F and avoid the problem outlined in Remark 3 above is to smooth the empirical $\hat{F}$ (see Silverman and Young 1987). A naive smoothed estimator is provided by a Gaussian kernel density estimate

$$\hat{f}_h(t) = \frac{1}{nh} \sum_{i=1}^{n} \phi\left(\frac{t - \hat{\theta}_i}{h}\right), \tag{4.14}$$

where the smoothing parameter h is fixed and ϕ is the standard normal probability density function.[9] Unfortunately, this kernel estimate does not take into account the boundary condition that $t < 1$, and can be shown to be biased and inconsistent since the support of F is bounded.

The *reflection method* described by Silverman (1986) is a simple tool to overcome this difficulty. Let each point $\hat{\theta}_i \le 1$ be reflected by its symmetric image $2 - \hat{\theta}_i \ge 1$, $i = 1, \ldots, n$, and then estimate the kernel density from the resulting set of $2n$ points using

[9] The smoothing parameter h is sometimes called the bandwidth of the kernel density estimator in (4.14). Larger values of h produce more smoothing in the density estimate, while smaller values of h eventually lead to multiple modes in the estimated density; consequently, the bias of $\hat{f}_h(t)$ increases and the variance decreases with h. For empirical applications, appropriate values of h can be found by maximizing the likelihood cross-validation function as discussed by Silverman (1986). Note that least-squares cross-validation is unlikely to yield meaningful values for h since $\{\hat{\theta}_1, \ldots, \hat{\theta}_n, 2 - \hat{\theta}_1, \ldots, 2 - \hat{\theta}_n\}$ will typically contain numerous elements equal to unity (see Silverman, p. 51–52, for a discussion of this point).

$$\hat{G}_h(t) = \frac{1}{2nh} \sum_{i=1}^{n} \left[\phi\left(\frac{t - \hat{\theta}_i}{h}\right) + \phi\left(\frac{t - 2 + \hat{\theta}_i}{h}\right) \right]. \quad (4.15)$$

Now define

$$\hat{F}_{s,h}(t) = \begin{cases} 2\hat{G}_h(t) & \text{if } t \le 1, \\ 0 & \text{otherwise.} \end{cases} \quad (4.16)$$

It can be proven that $\hat{F}_{s,h}(t)$ is a consistent estimator of F for all $t \le 1$ (see Schuster 1985).

The problem of generating samples $\theta_1^*, \ldots, \theta_n^*$ from $\hat{F}_{s,h}(t)$ is very simple. Let $\beta_1^*, \ldots, \beta_n^*$ be a simple bootstrap sample from $\hat{\theta}_1, \ldots, \hat{\theta}_n$ (obtained by drawing with replacement from $\hat{\theta}_1, \ldots, \hat{\theta}_n$). It is easy to show (by the convolution theorem; Efron and Tibshirani 1993) that

$$t_i = \beta_i^* + h\epsilon_i^* \sim \hat{G}_{1,h}(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} \phi\left(\frac{t - \hat{\theta}_i}{h}\right), \quad (4.17)$$

where ϵ_i^* is a random deviate drawn from the standard normal. Similarly, let t_i^R be the reflection of t_i w.r.t. 1. Then we have

$$t_i^R = 2 - \beta_i^* - h\epsilon_i^* \sim \hat{G}_{2,h}(t)$$
$$= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} \phi\left(\frac{t - 2 + \hat{\theta}_i}{h}\right), \quad (4.18)$$

and $\hat{G}_h(t)$ in (4.15) may be written as:

$$\hat{G}_h(t) = \tfrac{1}{2}\hat{G}_{1,h}(t) + \tfrac{1}{2}\hat{G}_{2,h}(t). \quad (4.19)$$

Now consider the following random generator:

$$\tilde{\theta}_i^* = \begin{cases} \beta_i^* + h\epsilon_i^* & \text{if } \beta_i^* + h\epsilon_i^* \le 1, \\ 2 - \beta_i^* - h\epsilon_i^* & \text{otherwise.} \end{cases} \quad (4.20)$$

