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

by

Léopold Simar** **and Paul W. Wilson*****

April 1995

Abstract

Efficiency scores of production units are generally measured relative to an estimated production frontier. Nonparametric estimators (DEA, FDH, ...) 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 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.

* 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.

** Institut de Statistique and CORE, Université Catholique de Louvain, Voie du Roman Pays, 34, Louvain-la-Neuve, Belgium.

Research support from the contract "Projet d'Actions de Recherche Concertées" (PARC No. 93/98-164) of the Belgian Government is gratefully acknowledged.

*** Department of Economics, University of Texas, Austin, Texas 78712, USA. Research support from the Management Science Group, US Department of Veterans Affairs, is gratefully acknowledged.

1 Introduction

The idea of estimating production efficiency scores in a nonparametric setup dates back to the original work of Farrell (1957). The efficiency of production units is generally measured relatively to a production frontier, defined as the geometrical locus of optimal productions. This frontier may be estimated (nonparametrically) from a set of n observed production units.

The DEA (Data Envelopment Analysis) approach develops Farrell's ideas and is based on linear programming techniques: the frontier is a boundary of the convex hull of the set of observed points. It relies on convexity assumptions of the attainable set of productions (see e.g.: Charnes, Cooper and Rhodes (1978) or Färe, Grosskopf and Lovell (1985). The FDH (Free Disposal Hull) initiated by Deprins, Simar and Tulkens (1984) extends the idea, allowing non convex production sets: the attainable set is the set of minimal volume containing all the observations. It relies only on disposability assumptions on inputs and outputs.

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, Simar and Tsybakov (1992) and (1995) have shown the consistency of FDH and DEA estimators under very weak general conditions but the obtained rates of convergence are, as in many nonparametric estimators, very slow.

This is why the bootstrap methodology seems an attractive tool to analyze the sensitivity of measured efficiency scores to the sampling variations. It was introduced in the frontier framework by Simar (1992), for parametric, nonparametric and semiparametric models in the presence of panel data; Hall, Härdle and Simar (1991) have investigated the consistency of the bootstrap distributions in the context of parametric frontier estimation.

The bootstrap, introduced by Efron (1979), is based on the idea of repeatedly simulating the data generating process (DGP) (usually through resampling) and applying the original estimator to mimic the sampling distribution of the original estimator. In principle, this can be done for any statistic (estimator) defined on the data, provided the underlying DGP is properly simulated. (For more details on the bootstrap, see Efron, 1982; Hall, 1992; or Efron and Tibshirani, 1993.)

The primary difficulty in applying bootstrap methods in complex situations, such as the case of nonparametric frontier estimation, lies in simulating the DGP. In the case of nonparametric frontier estimation, one must first clearly define what the DGP is.

The goal of this paper is to propose a bootstrap strategy that can be motivated through reasonable assumptions regarding 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. 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 section 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 section 5 to the particular case of nonparametric efficiency estimation. Two algorithms are proposed any compared with other possible approaches. Section 6 provides an empirical illustration of the bootstrap algorithms.

2 Data Generating Process and Bootstrap: The General Setup

Given a list of p inputs ($x \in I\!R_+^p$) and of q outputs ($y \in I\!R_+^q$), 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).

$$\Psi = \{(x, y) \in I\!R_+^{p+q} \mid x \text{ can produce } y\} \tag{2.1}$$

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

$$X(y) = \{x \in I\!R_+^p \mid (x, y) \in \Psi\} \tag{2.2}$$

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

$$Y(x) = \{y \in I\!R_+^q \mid (x, y) \in \Psi\}. \tag{2.3}$$

The relations between the two sets, along with standard assumptions one may reasonably make on them, are discussed in section 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 $\partial X(y)$ (resp. $\partial Y(x)$):

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

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

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

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

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

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 $\theta_k \leq 1$ 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 by $x^{\partial}(y_k)$ the efficient level of input corresponding to the output level y_k

$$x^{\partial}(y_k) = \theta_k x_k \tag{2.8}$$

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

Similarly, $\beta_k \geq 1$ 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 $\partial Y(x_k)$). Simar and Deprins (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 $\mathcal{X} = \{(x_i, y_i) | i = 1, \ldots, n\}$. This sample defines, by some method $\mathcal{M}$, the estimators $\widehat{\Psi}$, $\widehat{X}(y)$, and $\widehat{\partial X(y)}$. 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 $\widehat{\Psi}$, $\widehat{X(y)}$, $\widehat{\partial X(y)}$, 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 $\mathcal{X}^* = \{(x_i^*, y_i^*), \quad i = 1, \ldots, n\}$ generated by $\hat{\mathcal{P}}$. This pseudo-sample defines, by the same method $\mathcal{M}$, the corresponding quantities $\widehat{\Psi}^*, \widehat{X^*(y)}, \widehat{\partial X^*(y)}$. 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 $\widehat{\Psi}^*$, $\widehat{X}^*(y)$, and $\widehat{\partial X^*(y)}$ 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 $\widehat{\Psi}_b^*$, $\widehat{X}_b^*(y)$, and $\widehat{\partial X_b^*(y)}$, $b = 1, \ldots, B$. 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 key to the bootstrap method is that if $\hat{\mathcal{P}}$ is a reasonable estimator of $\mathcal{P}$, the known bootstrap distributions 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 $\hat{\theta}_k^*$ are defined by (2.6), (2.9) and (2.10). To be more explicit, analogy defined by (2.11) is valid provided $\hat{\mathcal{P}}$ is a consistent estimator of $\mathcal{P}$ (see **e.g.** Hall (1992)).

