A General Methodology for Bootstrapping in Nonparametric Frontier Models

Léopold Simar*

and

Paul W. Wilson**

April, 1998

Abstract

The Data Envelopment Analysis method has been extensively used in the literature to provide measures of firms' technical efficiency. These measures allow rankings of firms by their apparent performance. The underlying frontier model is nonparametric since no particular functional form is assumed for the frontier model. Since the observations result from some data-generating process, the statistical properties of the estimated efficiency measures are essential for their interpretations. In the general multi-output, multi-input framework, the bootstrap seems to offer the only means of inferring these properties (i.e. to estimate the bias and variance, and to construct confidence intervals). This paper proposes a general methodology for bootstrapping in frontier models, extending the more restrictive method proposed in Simar and Wilson (1998a) by allowing for heterogeneity in the structure of efficiency. A numerical illustration with real data is provided to illustrate the methodology.

Key words: DEA, Bootstrapping, Nonparametric estimation, Efficiency, Frontier models.

*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

A large literature concerning the measurement of efficiency in production has developed since Debreu (1951) and Farrell (1957) provided basic definitions for technical and allocative efficiency in production. One large section of this literature focuses on linear-programming based measures of efficiency along the lines of Charnes *et al.* (1978, 1979), Deprins *et al.* (1984), and Färe *et al.* (1985).[1] Among this part of the literature, those approaches which rely on convexity assumptions are known as Data Envelopment Analysis (DEA).

Lovell (1993) and others have labeled these approaches to efficiency measurement as *deterministic,* as if to suggest that DEA models have no statistical underpinnings. Although there have been numerous applications of DEA models to examine efficiency in a variety of industries, previous studies invariably present point estimates of inefficiency, with no measure and only slight or no discussion of uncertainty surrounding these estimates.[2] Yet, DEA models measure efficiency relative to a nonparametric, maximum likelihood *estimate* of an unobserved *true* frontier, conditional on observed data resulting from an underlying (and unobserved) data-generating process (DGP).[3] Consequently, estimates of efficiency from DEA models are subject to uncertainty due to sampling variation.

Banker (1993) proved the consistency of the *output*-oriented efficiency scores in the case of a *single* output, but gives no indication of the achieved rate of convergence. Korostelev *et al.* (1995a, 1995b) also analyzed the single-output problem and derive the speed of convergence for the estimated attainable production set (using either the Hausdorff metric or the Lebesgue measure of symmetric differences between the true and the estimated production sets), but not for the estimated measures of efficiency. The theory of statistical

[1] Estimation of efficiency using the free disposal hull (FDH) method of Deprins *et al.* typically is accomplished without solving linear programs, although the problem is sometimes formulated in terms of linear programs.

[2] See Lovell (1993) and Seiford (1996, 1997) for excellent biblographies of DEA applications.

[3] We discuss the nature of the DGP implicit in DEA models below; see also Simar (1996) and Kneip *et al.* (1998).

consistency in DEA models has been extended to the general multi-input *and* multi-output case for both input- and output-oriented efficiency measures in Kneip *et al.* (1998), where the rates of convergence are also derived.

Due to the complexity and multidimensional nature of DEA estimators, the sampling distributions of the estimators are not easely available. In the very particular case of one-input and one-output, Gijbels *et al.* (1996) derived the asymptotic sampling distribution of the DEA estimator, with an expression for its asymptotic bias and variance. However, in the more useful multi-ouput and multi-input case, the bootstrap methodology seems, so far, to be the only way to investigate sampling properties of DEA estimators. Simar and Wilson (1998a) proposed a bootstrap strategy for analyzing the sensitivity of the efficiency measures to sampling variation providing confidence intervals and corrections for the bias inherent to the DEA procedure. The method has been applied to time dependence structures for estimating Malmquist indices (Simar and Wilson, 1998b). However, the methodology in both of the papers relies on some restrictive homogeneity assumptions on the distribution of efficiency among firms.

This paper presents an extension of the method to allow for more general DGP's, in particular for less restrictive efficiency structures. Section 2 introduces some basic concepts and notation along with a brief introduction to DEA estimators. In section 3 we analyse the underlying statistical model and explicitly state our assumptions regarding the DGP. In section 4 we propose a general bootstrap procedure. An empirical illustrations are given in section 5, and conclusions are discussed in section 6. Although the discussion throughout is in terms of the Shephard (1970) input distance functions, it is straightforward to adapt the techniques presented below to output distance functions or to other nonparametric models (e.g., FDH models), including DEA measures of allocative efficiency and overall efficiency such as those described by Färe *et al.* (1985).

2. Frontier Analysis and DEA Estimators

2.1 The Frontier Model

Given column vectors of p inputs (denoted by $x \in \mathbb{R}_+^p$) and of q outputs (denoted by $y \in \mathbb{R}_+^q$), the activity of a productive organization can be described by means of the production set Ψ of physically attainable points (x, y):

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

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

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

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

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

Clearly,

$$x \in X(y) \Leftrightarrow y \in Y(x). \tag{2.4}$$

The relations between these two sets, along with standard assumptions one may reasonably make on them, are discussed in section 9.1 of Shephard (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 disposability assumptions correspond to some monotonicity of the frontier:

$$y_1 \leq y_2 \Rightarrow X(y_2) \subseteq X(y_1) \tag{2.5}$$

$$x_1 \leq x_2 \Rightarrow Y(x_1) \subseteq Y(x_2). \tag{2.6}$$

where the inequality between the vectors is understood to be component-wise.

The Farrell efficiency boundaries are subsets of $X(y)$ (and $Y(x)$, respectively) denoted by $\partial X(y)$ and $\partial Y(x)$, respectively:

$$\partial X(y) = \{x \mid x \in X(y), \theta x \notin X(y) \quad \forall\, 0 < \theta < 1\} \tag{2.7}$$

$$\partial Y(x) = \{y \mid y \in Y(x), \beta y \notin X(y) \quad \forall\, \beta > 1\}. \tag{2.8}$$

These may be used to define the Farrell input and output measures of efficiency (respectively) for a given point (x, y) in Ψ:

$$\theta(x, y) = \inf\{\theta \mid \theta x \in X(y)\} \tag{2.9}$$

$$\lambda(x, y) = \sup\{\lambda \mid \lambda y \in Y(x)\}. \tag{2.10}$$

We will discuss only the input-oriented case to conserve space; the output-oriented case largely involves a straightforward translation of the notation in what follows.

Equivalently, Farrell's input efficiency may be described by the Shephard input distance function

$$\delta(x, y) = (\theta(x, y))^{-1} = \sup\{\delta \mid \frac{x}{\delta} \in X(y)\}. \tag{2.11}$$

The input distance function δ gives a normalized measure of the distance from a point (x, y) to the frontier $\partial X(y)$, holding output and the direction of the input vector fixed. Since the direction of the input vector is held fixed, δ is said to be a *radial* measure; it gives the maximum feasible, proportionate reduction of inputs for a firm operating at $(x, y) \in \Psi$. Clearly, $\delta(x, y) \geq 1$ if and only if $x \in X(y)$; if $\delta(x, y) = 1$, then $(x, y) \in \partial X(y)$ and the point (x, y) is said to be *input-efficient*. It will be useful later to denote the efficient level of input, corresponding to the output level y and the input vector direction determined by x, as

$$x^{\partial}(y) = \frac{x}{\delta(x, y)}. \tag{2.12}$$

Note that $x^{\partial}(y)$ is the intersection of $\partial X(y)$ and the ray $(\theta x, y), \theta \in [0, \infty]$.