Using (4.16)–(4.19), it is straightforward to prove

$$\tilde{\theta}_i^* \sim \hat{F}_{s,h}(t). \quad (4.21)$$

Finally, it may be proved by standard manipulations that the obtained bootstrap random variable $\tilde{\theta}_i^*$ has the following properties:

$$E(\tilde{\theta}_i^* \mid \hat{\theta}_1, \ldots, \hat{\theta}_n) = \hat{\mu}, \quad (4.22)$$

$$V(\tilde{\theta}_i^* \mid \hat{\theta}_1, \ldots, \hat{\theta}_n) = \hat{\sigma}_{\hat{\theta}}^2 + h^2, \quad (4.23)$$

where $\hat{\sigma}_{\hat{\theta}}^2$ is the sample variance of $\hat{\theta}_1, \ldots, \hat{\theta}_n$, i.e.,

$$\hat{\sigma}_{\hat{\theta}}^2 = \frac{1}{n} \sum_{i=1}^{n} (\hat{\theta}_i^2 - \bar{\hat{\theta}})^2, \quad (4.24)$$

and $\hat{\mu}$ is the sample mean of the $\hat{\theta}_1, \ldots, \hat{\theta}_n$. As is typical when kernel estimators are used, the variance of the generated bootstrap sequence must be corrected by computing

$$\theta_i^* = \bar{\beta}^* + \frac{1}{\sqrt{1 + h^2/\hat{\sigma}_{\hat{\theta}}^2}} (\tilde{\theta}_i^* - \bar{\beta}^*), \quad (4.25)$$

where $\bar{\beta}^* = (1/n) \sum_{i=1}^{n} \beta_i^*$.

It may be proved by straightforward manipulation that

$$E(\theta_i^* \mid \hat{\theta}_1, \ldots, \hat{\theta}_n) = \hat{\mu}, \quad \text{and} \quad (4.26)$$

$$V(\theta_i^* \mid \hat{\theta}_1, \ldots, \hat{\theta}_n) = \hat{\sigma}_{\hat{\theta}}^2 \left(1 + \frac{h^2}{n(\hat{\sigma}_{\hat{\theta}}^2 + h^2)}\right). \quad (4.27)$$

Thus, the sequence θ_i^* obtained by the smoothed bootstrap has better properties than $\tilde{\theta}_i^*$ in the sense that the variance of θ_i^* is asymptotically correct. The smoothed bootstrap steps for generating $\theta_1^*, \ldots, \theta_n^*$ from $\hat{\theta}_1, \ldots, \hat{\theta}_n$ are summarized by the following steps:

(1) Generate $\beta_1^*, \ldots, \beta_n^*$ from $\hat{F}$ (drawing with replacement from $\hat{\theta}_1, \ldots, \hat{\theta}_n$).

(2) Define the sequence $\tilde{\theta}_1^*, \ldots, \tilde{\theta}_n^*$ using (4.20).

(3) Define the bootstrap sequence $\theta_1^*, \ldots, \theta_n^*$ using (4.25).

5. Sensitivity Analysis of the Original Efficiency Scores

In the usual application, the researcher is confronted with a set of observations $\mathcal{X} = \{(x_i, y_i) \mid i = 1, \ldots, n\}$ corresponding to n production units. For each of the n observed units, we wish to analyze the sensitivity of the efficiency scores estimated by $\hat{\theta}_1, \ldots, \hat{\theta}_n$. The procedure in §4 may be followed by allowing each observation (x_k, y_k) $k = 1, \ldots, n$ to replace (x_0, y_0) sequentially. This allows us to analyze the sensitivity of the distance from a fixed point (x_k, y_k) to the estimated frontier $\widehat{\partial X(y_k)}$, relative to the sampling variation of the estimator of the frontier, taking into account the entire set of observations $\mathcal{X}$. For the DEA approach, the complete bootstrap algorithm is summarized by the following steps:

(1) For each (x_k, y_k) $k = 1, \ldots, n$ compute $\hat{\theta}_k$ by the linear program (3.3).

(2) Using the smooth bootstrap of §4, generate a random sample of size n from $\hat{\theta}_i$, $i = 1, \ldots, n$ providing $\theta^*_{1b}, \ldots, \theta^*_{nb}$.

(3) Compute $\mathcal{X}^*_b = \{(x^*_{ib}, y_i)\ i = 1, \ldots, n\}$, where $x^*_{ib} = (\hat{\theta}_i / \theta^*_{ib}) x_i$, $i = 1, \ldots, n$.