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 $\hat{\theta}_{k,b}^*$:

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

Therefore, a bias-corrected estimator of θ_k is:

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

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

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

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 d.f. to be centered on $\tilde{\theta}_k$, the bias corrected estimator of θ_k. Therefore, the empirical d.f. 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 consider now the empirical d.f. 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:

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

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

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

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

where

$\alpha_1 = \Phi(2\hat{z}_0 + z^{(\alpha)})$
$\alpha_2 = \Phi(2\hat{z}_0 + z^{(1-\alpha)})$
$\hat{z}_0 = \Phi^{-1}\left(\frac{\#(\tilde{\theta}^*_{k,b} < \tilde{\theta}_k)}{B}\right)$

Φ is the standard normal cumulative d.f. and $\Phi(z^{(\alpha)}) = \alpha$.

Roughly speaking, $\hat{z}_0$ measures the discrepancy between the median of $\tilde{\theta}^*$ and $\tilde{\theta}$ 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 presents briefly the usual nonparametric estimators proposed in the literature, then section 4 will address the basic issue.

3 Nonparametric Frontier Estimation

The DEA approach is based on Farrell's (1957) ideas. It relies on the assumption of the convexity of Ψ (Charnes, Cooper and Rhodes (1978)). It 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*:

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

Then we obtain from (2.4) the **input-efficient boundary** for the output level y †

$$\widehat{\partial X(y)} = \{x | y \leq \sum_{i=1}^n \gamma_i y_i; \theta x < \sum_{i=1}^n \gamma_i x_i; \theta < 1; \sum_{i=1}^n \gamma_i = 1; \gamma_i \geq 0, i = 1, \ldots, n\}. \tag{3.2}$$

*This is the definition of $\hat{\Psi}_{DEA}$ with varying returns to scale. If the equality constraint in (3.1) is replaced by the inequality $\sum_{i=1}^n \gamma_i \leq 1$, this add the origin in the feasible set and implies decreasing returns to scale. For constant returns to scale this constraint is suppressed.

†Note that $\widehat{\partial X(y)}$ is only defined for y such that:

$$y \leq \sum_{i=1}^n \gamma_i y_i; \sum_{i=1}^n \gamma_i = 1; \gamma_i \geq 0 \quad i = 1, \ldots, n.$$

In particular, if $q = 1$, it is not defined if $y > \max(y_1, \ldots, y_n)$.

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

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

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

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

Note that $\hat{\Psi}_{DEA} \subseteq \Psi$, and so $\widehat{\partial X(y)}$ is an downward-biased estimator of $\partial X(y)$. For the k^{th} observed production unit $(x_k, y_k) \in \hat{\Psi}_{DEA}$, $\hat{\theta}_k \leq 1$ is an upward-biased estimator of θ_k. If we choose an (unobserved) point $(x', y') \in \Psi$ but $\notin \hat{\Psi}_{DEA}$ (see footnote 2), then $\hat{\theta}' > 1$, confirming the downward bias of $\widehat{\partial X(y)}$. 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 $\hat{\Psi}_{DEA}$, computed from $\mathcal{X}$.

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

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

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

It is proved in KST (1995) that $E_{\mathcal{P}}(n^{2/p+2} d_{\Delta}(\Psi, \hat{\Psi}_{DEA}))$ is asymptotically bounded. This means that for large values of n, the discrepancy between Ψ and $\hat{\Psi}_{DEA}$ is $\mathcal{O}_p(n^{-\frac{2}{p+2}})$. It is further proved that **no other estimator**, in the class of convex sets with monotone boundaries **can converge with 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:

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

$\hat{\Psi}_{FDH}$ 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 $\hat{\Psi}_{DEA}$, note that (3.6) may be rewritten

as:

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

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

The consistency of $\hat{\Psi}_{FDH}$ has been investigated for the case $q = 1$ by KST (1992) and (1995). Here, $\hat{\Psi}_{FDH}$ 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 $\hat{\Psi}_{FDH}$ is analyzed in KST (1995) 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 $E_{\mathcal{P}}(n^{\frac{1}{p+2}} d_\Delta(\Psi, \hat{\Psi}_{FDH}))$ is asymptotically bounded. For large n, $d_\Delta(\Psi, \hat{\Psi}_{FDH})$ is $\mathcal{O}_p(n^{-1/p+1})$ and again **no other estimator**, in the class of 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 (1992), the Hausdorff metric is used to appreciate 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 $d(z, A) = \min_{w \in A} |z - \omega|$ is the Euclidean distance between a point z and a set A. It is there shown that $E_{\mathcal{P}}\left(\left(\frac{n}{\log n}\right)^{1/p+1} d_H(\Psi, \hat{\Psi}_{FDH}) \right)$ is asymptotically bounded. This again is a very low rate of convergence when p is large.

4 The Bootstrap

From section 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

[‡]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:

$$\hat{\theta}_k = \min\{\theta \mid y_k \leq y_i; \theta x_k \geq x_i, (x_i, y_i) \in \mathcal{X}\}.$$

probability $\frac{1}{n}$ 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 in $\mathcal{X}$. The strategy is certainly attractive since it is robust w.r.t. assumptions made on the DGP. This is in the line of the "correlation model" proposed by Freedman (1981) in regression frameworks. However, in the frontier framework, this does not appear to provide a reasonable estimate of the DGP.