Since radial distances are used, we will often refer to the polar coordinates of x defined by its modulus $\omega = \omega(x) \in \mathbb{R}_+$ where $\omega(x) = \sqrt{(x'x)}$, and its angle $\eta = \eta(x) \in [0, \frac{\pi}{2}]^{p-1}$ where, for $j = 1, \ldots, p-1$, $\eta_j = \arctan(\frac{x_{j+1}}{x_1})$ if $x_1 > 0$ or $\eta_j = \frac{\pi}{2}$ if $x_1 = 0$.

Typically, Ψ, $X(y)$, and $\partial X(y)$ are unknown; hence, for an observed firm producing at (x, y), $\delta(x, y)$ is also unknown. The DEA technique provides a consistent estimator of $\delta(x, y)$ from a random sample $\mathcal{X} = \{(x_i, y_i) | i = 1, \ldots, n\}$.

2.2 The DEA Approach

The DEA approach involves measurement of efficiency for a given firm at (x, y) relative to the boundary of either the convex or conical hulls of the data $\mathcal{X} = \{(x_i, y_i), i = 1, \ldots, n\}$ intersected with the free-disposal hull. The intersection of the convex and free-disposal hulls is given by[4]

$$\widehat{\Psi} = \{(x, y) \in \mathbb{R}^{p+q} \mid y \leq \sum_{i=1}^{n} \gamma_i y_i,\ x \geq \sum_{i=1}^{n} \gamma_i x_i,\ \sum_{i=1}^{n} \gamma_i = 1,\ \gamma_i \geq 0,\ i = 1, \ldots, n\}. \quad (2.13)$$

Replacing Ψ in (2.2) and (2.7) with $\widehat{\Psi}$ yields estimates of the input requirement set and the input-efficient boundary for the output level y, respectively:

$$\widehat{X}(y) = \{x \in \mathbb{R}^p | (x, y) \in \widehat{\Psi}\} \quad (2.14)$$

$$\partial\widehat{X}(y) = \{x \in \mathbb{R}^p | x \in \widehat{X}(y), \theta x \notin \widehat{X}(y) \quad \forall 0 < \theta < 1\}. \quad (2.15)$$

Finally, for any given point (x, y), the estimator $\widehat{\delta}(x, y)$ of $\delta(x, y)$ is obtained by substituting $\widehat{X}(y)$ from (2.14) for $X(y)$ in (2.11), yielding:

$$\widehat{\delta}(x, y) = \sup\{\delta | x/\delta \in \widehat{X}(y)\} \quad (2.16)$$

To make this operational, we rewrite (2.16) as a linear program:

$$\left(\widehat{\delta}(x, y)\right)^{-1} = \min\{\theta > 0 \mid y \leq \sum_{i=1}^{n} \gamma_i y_i,\ \theta x \geq \sum_{i=1}^{n} \gamma_i x_i,\ \sum_{i=1}^{n} \gamma_i = 1,\ \gamma_i \geq 0,\ i = 1, \ldots, n\}. \quad (2.17)$$

Analogous to (2.12) we have an estimator of the input-efficient level of inputs,

$$\widehat{x}^{\partial}(y) = \frac{x}{\widehat{\delta}(x, y)}. \quad (2.18)$$

[4] Use of the convex hull to estimate Ψ allows for the possibility of varying returns to scale. If the underlying technology is one of constant returns to scale, we could delete the constraint $\sum_{i=1}^{n} \gamma_i = 1$ in (2.13) to obtain the conical hull (intersected with the free-disposal hull)(intersected with the free-disposal hull) as an estimate of Ψ. Alternatively, replacing the equality constraint in (2.13) by the inequality $\sum_{i=1}^{n} \gamma_i \leq 1$ assumes nonincreasing returns to scale. Since we do not know the true production set Ψ, we use the less-constrained convex hull to estimate Ψ.

Note that $\widehat{\Psi} \subseteq \Psi$, and so $\partial\widehat{X}(y)$ is an upward-biased estimator of $\partial X(y)$, and $\widehat{\delta}(x, y)$ is a downward-biased estimator of $\delta(x, y)$ so that

$$\widehat{\delta}(x, y) \leq \delta(x, y) \quad \forall\ (x, y) \in \Psi. \tag{2.19}$$

Moreover,

$$\widehat{\delta}(x, y) \geq 1 \quad \forall\ (x, y) \in \widehat{\Psi}. \tag{2.20}$$

Thus, for each of the sample points in $\mathcal{X}$, we have

$$\widehat{\delta}(x_i, y_i) \geq 1 \quad \forall\ i = 1, \dots, n. \tag{2.21}$$

In order to perform statistical inference on the estimated input distances, we must analyse the behavior of the difference $(\widehat{\delta}(x, y) - \delta(x, y))$ for a given (x, y) by investigating its sampling distribution.

3. The Statistical Model and Consistency of DEA

Our statistical model is defined through the following assumptions which were used by Kneip *et al.* (1998). These assumptions serve to characterize the DGP, and augment Shephard's (1970) assumptions, which are more concerned with the nature of the underlying production set.

Assumption A1. $\{(x_i, y_i),\ i = 1, \dots, n\}$ *are i.i.d. random variables on the convex production set* Ψ*.*

For the input-oriented case presented here, consider a point $(x, y) \in \Psi$. Then for a given output level y, the corresponding input level x falls on the ray $(\delta x^{\partial}(y), y)$, $\delta \in [1, \infty]$; the deviation of (x, y) away from $\partial X(y)$ is assumed to result from technical inefficiency.

Assumption A2. *Outputs* y *possess a density* $f(\cdot)$ *whose bounded support* $\mathcal{D} \subseteq \mathbb{R}^q_+$ *is compact.*

Due to the radial nature of the inefficiency, the conditional pdf of x for a given y is more naturally introduced through the polar coordinates of x.

Assumption A3. *For all $y \in \mathcal{D}$, $\eta = (\eta_1, \dots, \eta_{p-1})$ has a conditional pdf $f(\eta \mid y)$ on $[0, \frac{\pi}{2}]^{p-1}$ and conditional on (y, η), the modulus ω has a density $f(\omega \mid y, \eta)$.*

Note that, for a given (y, η) the efficient input-level $x^{\partial}(y)$ defined by (2.12) has modulus

$$\omega(x^{\partial}(y)) = \inf\{\omega \in \mathbb{R}_+ \mid f(\omega \mid y, \eta) > 0\}. \tag{3.1}$$

The relation between $\omega(x)$ and the input distance function $\delta(x, y)$ is given by

$$\delta(x, y) = \frac{\omega(x)}{\omega(x^{\partial}(y))}. \tag{3.2}$$

Assumption $A3$ induces, by (3.2), a conditional pdf for $\delta(x, y)$ given (y, η), namely $f(\delta \mid y, \eta)$, with support $[1, \infty)$[5].

In order to achieve consistency in any nonparametric estimation of spatial boundaries, the DGP must insure that points will be observed near the frontier when n is sufficiently large. This imposes, in our case, an assumption on the conditional pdf of ω given (y, η):

Assumption A4. *For all $y \in \mathcal{D}$ and all $\eta \in [0, \frac{\pi}{2}]^{p-1}$, there exist constants $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that $\forall\, \omega \in [\omega(x^{\partial}(y)), \omega(x^{\partial}(y)) + \epsilon_2]$, $f(\omega \mid y, \eta) \geq \epsilon_1$.*

Again, this implies by (3.2) that $f(\delta \mid y, \eta) \geq \epsilon_1 \;\forall\, \delta \in [1, 1 + \epsilon_2]$.