(4) Compute the bootstrap estimate $\hat{\theta}^*_{k,b}$ of $\hat{\theta}_k$ for $k = 1, \ldots, n$ by solving

$$\hat{\theta}^*_{k,b} = \min\left\{ \theta \,|\, y_k \le \sum_{i=1}^{n} \gamma_i y_i,\ \theta x_k \ge \sum_{i=1}^{n} \gamma_i x^*_{k,b};\ \theta > 0; \sum_{i=1}^{n} \gamma_i = 1;\ \gamma_i \ge 0,\ i = 1, \ldots, n \right\}.$$

(5) Repeat steps 2–4 B times to provide for $k = 1, \ldots, n$ a set of estimates

$$\{\hat{\theta}^*_{k,b},\ b = 1, \ldots, B\}.$$

For large datasets, the choice of B will be constrained by available computer resources. Hall (1986) suggests setting $B = 1000$ to ensure adequate coverage of the confidence intervals.

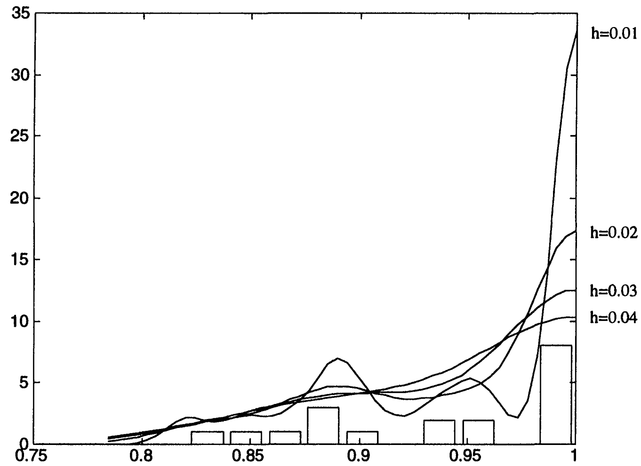
6. Empirical Illustration

To illustrate the methodology proposed in §5, we use data from Färe et al. (1989) on 19 electric utilities operating in 1978. The data contain information on one output (electric power, measured in KWh) and three inputs (labor, measured by average annual employment; fuel; and capital, represented by installed capacity measured in MW).

Figure 1 shows the histogram of $\hat{\theta}_1, \ldots, \hat{\theta}_{19}$, along with its smooth version for selected bandwidths h ($h = 0.07, 0.014, 0.028$). As expected, small values of h give smooth density estimates, which follow the empirical density function and place too much weight near the upper bound 1. Large values of h provide oversmooth density estimates with long tails at the left (below the smallest observed value of $\hat{\theta}$). Maximizing the likelihood cross-validation function described by Silverman (1986) suggests an optimal value of 0.014 for h.

Table 1 shows the results for the bootstrap exercise for $B = 1000$ and $h = 0.014$. Column 1 indicates the firm number, while columns 2–6 give the original DEA efficiency estimate, the bias-corrected estimate, the bootstrap bias estimate, the median of the bootstrapped values, and their standard deviation, respectively. The last four columns provide 95% confidence intervals for the bias-corrected efficiency estimates. The first confidence interval is based on the bias-correction formula in (2.18), while the second confidence interval was computed from the median-centering device represented in (2.19). Since in each case the median of $\tilde{\theta}^*_{k,b}$ is close to $\hat{\theta}_k$, the two sets of confidence intervals are similar.

Figure 1 **Kernel Estimates of F**



The results in Table 1 reveal the sensitivity of the efficiency measures w.r.t. sampling variation. The results indicate that one should be careful in making relative comparisons of the performances among firms based on the original DEA efficiency scores $\hat{\theta}_k$. For example, Firm 1 has a DEA efficiency score $\hat{\theta}_1 = 0.8692$, while Firm 2 is ostensibly efficient with $\hat{\theta} = 1.0$. With the bias-corrected measure in column 3, the difference is less dramatic, but still substantial. However, the last four columns show that the confidence intervals for the efficiency of the two firms overlap to a large degree. Thus, we would not say that the two firms are significantly different in terms of their technical efficiency.