First, it does not reflect Farrell's view of input inefficiency resulting from, for a given value of y, corresponding values of x which deviate from $x^{\partial}(y)$ in a radial direction. 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 2). In the regression framework, it is often preferable to bootstrap on the "residuals" (see Freedman (1981), Wu (1986), and 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 $x^{\partial}(y)$. 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 $x_i = x^{\partial}(y_i)/\theta_i$. Suppose that the process generating inefficiencies θ_i is the following:

$$(\theta_1, \ldots, \theta_n) \sim \text{ iid } F, \tag{4.1}$$

where F is a d.f. 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 $x^{\partial}(y_i)$ and F:

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

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

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

$$\hat{F}\text{: puts a probability } \frac{1}{n} \text{ on } \hat{\theta}_i \quad i = 1, \ldots, n. \tag{4.3}$$

Now defining

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

we know how to 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 $i = 1, \ldots, n$.

To generate $\mathcal{X}^*$, we first select at random with replacement θ_i^* $i = 1, \ldots, n$ from $\hat{\theta}_1, \ldots, \hat{\theta}_n$:

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

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

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

The general principles of section 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), $\partial \widehat{X^*(y_k)}$, $\hat{x}^*(y_k)$, 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):

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

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

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

Note that $\hat{\Psi}^*_{DEA} \subseteq \hat{\Psi}_{DEA}$, which in the bootstrap world mimics the original fact $\hat{\Psi}_{DEA} \subseteq \Psi$. Then $\hat{\theta}_0^*$ is obtained by solving the linear program (3.3):

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

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):

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

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 section 2 (correction for bias, percentile confidence interval, ...)

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 section 5.

Remark (1): It should be noted that $\hat{\theta}_0$, and consequently $\hat{\theta}_0^*$, are only defined if $y_0 \leq \sum_{i=1}^n \gamma_i y_i$; $\sum_{i=1}^n \gamma_i = 1$, $\gamma_i \geq 0$, $i = 1, \ldots, n$ (see footnote 2). Note also that $\hat{\theta}_0$

(as $\hat{\theta}_0^*$) may be less, equal or larger than one (see section 3), but if $\hat{\theta}_0 \geq 1$, then $\hat{\theta}_0^* \geq 1$ with probability one.

Remark (2): Other DGP's

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 $\hat{F}_i$ is not available from one single 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

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

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:

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

This is the spirit of Simar (1992), and, as observed there, is certainly a more comfortable scenario in which to infer the efficiency of a given unit, provided one is willing to assume 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 one, $\hat{F}$ 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

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

where the smoothing parameter h is fixed and ϕ is the standard normal p.d.f. Unfortunately, this kernel estimate does not take into account the boundary condition that $t < 1$. (The density estimated by (4.14) can be shown to be inconsistent and asymptotically biased when the support of F is bounded). **The reflection method** described by Silverman (1986) is a simple tool to overcome this difficulty.

Consider that each point $\hat{\theta}_i \leq 1$ is reflected by its symmetric image $2-\hat{\theta}_i \geq 1$, $i = 1, \ldots, n$ and then estimate the kernel density from this set of $2n$ points. From (4.14),

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

Now define

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

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

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 formula; Efron and Tibshirani (1993)) that

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

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

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

Note that $\hat{G}_h(t)$ given by (4.15) may be written as:

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

Consider now the following random generator:

$$\tilde{\theta}_i^* = \begin{cases} \beta_i^* + h\varepsilon_i^* & \text{if } \beta_i^* + h\varepsilon_i^* \leq 1 \\ 2 - \beta_i^* - h\varepsilon_i^* & \text{otherwise.} \end{cases} \tag{4.20}$$

It is straightforward to prove using (4.16)-(4.19) that:

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

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

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

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

where $\widehat{\sigma}_{\hat{\theta}}^2$ is the "plug in" estimator of the variance of $\hat{\theta}_1, \ldots, \hat{\theta}_n$, i.e.,

$$\widehat{\sigma}_{\hat{\theta}}^2 = \frac{1}{n} \sum_{i=1}^{n} (\hat{\theta}_i^2 - \widehat{\overline{\theta}})^2. \tag{4.24}$$

As is typical when kernel estimators are used, the variance of the generated bootstrap sequence must be corrected by computing

$$\theta_i^* = \overline{\beta}^* + \frac{1}{\sqrt{1 + \dfrac{h^2}{\widehat{\sigma}_{\hat{\theta}}^2}}} (\tilde{\theta}_i^* - \overline{\beta}^*), \tag{4.25}$$

where $\overline{\beta}^* = \dfrac{1}{n} \sum_{i=1}^{n} \beta_i^*$.

It may be proven by straightforward manipulation that

$$E(\theta_i^* \mid \hat{\theta}_1, \ldots, \hat{\theta}_n) = \widehat{\overline{\theta}} \tag{4.26}$$

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

Thus, the sequence θ_i^* obtained by the smoothed bootstrap has better properties than $\tilde{\theta}_i^*$. 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 $\hat{\theta}_1, \ldots, \hat{\theta}_n$. We consider two approaches, which differ primarily in how the reference set is constructed: the Leave-One-Out (LOO) Bootstrap, and the Complete Bootstrap.

5.1 The Leave-One-Out Bootstrap

Given a set of n observations $\mathcal{X}$ as described above, consider the problem of analyzing the sensitivity of the measured efficiency score for the k^{th} observation (x_k, y_k). The LOO bootstrap involves using the reference set

$$\mathcal{X}_{(k)} = \{(x_i, y_i) \mid i = 1, \ldots, k-1, k+1, \ldots, n\}, \tag{5.1}$$

rather than $\mathcal{X}$ to compute the efficiency score for the k unit.