In addition, Kneip *et al.* (1998) make two additional assumptions regarding the smoothness of the frontier in order to derive the rates of convergence; these may be written in terms of $\delta(x, y)$. For sake of simplicity, we assume here:[6]

Assumption A5. *The distance function $\delta(x, y)$ is differentiable in both its arguments.*

Consider now a fixed point (x, y) and a sample $\mathcal{X} = \{(x_i, y_i) | i = 1, \dots, n\}$ generated by the DGP defined by *A1* to *A4* in terms of polar coordinates, where the modulus is defined through the distance function with respect to the frontier by (3.2). An observation

[5] Even with this nonrestrictive assumption, consistency of $\widehat{\delta}(x, y)$ for a given (x, y) is achieved. Simar and Wilson (1998a, 1998b) imposed independence between δ and (y, η) so that $f(\delta \mid y, \eta) = f(\delta)$. We avoid this restriction here, allowing the distribution of radial inefficiency to vary over the production set.

[6] For more details on the role of the smoothness of the frontier on the rates of convergence, see Kneip *et al.* (1998). Assumption A5 here is slightly stronger than assumptions A5–A6 in Kneip *et al.,* but can be relaxed easily; the only cost is an increase in notational complexity.

$(x_i, y_i) \in \mathbb{R}^{p+q}$ has as its polar coordinate representation

$$(y_i, \eta_i, \delta_i) \in \mathcal{A} = \mathbb{R}_+^q \times [0, \pi/2]^{p-1} \times [1, \infty), \tag{3.3}$$

where

$$\delta_i = \frac{\omega(x_i)}{\omega(x^\delta(y_i))}. \tag{3.4}$$

The DGP is completely defined through the density of (y_i, η_i, δ_i) on $\mathcal{A}$:

$$f(y_i, \eta_i, \delta_i) = f(\delta_i \mid y_i, \eta_i) f(\eta_i \mid y_i) f(y_i). \tag{3.5}$$

Kneip *et al.* (1998) proved that for a fixed point (x, y),

$$\widehat{\delta}(x, y) - \delta(x, y) = \mathcal{O}_P(n^{-\frac{2}{p+q+1}}) \tag{3.6}$$

Thus, $\widehat{\delta}(x, y)$ is a consistent estimator of $\delta(x, y)$, but the rate of convergence is low; furthermore, by construction, $\widehat{\delta}(x, y)$ is downward-biased. Unfortunately, very few results exist on the sampling distribution of $\widehat{\delta}(x, y)$. Gijbels *et al.* (1996) derived the asymptotic distribution of $\widehat{\delta}(x, y)$ in the special case of one input and one output ($p = q = 1$) along with an analytic expression for its large sample bias and variance. This would allow one to construct a bias-corrected estimator and confidence intervals for this special case. Unfortunately, in the more general multivariate setting, the radial nature of the distance function and the complexity of the estimated frontier complicates the derivations; so far, the bootstrap appears to offer the only way to approximate this asymptotic distribution. The next section proposes a general methodology for bootstrapping the distribution of $\widehat{\delta}(x, y) - \delta(x, y)$.

4. The Bootstrap

4.1 The Principles

Let $\mathcal{P}$ denote the DGP defined by assumptions *A1–A4*, from which the random sample $\mathcal{X} = \{(x_i, y_i) \mid i = 1, \dots, n\}$ is obtained. Consider again a fixed point (x, y). From (2.16), we can obtain an estimate $\widehat{\delta}(x, y)$ of $\delta(x, y)$. Typically, we would choose this fixed point to correspond to one of the points in $\mathcal{X}$, but this is not necessary.

Given a consistent estimator $\widehat{\mathcal{P}}$ of $\mathcal{P}$ estimated from the data $\mathcal{X}$, consider now a new data set $\mathcal{X}^* = \{(x_i^*, y_i^*),\ i = 1, \dots, n\}$ drawn from $\widehat{\mathcal{P}}$. The convex hull of $\mathcal{X}^*$ gives an estimator $\widehat{\Psi}^*$ of $\widehat{\Psi}$, which from the perspective of $\mathcal{X}^*$, is the true set of possible values in $\mathcal{X}^*$, and which in our original setting was an estimate of Ψ defined in (2.1). Specifically, we have

$$\widehat{\Psi}^* = \{(x, y) \in \mathbb{R}^{p+q} \mid y \le \sum_{i=1}^{n} \gamma_i y_i^*,\ x \ge \sum_{i=1}^{n} \gamma_i x_i^*,\ \sum_{i=1}^{n} \gamma_i = 1,\ \gamma_i \ge 0,\ i = 1, \dots, n\}. \tag{4.1}$$

Analogous to (2.14)–(2.15), corresponding to $\widehat{\Psi}^*$ we have

$$\widehat{X}^*(y) = \{x \in \mathbb{R}^p | (x, y) \in \widehat{\Psi}^*\} \tag{4.2}$$

and

$$\partial \widehat{X}^*(y) = \{x \in \mathbb{R}^p \mid x \in \widehat{X}^*(y),\ \theta x \notin \widehat{X}^*(y)\ \forall\ 0 < \theta < 1\}. \tag{4.3}$$

Replacing $\widehat{X}(y)$ with $\widehat{X}^*(y)$ in (2.16), we have

$$\left(\widehat{\delta}^*(x, y)\right) = \sup\{\delta \mid x/\delta \in \widehat{X}^*(y)\}. \tag{4.4}$$

which may be evaluated by solving the linear program

$$\begin{aligned}\left(\widehat{\delta}^*(x, y)\right)^{-1} = \min\{\theta > 0 \mid y \le \sum_{i=1}^{n} \gamma_i y_i^*,\ \theta x \ge \sum_{i=1}^{n} \gamma_i x_i^*,\\ \sum_{i=1}^{n} \gamma_i = 1,\ \gamma_i \ge 0,\ i = 1, \dots, n\}.\end{aligned} \tag{4.5}$$

Note that conditional on $\mathcal{X}$, the sampling distribution of $\widehat{\delta}^*(x,y)$ is (in principle) completely known since $\widehat{\mathcal{P}}$ is known, although it may be difficult to compute analytically. However, the sampling distributions are easily approximated by Monte Carlo methods. Using $\widehat{\mathcal{P}}$ to generate B samples $\mathcal{X}_b^*$, $b = 1, \ldots, B$, yields a set of pseudo estimates $\widehat{\delta}_b^*(x,y)$, $b = 1, \ldots, B$. The empirical density function of these bootstrap values gives a Monte Carlo approximation of the sampling distribution of $\widehat{\delta}^*(x,y)$, conditional on $\widehat{\mathcal{P}}$.