The two sets of confidence intervals based on (2.18) and (2.19) are very similar, with the median-centered intervals in the last two columns of Table 1 shifted slightly to the right relative to the mean centered intervals in columns 7–8. While use of mean centering is probably more common in other bootstrap settings, the median provides a more robust measure of location than the mean when distributions are skewed as with DEA efficiency scores. The present example, however,

Table 1 **Bootstrap with Bandwidth $h = 0.014$**

k	$\hat{\theta}_k$	$\tilde{\theta}_k$	bias_k	Median of $\tilde{\theta}^*_{k,b}$	Std. Dev.	Bias Corrected 2.5%	Bias Corrected 97.5%	Centered on $\tilde{\theta}_k$ 2.5%	Centered on $\tilde{\theta}_k$ 97.5%
1	0.8692	0.8519	0.0173	0.8480	0.0143	0.8360	0.8854	0.8372	0.9057
2	1.0000	0.9307	0.0693	0.9145	0.0614	0.8631	1.0564	0.8646	1.0800
3	1.0000	0.9457	0.0543	0.9396	0.0475	0.8932	1.0574	0.8937	1.0647
4	0.9307	0.9173	0.0133	0.9148	0.0094	0.9059	0.9400	0.9076	0.9536
5	1.0000	0.9349	0.0651	0.9177	0.0573	0.8717	1.0573	0.8730	1.0807
6	0.9071	0.8907	0.0165	0.8849	0.0151	0.8756	0.9286	0.8777	0.9552
7	0.8915	0.8759	0.0156	0.8693	0.0185	0.8615	0.9353	0.8652	0.9861
8	0.8210	0.8076	0.0135	0.8041	0.0118	0.7955	0.8387	0.7967	0.8630
9	0.8892	0.8624	0.0268	0.8488	0.0301	0.8370	0.9434	0.8405	0.9816
10	0.8469	0.8374	0.0095	0.8359	0.0064	0.8294	0.8541	0.8304	0.8598
11	0.9534	0.9423	0.0111	0.9396	0.0104	0.9325	0.9720	0.9339	1.0010
12	1.0000	0.9335	0.0665	0.9153	0.0591	0.8689	1.0484	0.8700	1.0749
13	0.9602	0.9434	0.0168	0.9389	0.0140	0.9282	0.9764	0.9304	1.0049
14	1.0000	0.9258	0.0742	0.9036	0.0676	0.8534	1.0786	0.8545	1.0872
15	1.0000	0.9334	0.0666	0.9085	0.0640	0.8683	1.0786	0.8697	1.1029
16	0.8885	0.8767	0.0117	0.8742	0.0091	0.8664	0.9022	0.8684	0.9169
17	1.0000	0.9378	0.0622	0.9179	0.0561	0.8774	1.0608	0.8783	1.0774
18	1.0000	0.9424	0.0576	0.9325	0.0499	0.8866	1.0527	0.8871	1.0646
19	0.9441	0.9327	0.0113	0.9305	0.0083	0.9228	0.9546	0.9238	0.9645

indicates little practical difference in the two approaches.

A graphical representation of the distribution of $\tilde{\theta}^*_{k,b}$, $b = 1, \ldots, 1{,}000$, $k = 1, \ldots, 19$ is shown in Figure 2, using box plots to facilitate the comparison among firms. For each firm, the box represents the 50% midrange values of $\tilde{\theta}^*_{k,b}$; the length of each box represents the interquantile range (IQR). The whiskers define the natural bounds of the distributions (the mean ±1.5 (IQR)), while the crosses represent outliers lying outside the natural bounds.

Looking at the IQRs, Firms 1, 7, 8, 9, 10, and 16 stand below the others (particularly Firm 8). Among the remaining firms, no firm obviously dominates the others, although the narrow range of the bootstrap values for Firms 4, 11, 13, and 19 should be noted.

Tables 2 and 3 present similar results obtained with different values of the bandwidth h; in Table 2, the bandwidth is reduced by half, while in Table 3 the band width is doubled relative to the value used in Table 1. The results do not appear very sensitive with respect to the different bandwidths, although for $h = 0.007$, more weight is given near the upper bound of θ, while for $h = 0.028$, the distributions are shifted slightly to the left. This is reassuring, since the literature on kernel estimation presents a variety of objective functions that could be optimized to choose the bandwidth h.