This approach has been used by Andersen and Petersen (1989) and by Lovell *et al.* (1993) to allow a complete ranking of n production units. Charnes *et al.* (1986, 1991) and Thrall (1993) have used this approach to identify separate classes of efficient observations. Wilson (1995) uses this approach to identify influential observations.

We may apply the bootstrap technique of section 4 by replacing (x_0, y_0) with (x_k, y_k) and replacing $\mathcal{X}$ with $\mathcal{X}_{(k)}$. This allows us to measure the sensitivity of the distance of the fixed point (x_k, y_k) to the estimated frontier $\widehat{\partial X_{(k)}}(y_k)$, relative to the sampling variation of the estimator evaluated from the $(n-1)$ remaining points $\mathcal{X}_{(k)}$.
The DGP, $\mathcal{P}_{(k)} = (\mathcal{P}_{1,(k)}, \ldots, \mathcal{P}_{k-1,(k)}, \mathcal{P}_{k+1,(k)}, \ldots, \mathcal{P}_{n,(k)})$, is in fact estimated (as in (4.4)) by

$$\hat{\mathcal{P}}_{j,(k)} = (\hat{x}^{\partial}_{(k)}(y_j), \hat{F}_{(k)}), \quad j = 1, \ldots, n; j \neq k, \tag{5.2}$$

where $\hat{x}^{\partial}_{(j)}(y_k)$ and $\hat{F}_{(k)}$ are defined as in section 4 with $\mathcal{X}$ replaced by the leave-one-out set $\mathcal{X}_{(k)}$. The key relation (2.10) here becomes:

$$(\hat{\theta}^*_k - \hat{\theta}_{k,(k)}) \mid \hat{\mathcal{P}}_{(k)} \sim (\hat{\theta}_{k,(k)} - \theta_k) \mid \mathcal{P}_{(k)}, \tag{5.3}$$

where $\hat{\theta}_{k,(k)}$ is the initial estimate of θ_k from the data set $\mathcal{X}_{(k)}$:

$$\hat{\theta}_{k,(k)} = \min\{\theta \mid \theta x_k \in \widehat{X_{(k)}(y_k)}\}. \tag{5.4}$$

The bootstrap reference sets are thus $\mathcal{X}^*_{(k)}$, generated as $\mathcal{X}^*$, by drawing with replacement in $\hat{\theta}_{j,(k)}$, $j = 1, \ldots, n$, $j \neq k$. This provides in the Monte-Carlo loop, $\hat{\theta}^*_{k,b}$ for $b = 1, \ldots, B$. For the DEA approach, the LOO bootstrap algorithm is summarized by the following steps:

[1] Select a production unit k in $\mathcal{X}$

[2] Define $\mathcal{X}_{(k)}$ and compute $\hat{\theta}_{j,(k)}$, $j = 1, \ldots, n$, by solving

$$\hat{\theta}_{j,(k)} = \min\{\theta \quad | \quad y_j \leq \sum_{\substack{i=1 \\ i \neq k}}^{n} \gamma_i y_i; \theta x_j \geq \sum_{\substack{i=1 \\ i \neq k}}^{n} \gamma_i x_i; \theta > 0;$$
$$\sum_{\substack{i=1 \\ i \neq k}}^{n} \gamma_i = 1; \gamma_i \geq 0, i = 1, \ldots, n, i \neq k\}$$

[3] Define the empirical d.f. $\hat{F}_{(k)}$ putting mass $\dfrac{1}{n-1}$ on $\hat{\theta}_{j,(k)},\ j = 1,\ldots,n;$ $j \neq k$.

[4] Generate a random sample of size $(n-1)$ from a smooth version of $\hat{F}_{(k)}$: $\theta^*_{j,(k),b};\ j = 1,\ldots,n;\ j \neq k$.

[5] Compute $\mathcal{X}^*_{(k),b} = \{(x^*_{jb}, y_j); j = 1,\ldots,n;\ \ j \neq k\}$

where $x^*_{jb} = \dfrac{\hat{\theta}_{j,(k)}}{\theta^*_{j,(k),b} x_j}$

[6] Compute the bootstrap estimate: $\hat{\theta}^*_{k,b}$

$$\hat{\theta}^*_{k,b} = \min\{\theta \quad | \quad y_k \leq \sum_{\substack{i=1 \\ i\neq k}}^{n} \gamma_i y_i; \theta x_k \geq \sum_{\substack{i=1 \\ i\neq k}}^{n} \gamma_i x^*_{i,b}; \theta > 0;$$
$$\sum_{\substack{i=1 \\ i\neq k}}^{n} \gamma_i = 1; \gamma_i \geq 0, i = 1,\ldots,n, i \neq k\}$$

[7] Repeat [4] - [6] B times, providing

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

[8] Repeat [1] - [7] for $k = 1,\ldots,n$.

This procedure has two important drawbacks. First, by omitting the k^{th} observation from $\mathcal{X}$ and estimating the reference set Ψ by $\hat{\Psi}_{(k)}$, important information about the frontier $\partial X(y_k)$ is lost. In cases where $\hat{\theta}_k = 1$, the estimator $\hat{\theta}_{k,(k)}$ will be biased upward even more than the corresponding estimator $\hat{\theta}_k$. More generally, for finite samples, $\hat{\mathcal{P}}_{(k)}$ gives a less precise estimate of the DGP $\mathcal{P}$ than does $\hat{\mathcal{P}}$; this problem becomes more severe as n becomes smaller. Secondly, for at least one $k \in \{1,\ldots,n\}$, $\hat{\theta}_{k,(k)}$ and therefore $\hat{\theta}^*_{k,b}$ will not be defined due to infeasible constraints in the linear programs in steps [2] and [6] above (see footnote 2).