The bootstrap method introduced by Efron (1979) (see also Efron, 1982, or Efron and Tibshirani, 1993) is based on the idea that if $\widehat{\mathcal{P}}$ is a consistent estimator of $\mathcal{P}$, the known bootstrap distributions will mimic the original unknown sampling distributions of the estimators of interest. More specifically,

$$\left(\widehat{\delta}^*(x,y) - \widehat{\delta}(x,y)\right) \mid \widehat{\mathcal{P}} \quad \stackrel{\text{approx}}{\sim} \quad \left(\widehat{\delta}(x,y) - \delta(x,y)\right) \mid \mathcal{P}. \tag{4.6}$$

A naive bootstrap would consist of sampling the pairs (x_i^*, y_i^*) with replacement from the original pairs in $\mathcal{X}$. In boundary problem such as this, the naive bootstrap yields inconsistent estimates (see Bickel and Freedman, 1981 or Beran and Ducharme, 1991 for discussions of this problem in the context of estimating the support of a univariate density; see Gijbels *et al.*, 1996 for a two dimensionnal case). To illustrate, consider a fixed point (x,y). With nonzero probability, the naive bootstrap estimate $\widehat{\delta}^*(x,y)$ will equal $\widehat{\delta}(x,y)$. This can be proven as follows. In $\mathbb{R}^{p+q}$, $\left(\widehat{X}(y), y\right)$ is a polyhedron defined by all the dominating facets of (x_i, y_i), $i = 1, \ldots, n$, where the dominating facets are given by $\left(\partial\widehat{X}(y), y\right)$ and are determined by $(p+q)$ observed efficient points. The probability that the bootstrap sample $\mathcal{X}^*$ contains the original dominating facet of (x,y) is then $\sum_{j=0}^{p+q} \begin{pmatrix} p+q \\ j \end{pmatrix} (-1)^j \left(1 - \frac{j}{n}\right)^n$, which has the limit $(1-e^{-1})^{p+q} > 0$ as $n \to \infty$. Consequently, in the naive bootstrap $\widehat{\delta}^*(x,y) = \widehat{\delta}(x,y)$ with nonzero probability; moreover, this problem does not go away as $n \to \infty$, and so the naive bootstrap is inconsistent.

One way to overcome this difficulty is to use a smoothed bootstrap, *i.e.,* to draw i.i.d. bootstrap samples $(x_i^*, y_i^*), i = 1, \ldots, n$, from a density $\widehat{f}(x,y)$ on $\widehat{\Psi}$ where $\widehat{f}(x,y)$ is a smooth, consistent estimator of the joint density of (x, y) on Ψ. In terms of the polar coordinates introduced in (3.3)–(3.4), this is equivalent to estimating the density $f(y, \eta, \delta)$ and drawing bootstrap samples $(y_i^*, \eta_i^*, \delta_i^*)$ from the estimated density. Unfortunately, $f(y, \eta, \delta)$ has bounded support as described by (3.3), and ordinary kernel density estimates are known to be biased and inconsistent near boundaries. Scott (1992) proposes the use of boundary kernels to deal with this problem in the univariate case, but it is not clear how this method could be extended to higher-dimensional spaces such as ours. Instead, we adopt the reflection method proposed by Silverman (1989) to estimate $f(y, \eta, \delta)$.[7] Since the frontier is unknown, δ is not directly observable in our sample. Therefore, we will use the estimator of $f(y, \eta, \widehat{\delta})$ from the set of points $\{(y_i, \eta_i, \widehat{\delta}_i),\ i = 1, \ldots, n\}$, where $\widehat{\delta}_i$ is the consistent DEA estimator of δ_i.

To estimate $f(y, \eta, \widehat{\delta})$ consistently, recall from (3.3) that $\widehat{\delta} \in [1, \infty]$. As noted earlier, in a given sample of size n we will likely observe many values of $\widehat{\delta}_i = 1$. To ensure consistency of our estimator $\widehat{f}$ of $f(y, \eta, \widehat{\delta})$, we reflect each point $(y_i, \eta_i, \widehat{\delta}_i)$ through the boundary at unity for $\widehat{\delta}_i$.[8] Let $\mathcal{Z}$ denote the $n \times (p + q)$ matrix

$$\mathcal{Z} = \begin{bmatrix} y_i & \eta_i & \widehat{\delta}_i \end{bmatrix} \tag{4.7}$$

so that the ith row of $\mathcal{Z}$ contains the observation for the ith firm expressed in polar coordinates. Then the matrix of points reflected about the boundary $\widehat{\delta}_i = 1$ is

$$\mathcal{Z}_R = \begin{bmatrix} y_i & \eta_i & 2 - \widehat{\delta}_i \end{bmatrix}. \tag{4.8}$$

[7] This idea was used in Simar and Wilson (1998a, 1998b) but there, the estimator of $\mathcal{P}$ incorporates homogeneity restrictions as described previously in footnote 5.

[8] The boundary conditions $y \geq 0$ and $\eta_j \in [0, \pi/2],\ j = 1, \ldots, p-1$ are not taken into account here for sake of simplicity and because in practice, these constraints are seldom binding. The problem mentioned above is related to the fact that for a given (x, y), the naive bootstrap frontier $\widehat{\partial X}^*(y)$ could be equal to the estimated frontier $\widehat{\partial X}(y)$ with positive probability. So only reflection about this boundary is necessary in practice. Our method could be adapted, if desired, to account for the boundary constraints on y and η but at a cost of complexity: $y_k,\ k = 1, \ldots, q$ would be reflected about 0 and $\eta_j \in [0, \pi/2],\ j = 1, \ldots, p-1$ would be reflected about zero and about $\pi/2$. This would lead to an augmented data matrix $\tilde{\mathcal{Z}}$ corresponding to (4.9) of size $(n * 2 * 2q * 3(p-1)) \times (p+q)$. We prefer, for simplicity, to delete any negative bootstrap value of y_i^* and any unfeasible value of η_i^* below, if by chance, such values should appear in the bootstrap algorithm.

Let z_i denote the ith row of $\mathcal{Z}$, and z_{Ri} denote the ith row of $\mathcal{Z}_R$. The "augmented" set of points is now given by the $2n \times (p+q)$ matrix

$$\tilde{\mathcal{Z}} = \begin{bmatrix} \mathcal{Z} \\ \mathcal{Z}_R \end{bmatrix}. \tag{4.9}$$

We wish to estimate the unknown density of the $2n$ unbounded points represented by $\tilde{\mathcal{Z}}$. We use a multivariate Gaussian kernel function scaled to have the same shape as the cloud of points in $(p+q)$-space represented in $\tilde{\mathcal{Z}}$. Specifically, let $\widehat{\Sigma}_1$ be an estimate of the covariance matrix of $\mathcal{Z}$, which can be written in partitioned form as

$$\widehat{\Sigma}_1 = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \tag{4.10}$$

where S_{11} is $(p+q-1) \times (p+q-1)$, $S_{12} = S_{21}'$ is $(p+q-1) \times 1$, and S_{22} is scalar.[9] Accordingly, the corresponding estimate of the covariance matrix of the reflected data is

$$\widehat{\Sigma}_2 = \begin{bmatrix} S_{11} & -S_{12} \\ -S_{21} & S_{22} \end{bmatrix}. \tag{4.11}$$

Finally, let $K_\ell(\cdot)$ be a $(p+q)$-variate Gaussian density with zero mean and shape $\widehat{\Sigma}_\ell$ for $\ell = 1, 2$; *i.e.*,

$$K_\ell(x) = (2\pi)^{-\frac{p+q}{2}} \left(\det(\widehat{\Sigma}_\ell)\right)^{-\frac{1}{2}} \exp(-\frac{1}{2} x' \widehat{\Sigma}_\ell^{-1} x), \quad \ell = 1, 2. \tag{4.12}$$