Figure 2 **Box Plots of Firm Efficiencies**

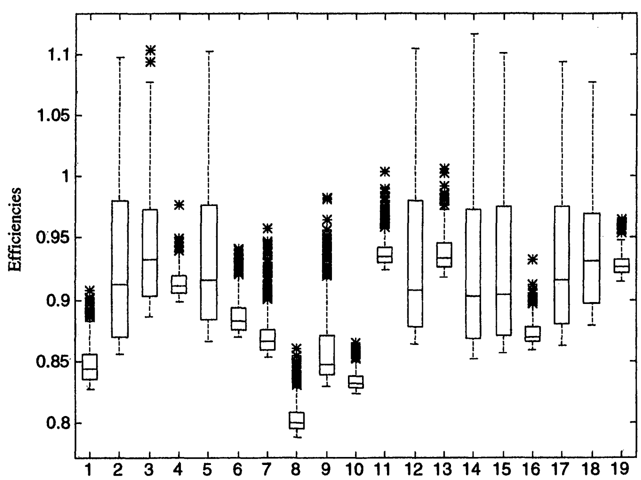


Table 2 **Bootstrap with Bandwidth** $h = 0.007$

k	$\hat{\theta}_k$	$\tilde{\theta}_k$	bias_k	Median of $\tilde{\theta}^*_{k,b}$	Std. Dev.	2.5% Bias Corrected	97.5% Bias Corrected	2.5% Centered on $\tilde{\theta}_k$	97.5% Centered on $\tilde{\theta}_k$
1	0.8692	0.8560	0.0132	0.8498	0.0142	0.8433	0.8909	0.8443	0.9170
2	1.0000	0.9335	0.0665	0.9192	0.0629	0.8677	1.0674	0.8685	1.0856
3	1.0000	0.9489	0.0511	0.9466	0.0492	0.8984	1.0686	0.8986	1.0750
4	0.9307	0.9215	0.0092	0.9182	0.0088	0.9131	0.9448	0.9149	0.9632
5	1.0000	0.9377	0.0623	0.9251	0.0588	0.8762	1.0671	0.8768	1.0862
6	0.9071	0.8945	0.0126	0.8875	0.0148	0.8824	0.9350	0.8840	0.9653
7	0.8915	0.8793	0.0122	0.8717	0.0191	0.8676	0.9432	0.8704	0.9857
8	0.8210	0.8111	0.0099	0.8065	0.0116	0.8018	0.8446	0.8029	0.8826
9	0.8892	0.8657	0.0235	0.8490	0.0312	0.8427	0.9501	0.8444	0.9852
10	0.8469	0.8410	0.0059	0.8394	0.0054	0.8357	0.8565	0.8369	0.8720
11	0.9534	0.9462	0.0072	0.9430	0.0098	0.9394	0.9767	0.9406	1.0302
12	1.0000	0.9365	0.0635	0.9221	0.0603	0.8738	1.0547	0.8743	1.0812
13	0.9602	0.9479	0.0123	0.9426	0.0135	0.9363	0.9821	0.9389	1.0289
14	1.0000	0.9281	0.0719	0.9085	0.0692	0.8570	1.0789	0.8576	1.0837
15	1.0000	0.9360	0.0640	0.9157	0.0654	0.8727	1.0811	0.8734	1.0977
16	0.8885	0.8806	0.0079	0.8778	0.0084	0.8732	0.9059	0.8749	0.9245
17	1.0000	0.9406	0.0594	0.9265	0.0577	0.8818	1.0632	0.8823	1.0874
18	1.0000	0.9461	0.0539	0.9392	0.0507	0.8929	1.0584	0.8931	1.0703
19	0.9441	0.9366	0.0075	0.9342	0.0073	0.9297	0.9574	0.9308	0.9737