Both of these drawbacks are avoided in the Complete Bootstrap.

5.2 The Complete Bootstrap

There is no need to delete observations from the reference set in order to examine the sensitivity of efficiency scores. The procedure in section 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] Define the empirical d.f. $\hat{F}$ putting mass $\frac{1}{n}$ on $\hat{\theta}_i$, $i = 1, \ldots, n$

[3] Generate a random sample of size n from a smoothed version of $\hat{F}$:

$$\theta^*_{1b}, \ldots, \theta^*_{nb}$$

[4] Compute $\mathcal{X}^*_b = \{(x^*_{ib}, y_i) \;\; i = 1, \ldots, n\}$
where $x^*_{ib} = \frac{\hat{\theta}_i}{\theta^*_{ib}} x_i$, $i = 1, \ldots, n$

[5] Compute the bootstrap estimate of $\hat{\theta}_k : \hat{\theta}^*_{k,b}$ for $k = 1, \ldots, n$, by solving

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

[6] Repeat [3] - [5] B times to provide for $k = 1, \ldots, n$ a set of estimates

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

6 Empirical Illustration

To illustrate the two approaches proposed in section 5, we use data from Färe, Grosskopf, and Kokkelenberg (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 ($h = 0.01,\ 0.02,\ 0.03,\ 0.04$). As expected, small values of h give smooth density estimates which follow the empirical d.f. 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 $\hat{\theta}$). We choose $h = 0.02$ which provides a reasonably smooth estimate of F. Table 1 shows the results for the complete bootstrap for $B = 500$.

The last four columns of Table 1 provides the confidence intervals: $(\tilde{\theta}^{*(0.025)}_k$, $\tilde{\theta}^{*(0.975)})$ the 95% bias corrected confidence interval for θ (see (2.18)), and its correction for centering the intervals on $\tilde{\theta}_k$ (see (2.19)).

Note that here, the median of $\tilde{\theta}^*_{k,b}$ being not so far from $\tilde{\theta}_k$ the latter confidence intervals are very similar to the preceding ones. This table allows to appreciate the sensitivity of the efficiency measures w.r.t. the sampling variations. It shows clearly how careful should be the relative comparisons of the performances among firms based on $\hat{\theta}_k$ (compare the first two rows of Table 1). A graphical representation of the distribution of $\tilde{\theta}^*_{k,b}$ $b = 1, \ldots, 500$, $k = 1, \ldots, 19$ is displayed on figure 2.

The box-plots there facilitate the comparison among firms. For each firm the box represents the 50% mid range values of $\tilde{\theta}^*_{k,b}$; its length is the interquantile range (IQR). The whiskers define the natural bounds of the distributions (the mean ± 1.5 (IQR)), the crosses represent outliers standing out of the regular bounds.

Looking at the IQR's, the firms $\#1, 7, 8, 9, 10, 16$ stand below the other ones, (particularly the firm #8). No firm looks like dominating the others, although the regular good behavior of firms $\#4, 11, 13$ and 19 should be stressed.

The sensitivity of the results to the chosen bandwidth h in the bootstrap may be appreciated by looking to Tables 2 and 3. The results are not really sensitive to this choice although for $h = 0.01$, more weight is given near the upper bound of θ and on the contrary for $h = 0.04$ all the distributions are slightly moved to the left. Table 4 illustrates the results of the Leave-one out bootstrap. Note that the sensitivity analysis cannot be performed for the unit #15, note also that a few $\hat{\theta}_k$ are larger than one. Note also that for firms such that $\hat{\theta}_k < 1$, the results are very similar (as expected) to the complete bootstrap.

References

[1] Andersen, P. and N.C. Petersen (1989), A Procedure for Ranking Efficient Units in Data Envelopment Analysis, unpublished working paper, Department of Management, Odense University, Odense, Denmark.

[2] Charnes, A., Cooper W.W. and E. Rhodes (1978), Measuring the inefficiency of decision making units, *European Journal of Operational Research* 2, 429-444.

[3] Charnes, A., Cooper W.W. and R.M. Thrall (1986), Identifying and classifying scale and technical efficiencies in data envelopment analysis, *Operations Research Letters* 5, 105-116.

[4] Charnes, A., Cooper W.W. and R.M. Thrall (1991), A structure or classifying and characterizing efficiency in data envelopment analysis, *Journal of Productivity Analysis* 2, 197-237.

[5] Deprins, D. and L. Simar (1983), On Farrell measures of technical efficiency, *Recherches Economiques de Louvain* 49(2), 123-137.

[6] Deprins, D., Simar, L. and H. Tulkens (1984), Measuring labor inefficiency in post offices, in *The Performance of Public Enterprises: Concepts and measurements*. M. Marchand, P. Pestieau and H. Tulkens (eds.), Amsterdam, North-Holland, 243-267.

[7] Efron, B. (1979), Bootstrap methods: another look at the jackknife, *Annals of Statistics* 7, 1-26.

[8] Efron, B. (1982), *The jackknife, the bootstrap and other resampling plans* 38, *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM.

[9] Efron, B. and R.J. Tibshirani (1993), *An Introduction to the Bootstrap*. Chapman and Hall, London.

[10] Färe, R., Grosskopf, S. and E.C. Kokkelenberg (1989), Measuring plant Capacity, utilization and technical change: A nonparametric approach,*International Economic Review* 30, 655-666.

[11] Färe, R., Grosskopf, S. and C.A.K. Lovell (1985), *The Measurement of Efficiency of Production*. Boston, Kluwer-Nijhoff Publishing.