A consistent kernel estimator of the density of $z = (y, \eta, \delta)$ over the $2n$ observations in $\tilde{\mathcal{Z}}$ is given by

$$\tilde{f}_h(z) = \frac{1}{2nh^{(p+q)}} \sum_{i=1}^{n} \left[K_1\left(\frac{z - z_i}{h}\right) + K_2\left(\frac{z - z_{Ri}}{h}\right) \right], \tag{4.13}$$

[9] The data represented by the rows of $\mathcal{Z}$ may not fall in a regular ellipsoid due to the typical large number of cases where $\widehat{\delta}_i = 1$. From the viewpoint of the sample covariance matrix, this situation resembles the problem of outliers; furthermore, the sample covariance matrix is quite sensitive to outliers. A robust estimator with high breakdown point would be a better choice for $\widehat{\Sigma}_1$. We use the M-estimator of the covariance matrix described by Campbell (1980). Alternatively, one could use the minimum volume ellipsoid estimator proposed by Rousseeuw (1985) when $(p+q)$ is large, since the breakdown point of M-estimators decreases as the dimensionality increases (see Hampel *et al.*, 1986, ch.5, for discussion of this point). Unfortunately, the minimum volume ellipsoid estimator is more difficult to compute.

where h is the bandwidth.

Then a consistent estimator $\widehat{f}_h$ of f, with bounded support $\mathcal{A}$ defined in (3.3), is obtained by setting

$$\widehat{f}_h(z) = \begin{cases} 2\tilde{f}_h(z) & \text{if } z \in \mathcal{A} \\ 0 & \text{otherwise.} \end{cases} \tag{4.14}$$

The remaining question regards the proper choice of the bandwidth h. Consistency of $\tilde{f}_h(z)$, and hence $\widehat{f}_h(z)$, requires $h \to 0$ as $n \to \infty$, but not to quickly; in particular we need $h = \mathcal{O}(n^{-\frac{1}{p+q+4}})$. One possibility is to use the normal reference rule (*e.g.,* see Scott, 1992), which assigns

$$\widehat{h} = \left(\frac{4}{p+q+2}\right)^{\frac{1}{p+q+4}} n^{-\frac{1}{p+q+4}}. \tag{4.15}$$

This bandwidth minimizes the approximate mean integrated square error (MISE) when the data are normally distributed, and have been prewhittened to have unit variance and zero covariance. Unfortunately, the data in $\tilde{\mathcal{Z}}$ will almost certainly be highly nonnormal, and so using (4.15) to determine the bandwith will almost certainly fail to minimize the MISE. A better approach is to use the data to choose a value $\widehat{h}$ that minimizes an approximation to MISE; we provide such a method in the appendix.

To generate a random sample of size n from $\widehat{f}_h(z)$ in (4.11) is very easy; an algorithm is given in the next section. This provides pseudo observations $\{(y_i^*, \eta_i^*, \delta_i^*) \in \mathcal{A},\ i = 1, \ldots, n\}$, which can be translated back to the rectangular coordinates $\{(x_i^*, y_i^*),\ i = 1, \ldots, n\}$. To demonstrate this, consider the draw $(y_i^*, \eta_i^*, \delta_i^*)$ from $\widehat{f}_h(z)$ in (4.11). Given the output level y_i^* and the angle η_i^*, we must determine the location of the original estimated frontier $\widehat{\partial X}(y_i^*)$ in order to compute x_i^*. This is accomplished by solving an additional linear program.

Let $\tilde{x}$ be any point in the $x-$space on the ray with angle η_i^* (for instance, take $\tilde{x}_1 = 1$ and $\tilde{x}_{j+1} = \tan(\eta_{ij}^*)$ for $j = 1, \ldots, p-1$). For this point $(\tilde{x}, y_i^*)$, use (2.17) to compute $\widehat{\delta}(\tilde{x}, y_i^*)$. Given the output level y_i^* and the angle η_i^* contained in our draw from $\widehat{f}_h(z)$, we can replace $\widehat{\delta}(x, y)$ with $\widehat{\delta}(\tilde{x}, y_i^*)$ in (2.18) and the x appearing in the numerator of (2.18) with $\tilde{x}$ to obtain

$$\widehat{x}^{\partial *}(y_i^*) = \frac{\tilde{x}}{\widehat{\delta}(\tilde{x}, y_i^*)}, \tag{4.16}$$

which gives the (true) input-efficient level of inputs in the bootstrap world.[10] In other words, $\widehat{x}^{\partial *}(y_i^*)$ gives the value of the input vector with angle η_i^*, corresponding to output level y_i^*, and lying on the estimated frontier $\widehat{\partial X}(y_i^*)$.

Finally we can define the Cartesian coordinates x_i^* by computing

$$x_i^* = \delta_i^* \widehat{x}^{\partial *}(y_i^*). \tag{4.17}$$

Thus, x_i^* is formed by taking a random deviation away from the input vector $\widehat{x}^{\partial *}(y_i^*)$ lying on the estimated frontier $\widehat{\partial X(y_i^*)}$, consistent with our assumptions *A1-A4* regarding the underlying DGP.

Now, having constructed the pseudo-sample $\mathcal{X}^* = \{(x_i^*, y_i^*), \quad i = 1, \ldots, n\}$ we can compute the bootstrap value $\widehat{\delta}^*(x, y)$ by solving the linear program (4.5).[11] To clarify, we detail the various steps we have described above in algorithmic form in the next section.

4.2 The Algorithm

As seen from the previous section, the critical element of our bootstrap procedure involves generation of pseudodata $(y_i^*, \eta_i^*, \delta_i^*) \in \mathcal{A}$ from the density estimate $\widehat{f}_h(\cdot)$ given by (4.11). Our procedure consists of eleven steps, as detailed below:

[1] From the original data set $\mathcal{X}$, compute $\widehat{\delta}_i = \widehat{\delta}(x_i, y_i) \; \forall \; i = 1, \ldots, n$ using (2.16).

[2] Translate the data into polar coordinates: $(y_i, \eta_i, \widehat{\delta}_i) \; \forall \; i = 1, \ldots, n$, and form the augmented matrix $\tilde{\mathcal{Z}}$ as in (4.7)–(4.9).

[3] Compute the estimated covariance matrices $\widehat{\Sigma}_1$, $\widehat{\Sigma}_2$ as in (4.10)–(4.11), and the lower triangular matrices L_1 and L_2 such that $\widehat{\Sigma}_1 = L_1 L_1'$ and $\widehat{\Sigma}_2 = L_2 L_2'$ via the Cholesky decomposition.

[10] Replace (x, y) with $(\tilde{x}, y_i^*)$ in (2.17) and solve the resulting linear program to obtain $\widehat{\delta}(\tilde{x}, y_i^*)$.

[11] The linear program (4.5) will not have a solution if $y_i > \max_j(y_j^*)$; the solution is infeasible in this case since (x, y) lies above the bootstrap frontier. This is a problem of finite samples.

[4] Choose an appropriate bandwith h as described in the appendix using the information in $\tilde{\mathcal{Z}}$, $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$.

[5] Draw n rows randomly, with replacement from the augmented matrix $\tilde{\mathcal{Z}}$ and denote the result by the $n \times (p+q)$ matrix $\tilde{\mathcal{Z}}^*$; compute $\bar{z}^*$, the $1 \times (p+q)$ row vector containing the means of each column of $\tilde{\mathcal{Z}}^*$.