Table 3 **Bootstrap with Bandwidth** $h = 0.028$

k	$\hat{\theta}_k$	$\tilde{\theta}_k$	bias_k	Median of $\tilde{\theta}^*_{k,b}$	Std. Dev.	2.5% Bias Corrected	97.5% Bias Corrected	2.5% Centered on $\tilde{\theta}_k$	97.5% Centered on $\tilde{\theta}_k$
1	0.8692	0.8452	0.0240	0.8429	0.0151	0.8248	0.8792	0.8258	0.8856
2	1.0000	0.9255	0.0745	0.9060	0.0594	0.8554	1.0534	0.8586	1.0711
3	1.0000	0.9400	0.0600	0.9330	0.0448	0.8847	1.0469	0.8858	1.0591
4	0.9307	0.9106	0.0201	0.9084	0.0107	0.8953	0.9369	0.8967	0.9415
5	1.0000	0.9298	0.0702	0.9133	0.0551	0.8640	1.0466	0.8670	1.0734
6	0.9071	0.8846	0.0226	0.8805	0.0153	0.8659	0.9219	0.8683	0.9393
7	0.8915	0.8697	0.0219	0.8639	0.0182	0.8512	0.9232	0.8543	0.9563
8	0.8210	0.8015	0.0195	0.7993	0.0125	0.7855	0.8323	0.7867	0.8410
9	0.8892	0.8563	0.0329	0.8465	0.0289	0.8272	0.9330	0.8309	0.9724
10	0.8469	0.8317	0.0152	0.8303	0.0079	0.8204	0.8500	0.8213	0.8547
11	0.9534	0.9357	0.0177	0.9334	0.0115	0.9215	0.9638	0.9228	0.9789
12	1.0000	0.9277	0.0723	0.9114	0.0577	0.8601	1.0458	0.8623	1.0657
13	0.9602	0.9360	0.0242	0.9329	0.0153	0.9158	0.9701	0.9178	0.9794
14	1.0000	0.9213	0.0787	0.9011	0.0660	0.8470	1.0774	0.8499	1.0994
15	1.0000	0.9279	0.0721	0.9053	0.0627	0.8602	1.0767	0.8629	1.1178
16	0.8885	0.8703	0.0181	0.8680	0.0105	0.8562	0.8965	0.8582	0.9052
17	1.0000	0.9328	0.0672	0.9150	0.0538	0.8704	1.0471	0.8722	1.0738
18	1.0000	0.9358	0.0642	0.9250	0.0488	0.8764	1.0468	0.8778	1.0620
19	0.9441	0.9263	0.0177	0.9244	0.0100	0.9126	0.9504	0.9137	0.9550

7. Conclusions

Although nonparametric efficiency measures are often criticized for lacking a statistical basis, we demonstrate that, in fact, nonparametric efficiency measures do have a statistical basis. One of their chief differences from stochastic, parametric models is the implicit nature of the DGP. By focusing on the underlying DGP, we are able to use bootstrap methods to analyze the sensitivity of nonparametric efficiency scores to sampling variation. While it remains true that nonparametric efficiency models do not easily allow for noise in the data, our bootstrapping method addresses at least part of the criticism inflicted on nonparametric frontier models.

All computations in our empirical examples were performed using FORTRAN code written by the authors. Computational times were trivial for the 19 observations in the Färe et al. (1989) data, although this would not be the case for large data sets. In general, the computational burden can be expected to increase both with the number of observations and the numbers of inputs and outputs. We have bootstrapped data sets with up to 322 observations and 8 dimensions in the input/output space, which required 3 hours and 46 minutes elapsed time on a SUN SPARCstation 20. More careful coding would likely reduce the required time, perhaps substantially. Given the continuing decline in the cost of CPU cycles, and the introduction of relatively inexpensive parallel processors, computational burdens for our procedure do not appear critical for most data sets that have been examined in the DEA literature.

The bootstrap estimates offered by our methodology offer several possible enhancements to typical DEA applications. For instance, one can use our methods to test for statistical significance among differences in firms' efficiency scores, much the way t ratios are used in classical regression studies. The bootstrap estimates can also be used to test hypotheses about the structure of the underlying technology, as in Simar and Wilson (1996). Our approach requires only minimal assumptions on the DGP, allowing one to avoid more restrictive assumptions as in Banker (1996). Finally, our approach can be used with cross-sectional data, unlike other nonparametric approaches that require panel data (e.g., Kneip and Simar 1996).[10]

[10] We have benefited from comments by and discussions with G. Ferrier, I. Gijbels, and M. Stinchcombe. We alone are, of course, responsible for any remaining errors or omissions. Research support to the first author from the contract "Projet d'Actions de Recherche Concertées" (PARC No. 93/98-164) of the Belgian Government is gratefully acknowledged. Research support to the second author from the Management Science Group, U.S. Department of Veterans Affairs, is also gratefully acknowledged.

References

Banker, R., "Maximum Likelihood, Consistency, and Data Envelopment Analysis: A Statistical Foundation," *Management Sci.*, 39 (1993), 1265–1273.

——, "Hypothesis Tests Using Data Envelopment Analysis," *J. Productivity Analysis*, 7 (1996), 139–159.

Charnes, A., W. W. Cooper, and E. Rhodes, "Measuring the Inefficiency of Decision Making Units," *European J. Oper. Res.*, 2 (1978), 429–444.