[12] Farrell, M.J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society* A(120), 253-281.

[13] Freedman, D.A. (1982), Bootstrapping regression models, *The Annals of Statistics* 9(6), 1218-1228.

[14] Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New-York.

[15] Hall, P., Härdle, W. and L. Simar (1991), Iterated bootstrap with application to frontier models, CORE Discussion paper 9121, UCL, to appear in the *Journal of Productivity Analysis*.

[16] Härdle, W. and E. Mammen (1991), Bootstrap methods in nonparametric regression in *Nonparametric Functional Estimation and Related Topics*. G. Roussas (eds.), Kluwer Academic Publishers, 111-123.

[17] Korostelev, A., Simar, L. and A. Tsybakov (1992), Efficient estimation of monotone boundaries, Discussion paper 9209, Institut de Statistique, UCL, to appear in the *Annals of Statistics*.

[18] Korostelev, A., Simar, L. and A. Tsybakov (1995), On estimation of monotone and convex boundaries, *Publications de l'ISUP* XXXIX(1), 3-18.

[19] Lovell, C.A.K., Walters L.C. and L.L. Wood (1993), Stratified models of education production using DEA and regression analysis in *Data Envelopment Analysis: Theory, Methods, and Applications*, eds. A. Charnes, W.W. Cooper, A.Y. Lewin and L.M. Seiford. New York: Quorom Books.

[20] Sheppard, R.W. (1970), *Theory of Cost and Production Function*. Princeton University Press, Princeton, New-Jersey.

[21] Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

[22] Silverman, B.W. and G.A. Young (1987), The bootstrap: smooth or not to smooth?, *Biometrika* 74, 469-479.

[23] Simar, L. (1992), Estimating efficiencies from frontier models with panel data: a comparison of parametric, non-parametric and semi-parametric methods with bootstrapping, *Journal of Productivity Analysis* 3, 167-203.

[24] Thrall, R.M. (1993), Duality, Classification, and Slacks in DEA, unpublished working paper, Jones Graduate School of Administration, Rice University, Houston, Texas.

[25] Wilson, P. (1994), A bootstrap methodology for nonparametric efficiency estimates, manuscript, Department of Economics, University of Texas, Austin.

[26] Wilson, P.W. (1995), Detecting influential observations in data envelopment analysis, *Journal of Productivity Analysis*, forthcoming.

[27] Wu, C.F.J (1986), Jackknife, bootstrap, and other resampling methods in regression analysis, *Annals of Statistics* 14, 1261-1295.