[6] Use a random number generator to generate an $n \times (p+q)$ matrix ϵ of i.i.d. standard normal pseudorandom variates; let $\epsilon_{i.}$ denote the i^{th} row of this matrix. Then compute the $n \times (p+q)$ matrix ϵ^* with i^{th} row $\epsilon^*_{i.}$ given by

$$\epsilon^*_{i.} = \epsilon_{i.} L'_\ell \tag{4.18}$$

so that $\epsilon^*_{i.} \sim N_{p+q}(0, \widehat{\Sigma}_\ell)$, where $\ell = 1$ if the ith row of $\tilde{\mathcal{Z}}^*$ was drawn from rows $1, \dots, n$ of $\mathcal{Z}$, or $\ell = 2$ if the ith row of $\tilde{\mathcal{Z}}^*$ was drawn from rows $(n+1), \dots, 2n$ of $\mathcal{Z}$.

[7] Compute the $n \times (p+q)$ matrix

$$\Gamma = (1+h^2)^{-\frac{1}{2}}(M\tilde{\mathcal{Z}}^* + h\epsilon^*) + i_n \otimes \bar{z}^* \tag{4.19}$$

where M is the usual $n \times n$ centering matrix ($M = I_n - \frac{1}{n} i_n i'_n$, where I_n is the identity matrix of order n and i_n a $n \times 1$ vector of ones) and $\otimes$ is the Kronecker product.[12]

[8] Partition Γ so that $\Gamma = [\gamma_{i1}, \gamma_{i2}, \gamma_{i3}]$, where $\gamma_{i1} \in \mathbb{R}^q_+$, $\gamma_{i2} \in [0, \pi/2]^{p-1}$, and $\gamma_{i3} \in (-\infty, \infty) \quad \forall\, i = 1, \dots, n$. Define the $n \times (p+q)$ matrix of bootstrap pseudodata $\mathcal{Z}^*$ such that the i^{th} row z^*_i of $\mathcal{Z}^*$ is given by

$$z^*_i = \begin{cases} (\gamma_{i1}, \gamma_{i2}, \gamma_{i3}) & \text{if } \gamma_{i3} \geq 1 \\ (\gamma_{i1}, \gamma_{i2}, 2 - \gamma_{i3}) & \text{otherwise.} \end{cases} \tag{4.20}$$

Thus, values $\gamma_{i3} < 1$ are reflected back across the boundary $\widehat{\delta}_i = 1$, ensuring that $z^*_i \in \mathcal{A} \quad \forall\, i = 1, \dots, n$.

[12] The rescaling factor $(1+h^2)^{-\frac{1}{2}}$ is required for the rows of Γ to have approximately the covariance structure $\widehat{\Sigma}_1$ of the original data in $\mathcal{Z}$; the centering correction is to keep the correct mean of $\tilde{\mathcal{Z}}$.

[9] Translate the polar coordinates in $\mathcal{Z}^*$ to Cartesian coordinates using (4.16)–(4.17); note that this requires solving n linear programs similar to (2.17), as discussed in the previous section and in footnote 10. This yields the bootstrap sample $\mathcal{X}^* = \{(x_i^*, y_i^*),\ i = 1, \ldots, n\}$. For observations where this results in linear programs with infeasible solutions, repeat steps [5]–[8].[13]

[10] For the given point (x, y), compute $\widehat{\delta}^*(x, y)$ by solving the DEA program (4.5).

[11] Repeat steps [5]–[10] B times to obtain a set of bootstrap estimates $\{\widehat{\delta}_b^*(x, y) \mid b = 1, \ldots, B\}$.

Step [11] amounts to a Monte-Carlo simulation. Unfortunately, the computational burden is not negligable, particularly for large n. Step [1] requires that n linear programs to be solved; step [9] requires solution of an additional n linear programs, and step [10] requires solution of a linear program for each point (x, y) being considered. Moreover, steps $[9]-[10]$ are repeated B times. If one were to consider the efficiency of each of the n points in the original sample $\mathcal{X}$, the total number of linear programs for such an application would be $n(2B + 1)$, which is of order $O(nB)$. For very large values of n, however, using the bootstrap to estimate confidence intervals, variance, bias, etc. for every observation might yield an overwhelming amont of information. One might gain more insight by identifying groups of similar observations, perhaps through a cluster analysis, and then bootstrapping from a representative point for each of these groups, perhaps located at the center of each group. Alternatively, the researcher might be interested in the performance of only a few firms within a large sample, in which case the computational costs will be much lower than indicated above.

Once the bootstrap values $\widehat{\delta}_b^*(x, y),\ b = 1, \ldots, B$ have been obtained, we can compute

[13] In some cases, the linear program described in footnote 10 may have infeasible solutions; this will happen when $y_i^* > \max_j(y_j)$. One approach would be to impose an additional boundary condition, namely, $y_i^* \leq \max_j(y_j)$, but this would result in a great increase in complexity and computational burden. We adopt the simpler, innocuous approach of deleting bootstrap values for which $y_i^* > \max_j(y_j)$, and then redrawing. Note that the inequalities here are understood to be element-wise. In our empirical example in Section 5, this problem arises in approximately 2.3 percent of bootstrap draws.

the bootstrap bias estimate for the original estimator $\widehat{\delta}(x,y)$ as

$$\widehat{bias}_B\left[\widehat{\delta}(x,y)\right] = B^{-1}\sum_{b=1}^{B}\widehat{\delta}_b^*(x,y) - \widehat{\delta}(x,y), \tag{4.21}$$

which is merely the empirical bootstrap analog of the definition of bias, *i.e.*, $E\left[\widehat{\delta}(x,y)\right] - \delta(x,y)$. Then a bias-corrected estimator of $\delta(x,y)$ can be computed as

$$\begin{aligned}\widehat{\widehat{\delta}}(x,y) &= \widehat{\delta}(x,y) - \widehat{bias}_B\left[\widehat{\delta}(x,y)\right] \\ &= 2\widehat{\delta}(x,y) - B^{-1}\sum_{b=1}^{B}\widehat{\delta}_b^*(x,y).\end{aligned} \tag{4.22}$$

Of course one should avoid using $\widehat{\widehat{\delta}}(x,y)$ if it has higher mean-square error than $\widehat{\delta}(x,y)$. The variance of the summation term on the right-hand side (r.h.s.) of the second line of (4.22) can be made arbitrarily small by increasing B. Yet, even if $B \to \infty$, the bias-corrected estimator $\widehat{\widehat{\delta}}(x,y)$ will have variance equal to four times that of the original, uncorrected estimator, $\widehat{\delta}(x,y)$. The sample variance of the bootstrap values $\widehat{\delta}_b^*(x,y)$ gives an estimate of the variance of $\widehat{\delta}(x,y)$; call this estimate $\widehat{\sigma}^2$. Then the estimated mean square error of $\widehat{\widehat{\delta}}(x,y)$ is $4\widehat{\sigma}^2$ if $B \to \infty$, and $\left[\widehat{\sigma}^2 + \left(\widehat{bias}_B[\widehat{\delta}(x,y)]\right)^2\right]$ for $\widehat{\delta}(x,y)$. Hence the bias-correction should not be used unless

$$\widehat{\sigma}^2 < \frac{1}{3}\left(\widehat{bias}_B[\widehat{\delta}(x,y)]\right)^2. \tag{4.23}$$

Moreover, since (4.23) contains only estimates of variance and bias rather the true values, caution would indicate that the bias correction in (4.22) be used only if the ratio $\frac{1}{3}\left(\widehat{bias}_B[\widehat{\delta}(x,y)]\right)^2/\widehat{\sigma}^2$ is well above unity.