Deprins, D. and L. Simar, "On Farrell Measures of Technical Efficiency," *Recherches Economiques de Louvain*, 49, 2 (1983), 123–137.

——, ——, and H. Tulkens, "Measuring Labor Inefficiency in Post Offices," in M. Marchand, P. Pestieau and H. Tulkens (Eds.), *The Performance of Public Enterprises: Concepts and Measurements*, North-Holland, Amsterdam, 1984, 243–267.

Efron, B., "Bootstrap Methods: Another Look at the Jackknife," *Ann. Statistics*, 7 (1979), 1–26.

——, "The Jackknife, the Bootstrap and Other Resampling Plans 38," *CBMS-NSF Regional Conf. Series in Applied Math.*, SIAM, Philadelphia, PA.

—— and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, London, 1993.

Färe, R., S. Grosskopf, and E. C. Kokkelenberg, "Measuring Plant Capacity, Utilization and Technical Change: A Nonparametric Approach," *International Economic Rev.*, 30 (1989), 655–666.

——, ——, and C. A. K. Lovell, *The Measurement of Efficiency of Production*, Kluwer-Nijhoff Publishing, Boston, MA, 1985.

Farrell, M. J., "The Measurement of Productive Efficiency," *J. Royal Statistical Society*, A(120) (1957), 253–281.

Freedman, D. A., "Bootstrapping Regression Models," *Ann. Statistics*, 9, 6 (1981), 1218–1228.

Grosskopf, S., "The Role of the Reference Technology in Measuring Productive Efficiency," *Economic J.*, 96 (1986), 499–513.

Hall, P., "On the Number of Bootstrap Simulations Required to Construct a Confidence Interval," *Ann. Statistics*, 14 (1986), 1453–1462.

——, *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York, 1992.

——, W. Härdle, and L. Simar, "Iterated Bootstrap with Application to Frontier Models," *J. Productivity Analysis*, 6 (1995), 63–76.

Härdle, W. and E. Mammen, "Bootstrap Methods in Nonparametric Regression," in G. Roussas (Ed.), *Nonparametric Functional Estimation and Related Topics*, Kluwer Academic Publishers, Boston, MA, 1991, 111–123.

Kneip, A. and L. Simar, "A General Framework for Frontier Estimation with Panel Data," *J. Productivity Analysis*, 7 (1996), 187–212.

Korostelev, A., L. Simar, and A. Tsybakov, "Efficient Estimation of Monotone Boundaries," *Ann. Statistics*, 23 (1995a), 476–489.

——, ——, and ——, "On Estimation of Monotone and Convex Boundaries," *Publications de l'ISUP*, 39 (1995b), 3–18.

Lovell, C. A. K., L. C. Walters, and L. L. Wood, "Stratified Models of Education Production Using DEA and Regression Analysis," in *Data Envelopment Analysis: Theory, Methods, and Applications*, A. Charnes, W. W. Cooper, A. Y. Lewin, and L. M. Seiford (Eds.), Quorom Books, New York, 1993.

Sheppard, R. W., *Theory of Cost and Production Function*, Princeton University Press, Princeton, NJ, 1970.

Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.

—— and G. A. Young, "The Bootstrap: Smooth or Not to Smooth?" *Biometrika*, 74 (1987), 469–479.

Simar, L., "Estimating Efficiencies from Frontier Models with Panel Data: A Comparison of Parametric, Non-parametric and Semi-parametric Methods with Bootstrapping," *J. Productivity Analysis*, 3 (1992), 167–203.

——, "Aspects of Statistical Analysis in DEA-Type Frontier Models," *J. Productivity Analysis*, 7 (1996), 177–185.

—— and P. W. Wilson, "Nonparametric Tests of Returns to Scale," Unpublished Manuscript, Department of Economics, University of Texas, Austin, TX, 1996.

Schuster, E. F., "Incorporating Support Constraints into Nonparametric Estimators of Densities," *Communications in Statistics—Theory and Methods*, 14 (1985), 1123–1136.

Wu, C. F. J., "Jackknife, Bootstrap, and Other Resampling Methods in Regression Analysis," *Ann. Statistics*, 14 (1986), 1261–1295.

Accepted by Richard M. Burton; received July 1995. This paper has been with the authors 6 months for 1 revision.