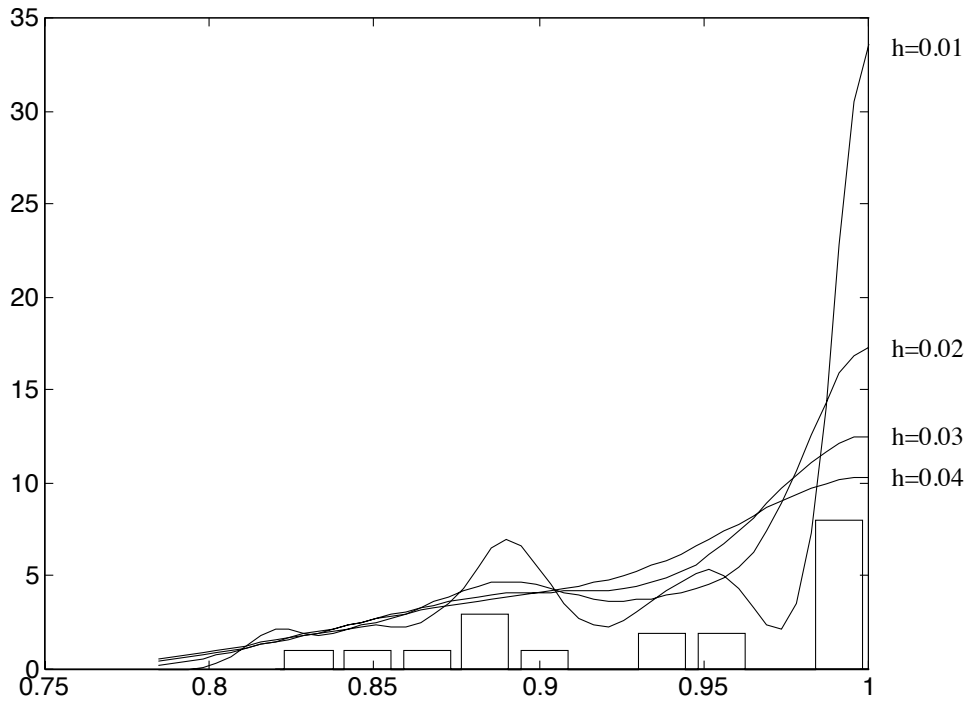

figure 1 kernel estimates of F

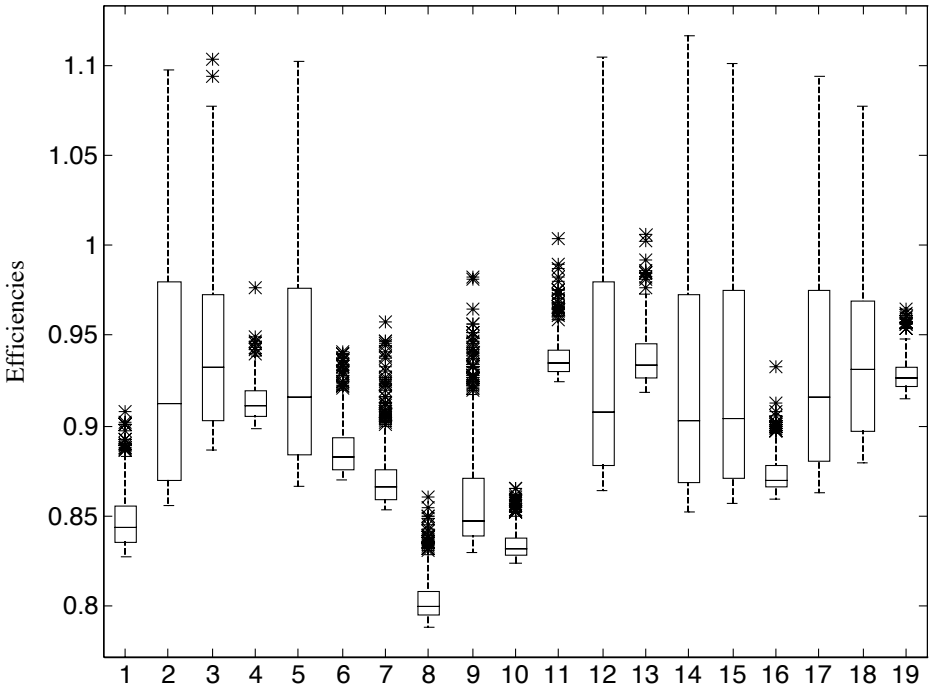

figure 2 Firm number

TABLE 1: COMPLETE BOOTSTRAP with bandwidth $h = 0.0200$

k	$\hat{\theta}_k$	$\tilde{\theta}_k$	$\widehat{bias}_k$	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	0.8475	0.0217	0.8440	0.0151	0.8292	0.8869	0.8308	0.9008
2	1.0000	0.9270	0.0730	0.9124	0.0592	0.8585	1.0463	0.8592	1.0630
3	1.0000	0.9427	0.0573	0.9325	0.0457	0.8892	1.0468	0.8900	1.0527
4	0.9307	0.9136	0.0170	0.9110	0.0101	0.9004	0.9372	0.9013	0.9458
5	1.0000	0.9319	0.0681	0.9168	0.0536	0.8687	1.0485	0.8698	1.0621
6	0.9071	0.8873	0.0199	0.8827	0.0144	0.8710	0.9283	0.8728	0.9375
7	0.8915	0.8717	0.0199	0.8660	0.0180	0.8550	0.9245	0.8577	0.9475
8	0.8210	0.8036	0.0174	0.8008	0.0120	0.7896	0.8379	0.7909	0.8496
9	0.8892	0.8593	0.0300	0.8475	0.0295	0.8329	0.9410	0.8364	0.9828
10	0.8469	0.8342	0.0127	0.8326	0.0072	0.8253	0.8525	0.8263	0.8609
11	0.9534	0.9376	0.0158	0.9355	0.0109	0.9255	0.9685	0.9267	0.9829
12	1.0000	0.9308	0.0692	0.9075	0.0596	0.8658	1.0578	0.8680	1.0741
13	0.9602	0.9383	0.0219	0.9344	0.0147	0.9211	0.9725	0.9225	0.9864
14	1.0000	0.9254	0.0746	0.9035	0.0652	0.8553	1.0731	0.8578	1.0848
15	1.0000	0.9276	0.0724	0.9047	0.0644	0.8592	1.0727	0.8610	1.0896
16	0.8885	0.8732	0.0153	0.8702	0.0100	0.8611	0.8979	0.8624	0.9082
17	1.0000	0.9303	0.0697	0.9161	0.0535	0.8649	1.0437	0.8662	1.0545
18	1.0000	0.9388	0.0612	0.9319	0.0465	0.8819	1.0436	0.8828	1.0543
19	0.9441	0.9285	0.0156	0.9266	0.0086	0.9167	0.9485	0.9178	0.9578

TABLE 2: COMPLETE BOOTSTRAP with bandwidth $h = 0.0100$

k	$\hat{\theta}_k$	$\tilde{\theta}_k$	$\widehat{bias}_k$	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	0.8533	0.0160	0.8468	0.0154	0.8385	0.8971	0.8399	0.9187
2	1.0000	0.9307	0.0693	0.9178	0.0620	0.8629	1.0511	0.8634	1.0799
3	1.0000	0.9480	0.0520	0.9432	0.0481	0.8972	1.0590	0.8973	1.0598
4	0.9307	0.9193	0.0114	0.9159	0.0095	0.9092	0.9424	0.9108	0.9661
5	1.0000	0.9356	0.0644	0.9250	0.0565	0.8730	1.0560	0.8733	1.0756
6	0.9071	0.8930	0.0141	0.8872	0.0143	0.8803	0.9335	0.8823	0.9591
7	0.8915	0.8768	0.0147	0.8691	0.0189	0.8631	0.9339	0.8655	0.9735
8	0.8210	0.8090	0.0121	0.8044	0.0122	0.7981	0.8464	0.7998	0.8598
9	0.8892	0.8637	0.0255	0.8480	0.0313	0.8395	0.9489	0.8411	0.9777
10	0.8469	0.8394	0.0075	0.8376	0.0061	0.8332	0.8561	0.8340	0.8730
11	0.9534	0.9436	0.0098	0.9405	0.0106	0.9351	0.9832	0.9362	1.0042
12	1.0000	0.9341	0.0659	0.9145	0.0624	0.8697	1.0600	0.8708	1.0744
13	0.9602	0.9446	0.0156	0.9392	0.0148	0.9309	0.9841	0.9332	1.0278
14	1.0000	0.9285	0.0715	0.9105	0.0678	0.8585	1.0755	0.8600	1.0853
15	1.0000	0.9303	0.0697	0.9114	0.0676	0.8617	1.0747	0.8627	1.0894
16	0.8885	0.8785	0.0100	0.8750	0.0093	0.8696	0.9045	0.8713	0.9389
17	1.0000	0.9340	0.0660	0.9242	0.0567	0.8694	1.0582	0.8703	1.0705
18	1.0000	0.9441	0.0559	0.9404	0.0488	0.8898	1.0436	0.8902	1.0566
19	0.9441	0.9345	0.0096	0.9323	0.0075	0.9262	0.9536	0.9277	0.9685