The bootstrap values $\widehat{\delta}_b^*(x,y)$ can also be used to construct confidence intervals for the true value $\delta(x,y)$, along the lines of Simar and Wilson (1998b). Recall that the idea behind the bootstrap is to approximate the unknown distribution of $\left(\widehat{\delta}(x,y) - \delta(x,y)\right)$ by the distribution of $\left(\widehat{\delta}^*(x,y) - \widehat{\delta}(x,y)\right)$ conditioned on the original data $\mathcal{X}$. The bootstrap values $\widehat{\delta}_b^*(x,y)$, $b = 1,\ldots,B$, together with the original estimate $\widehat{\delta}(x,y)$, can be used to obtain an empirical approximation to the second distribution.

If we knew the distribution of $\left(\widehat{\delta}(x,y)-\delta(x,y)\right)$, then it would be trivial to find values a_α, b_α such that

$$\text{Prob}\left(-b_\alpha \le \widehat{\delta}(x,y)-\delta(x,y) \le -a_\alpha\right) = 1-\alpha \tag{4.24}$$

for some small value of α, say .10 or .05. Since we do not know this distribution, we can use the bootstrap values to find values a^*_α, b^*_α such that the statement

$$\text{Prob}\left(-b^*_\alpha \le \widehat{\delta}^*(x,y)-\widehat{\delta}(x,y) \le -a^*_\alpha \mid \mathcal{X}\right) = 1-\alpha \tag{4.25}$$

is true with high probability.[14] Mechanically, this involves sorting the values

$$\left(\widehat{\delta}^*(x,y)-\widehat{\delta}(x,y)\right), \quad b=1,\ldots,B$$

by algebraic value, deleting $(\frac{\alpha}{2}\times 100)$-percent of the elements at either end of this sorted array, and then setting $-b^*_\alpha$ and $-a^*_\alpha$ equal to the endpoints of the resulting (sorted) array, with $a^*_\alpha \le b^*_\alpha$. When we say that (4.25) is "true with high probability," we mean that this can be made so by making the number of bootstrap replications, B, large enough; as $B\to\infty$, the probability that (4.25) is true approaches 1.[15]

Since

$$\left[\widehat{\delta}(x,y)-\delta(x,y)\right] \mid \mathcal{P} \quad \overset{\text{approx}}{\sim} \quad \left[\widehat{\delta}^*(x,y)-\widehat{\delta}(x,y)\right] \mid \widehat{\mathcal{P}}, \tag{4.26}$$

we obtain the bootstrap approximation

$$\text{Prob}\left(-b^*_\alpha \le \widehat{\delta}(x,y)-\delta(x,y) \le -a^*_\alpha \mid \mathcal{X}\right) \approx 1-\alpha \tag{4.27}$$

by substituting a^*_α and b^*_α for a_α and b_α in (4.24) and noting the conditioning in (4.25). Rearranging the terms in parentheses in (4.27) yields an estimated $(1-\alpha)$-percent confidence

[14] Note we have omitted the subscript b from $\widehat{\delta}^*$ in (4.25), to signify the random variable $\widehat{\delta}^*$, as opposed to one of its realizations, $\widehat{\delta}^*_b$.

[15] Of course, one could also choose $-b^*_\alpha$ and $-a^*_\alpha$ as the $100\times\tau_1$th and $100\times\tau_2$th percentiles of the sorted bootstrap differences, where $\tau_1 = \Phi(2\widehat{w}_0 + w^{(\alpha/2)})$, $\tau_2 = \Phi(2\widehat{w}_0 + w^{(1-\alpha/2)})$, $\widehat{w}_0 = \Phi^{-1}\left(\frac{\#\{\widehat{\delta}^*_b(x,y)<\widehat{\delta}(x,y)\}}{B}\right)$, $\Phi(\cdot)$ is the standard normal distribution function, and $w^{(\alpha)}$ is the $100\times\alpha$th percentile of the standard normal distribution so that $\Phi(w^{(\alpha)})=\alpha$. This procedure involves a median-bias correction, and was used in Simar and Wilson (1998a), and is discussed in Efron and Tibshirani (1993).

interval[16]

$$\widehat{\delta}(x,y) + a^*_\alpha \;\le\; \delta(x,y) \;\le\; \widehat{\delta}(x,y) + b^*_\alpha. \tag{4.28}$$

As a final note, we remark that by construction, $\delta(x,y) \ge \widehat{\delta}(x,y) \ge 1$. Similarly, we have $\widehat{\delta}(x,y) \ge \widehat{\delta}^*_b(x,y) \quad \forall\ b = 1,\ldots,B$. Necessarily then $0 \le a^*_\alpha \le b^*_\alpha$, and so the estimated confidence interval in (4.28) will only include the original estimate $\widehat{\delta}(x,y)$ on its lower boundary if $a^*_\alpha = 0$; more typically, $\widehat{\delta}(x,y)$ will fall outside the estimated confidence interval. This indeed should not be surprising, since $\widehat{\delta}(x,y)$ is necesarily a downward-biased estimator of $\delta(x,y)$. This result merely indicates that $\widehat{\delta}(x,y)$ and the bootstrap values $\widehat{\delta}^*_b(x,y)$ provide enough information about the true value $\delta(x,y)$ to suggest that $\delta(x,y)$ lies in some range above $\widehat{\delta}(x,y)$. In any case, given an estimated confidence interval of size $1-\alpha$ as in (4.28), it is not clear why the applied researcher would still be concerned with a point estimate.

In the next section, we provide an empirical exemple to illustrate the algorithm proposed above.

5. Empirical Example

As an illustration of our methods, we examine data reported by Charnes *et al.* (1981) on Program Follow Through (PFT), an experimental education program administered in US schools. In particular, Charnes *et al.* displayed data on five inputs and three outputs for 49 schools that implemented PFT, and 21 schools that did not, for a total of 70 observations.[17]

Step [4] in our bootstrap algorithm involves choosing the bandwidth parameter h as described in the Appendix. For the Charnes *et al.* data, the normal reference rule in (4.15) gives a bandwidth of 0.65025. Using the least-squares cross-validation procedure described

[16] In Simar and Wilson (1998a), we constructed confidence intervals for distance functions using an estimate of bias analagous to (4.21). The approach in this paper avoids introducing the extra noise contained in this estimate. The bias inherent in the distance function estimates is implicitly accounted for here since we use the bootstrap values to construct an empirical distribution of *differences* as in (4.26).

[17] See Charnes *et al.* for the actual data and a detailed discussion, including definitions of the input and output variables. Charnes *et al.* indicate that schools with similar circumstances were chosen for their data to allow comparison of those implementing PFT and those not doing so.

in the Appendix, we chose a bandwidth of 0.87946, which we used in our bootstrap estimation. The cross-validation procedure required just over 34 minutes elapsed time on a SUN Sparcstation 20.