TABLE 3: COMPLETE BOOTSTRAP with bandwidth $h = 0.0400$

k	$\hat{\theta}_k$	$\tilde{\theta}_k$	$\widehat{bias}_k$	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	0.8380	0.0312	0.8362	0.0155	0.8157	0.8776	0.8164	0.8829
2	1.0000	0.9209	0.0791	0.9094	0.0540	0.8530	1.0285	0.8544	1.0406
3	1.0000	0.9338	0.0662	0.9263	0.0423	0.8775	1.0259	0.8803	1.0534
4	0.9307	0.9038	0.0269	0.9014	0.0116	0.8863	0.9290	0.8881	0.9354
5	1.0000	0.9257	0.0743	0.9119	0.0499	0.8631	1.0310	0.8675	1.0666
6	0.9071	0.8771	0.0300	0.8739	0.0148	0.8570	0.9120	0.8584	0.9182
7	0.8915	0.8625	0.0290	0.8571	0.0179	0.8422	0.9091	0.8449	0.9284
8	0.8210	0.7947	0.0264	0.7922	0.0127	0.7775	0.8269	0.7791	0.8325
9	0.8892	0.8506	0.0386	0.8417	0.0276	0.8212	0.9223	0.8246	0.9601
10	0.8469	0.8253	0.0216	0.8239	0.0090	0.8131	0.8466	0.8139	0.8538
11	0.9534	0.9276	0.0258	0.9246	0.0121	0.9117	0.9548	0.9135	0.9643
12	1.0000	0.9247	0.0753	0.9112	0.0546	0.8606	1.0485	0.8650	1.0825
13	0.9602	0.9280	0.0322	0.9243	0.0161	0.9064	0.9676	0.9079	0.9775
14	1.0000	0.9200	0.0800	0.9005	0.0613	0.8520	1.0698	0.8560	1.1046
15	1.0000	0.9224	0.0776	0.9045	0.0592	0.8552	1.0665	0.8595	1.0915
16	0.8885	0.8638	0.0246	0.8616	0.0114	0.8477	0.8891	0.8497	0.8980
17	1.0000	0.9238	0.0762	0.9131	0.0490	0.8588	1.0301	0.8612	1.0383
18	1.0000	0.9294	0.0706	0.9219	0.0435	0.8703	1.0285	0.8721	1.0375
19	0.9441	0.9186	0.0255	0.9170	0.0103	0.9031	0.9431	0.9043	0.9475

TABLE 4: LEAVE-ONE OUT BOOTSTRAP with bandwidth $\boldsymbol{h} = 0.0200$

k	$\hat{\theta}_k$	$\tilde{\theta}_k$	$\widehat{bias}_k$	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	0.8423	0.0269	0.8413	0.0156	0.8199	0.8792	0.8201	0.8828
2	1.1309	1.0495	0.0814	1.0393	0.0543	0.9820	1.1620	0.9834	1.1826
3	1.0510	0.9777	0.0734	0.9621	0.0540	0.9126	1.1066	0.9158	1.1169
4	0.9307	0.9122	0.0185	0.9097	0.0101	0.8986	0.9353	0.9014	0.9494
5	1.0978	1.0291	0.0687	1.0143	0.0468	0.9716	1.1334	0.9755	1.1545
6	0.9071	0.8845	0.0226	0.8801	0.0144	0.8664	0.9201	0.8685	0.9295
7	0.8915	0.8669	0.0246	0.8581	0.0218	0.8466	0.9256	0.8496	0.9535
8	0.8210	0.8016	0.0195	0.7996	0.0115	0.7861	0.8313	0.7874	0.8443
9	0.8892	0.8442	0.0451	0.8297	0.0370	0.8042	0.9304	0.8066	0.9468
10	0.8469	0.8331	0.0138	0.8317	0.0069	0.8233	0.8483	0.8241	0.8533
11	0.9534	0.9360	0.0174	0.9337	0.0117	0.9222	0.9682	0.9240	0.9813
12	1.1465	1.0930	0.0535	1.0864	0.0343	1.0480	1.1733	1.0523	1.1893
13	0.9602	0.9326	0.0276	0.9287	0.0175	0.9096	0.9764	0.9115	0.9905
14	1.4857	1.3853	0.1004	1.3629	0.0763	1.2904	1.5332	1.2927	1.5558
15	-	-	-	-	-	-	-	-	-
16	0.8885	0.8714	0.0171	0.8693	0.0099	0.8584	0.8993	0.8607	0.9058
17	1.1026	1.0316	0.0709	1.0255	0.0439	0.9717	1.1323	0.9739	1.1436
18	1.0675	1.0044	0.0630	0.9910	0.0494	0.9462	1.1139	0.9485	1.1261
19	0.9441	0.9267	0.0174	0.9246	0.0099	0.9138	0.9500	0.9149	0.9560