Table 1 shows our original distance function estimates, $\widehat{\delta}(x_i, y_i)$, the corresponding bias estimates, and the estimated standard deviations across bootstrap replications for each observation ($\widehat{\sigma}_i$). The fifth column of the table gives the ratio $r_i = \frac{1}{3}\left(\widehat{bias}_B[\widehat{\delta}(x_i, y_i)]\right)^2/\widehat{\sigma}^2$ discussed in the Section 4, which may be used to assess whether the bias correction might increase mean-square error. The sixth column shows the bias-corrected distance function estimate, while the last two columns give the estimated 95 percent confidence bounds. Results were produced using $B = 2000$ bootstrap replications using Fortran code written by the authors, requiring approximately 47 minutes elapsed computation time on a SUN Sparcstation 20.

Bootstrap estimates could not be determined for observations 44 and 59. These observations were identified as outliers in Wilson (1993); indeed, on every bootstrap replication, these observations lie above the bootstrap frontier, making solutions to the linear program (4.5) infeasible.

Of 70 observations in the data, 37 appear ostensibly efficient as indicated by distance function estimates $\widehat{\delta}(x_i, y_i) = 1$. The remaining 33 observations have distance function estimates ranging from 1.0037 to 1.2447. If one were to rely on only these results, one might be tempted to conclude, as did Charnes *et al.*, that PFT schools were no different from non-PFT schools. Indeed, the average distance function estimate over the PFT schools (observations 1–49) is 1.0384, while the average estimate over the non-PFT schools (observations 50–70) is 1.0300, indicating only a negligible difference.

However, considering the stochastic nature of the estimation problem via our bootstrap method suggests a different picture of PFT. The estimated biases are negative, as expected, and in numerous cases quite large. The estimated variances, on the other hand, are frequently quite small, so that the ratios r_i frequently exceed unity, in many cases by a great deal. Thus, for example, observation 1, which appears ostensibly efficient if one views only the original distance function estimate, has a bias-corrected distance function estimate of 1.3682, suggesting that the same outputs could have been produced while scaling inputs back by more than 36 percent. The estimated 95 percent confidence interval for this observations suggests that inputs could have been reduced by between 29 percent and 59 percent.

Comparing the PFT and the non-PFT observations provides further insights. The PFT observations have an average bias-corrected distance function estimate of 1.2610, while the non-PFT schools have a corresponding average of 1.3104. Similarly, the estimated lower confidence bound averages 1.1999 for the PFT schools, and 1.2550 for the non-PFT schools. Perhaps more revealing, 20 of the PFT schools (40.8 percent) have estimated upper confidence bounds less than 1.3, while 7 of the non-PFT schools (24.1 percent) have estimated *lower* confidence bounds *greater* than 1.3. Finally, note that three of the PFT schools (observations 8, 16, and 23) have estimated confidence intervals that include unity, while none of the non-PFT schools have estimated lower confidence bounds less than 1.1. Thus, while we have not formally tested whether PFT was effective, our results are suggestive, and we would at this point be reluctant to draw the same conclusions as Charnes *et al.* did.

6. Conclusions

As noted in the Introduction, DEA methods have been widely applied to examine efficiency in a variety of production settings. Yet, despite a small but growing literature on the statistical properties of DEA estimators, most researchers have used these methods while ignoring the sampling noise in the resulting efficiency estimators, and continue to do so (*e.g.,* for a very recent example, see Ray and Desli, 1997). As our empirical example demonstrates, ignoring the statistical properties of DEA estimators and the uncertainty surrounding DEA estimates can lead to erroneous conclusions.

We have provided a general, computatonally tractible method for adapting bootstrap methods to the problem of nonparametric efficiency estimation. While it is true that with samples containing many hundreds or thousands of observations it will likely be infeasible to obtain bootstrap estimated confidence intervals for each observation as we have done in our empirical example, it is not clear why one would want to do so when the resulting volume of information would overwhelm any mortal practitioner. Rather, our method can be used to assess uncertainty about distance to the true production frontier from a relatively small number of points in the production set, cleverly chosen to reflect the location of most of the data, or of at least the most interesting part of the data.

Appendix: Choice of Bandwith

Least-squares cross-validation (e.g. see Silverman, 1986) provides a data-driven criterion for choosing the bandwidth h in kernel density estimation. As Silverman (1986) demonstrates, the usual cross-validation function gives an approximation to MISE; the goal is to choose h to minimize this approximation. As noted previously, the usual cross-validation function suffers from degenerate behavior when data have been discretized, allowing the cross-validation function to approach $-\infty$ as $h \to 0$.

We deal with this problem by minimizing an approximation to the mean weighted intergrated square error (MWISE) (see e.g. Härdle, 1990). The MWISE is defined by

$$\begin{aligned} E&\left[\int w(z)(f(z)-\widehat{f}_h(z))^2 dz\right] \\ &= E\left[\int w(z)\widehat{f}_h^2(z)dz - 2\int w(z)f(z)\widehat{f}_h(z)dz + \int w(z)f^2(z)dz\right], \end{aligned} \tag{A.1}$$

where $w(z)$ is some predefined weight function.

The last term of (A.1) does not depend on h, so the optimal choice of the bandwith (in the sense of minimizing MWISE) minimizes the following criterion :

$$C(h) = E\left[\int w(z)\widehat{f}_h^2(z)dz - 2\int w(z)f(z)\widehat{f}_h(z)dz\right]. \tag{A.2}$$

The idea of cross validation is to construct an estimate of $C(h)$ from the data $\{z_i,\ i = 1, \ldots, n\}$. Following Silverman, an unbiased estimator of $C(h)$ is given by

$$CV(h) = \int w(z)\widehat{f}_h^2(z)dz - \frac{2}{n}\sum_{i=1}^{n} w(z_i)\widehat{f}_{h,-i}(z_i), \tag{A.3}$$

where $\widehat{f}_{h,-i}(z_i)$ is the leave-one out estimator of $f(z_i)$ based on all the original observations except z_i, with bandwith h.

In order to avoid the discretization problem, we define the weight function $w(z)$ as

$$w(z) = \begin{cases} 0 & \text{if } \delta \in [1, 1+\epsilon] \\ 1 & \text{otherwise,} \end{cases} \tag{A.4}$$

where ϵ is small, say 10^{-6}. By choosing $\epsilon > 0$ but very small, we avoid the discretization problem by preventing observations where $\widehat{\delta}_i = 1$ from influencing the $CV(h)$ in (A.3). The value of ϵ should be chosen small enough so that only observations where $\widehat{\delta}_i = 1$ are eliminated.

In minimizing $CV(h)$ with respect to h, clearly there is no problem in computing the second term on the r.h.s. of (A.3), but the first term is more difficult. We can write this first term as

$$\int w(z)\widehat{f}_h^2(z)\,dz = \int_{z\in\mathcal{A}} \widehat{f}_h^2(z)\,dz - \int_{z\in\mathcal{A}_\epsilon} \widehat{f}_h^2(z)\,dz \tag{A.5}$$

where $\mathcal{A}_\epsilon = \mathbb{R}_+^q \times [0,\pi/2]^{p-1} \times [1, 1+\epsilon]$.

Consider the **first term** of the r.h.s. of (A.5):

$$\begin{aligned}
\int_{z\in\mathcal{A}} \widehat{f}_h^2(z)\,dz = &\int_{z\in\mathcal{A}} \frac{1}{nh^{(p+q)}} \sum_{i=1}^n \left[K_1\left(\frac{z-z_i}{h}\right) + K_2\left(\frac{z-z_{Ri}}{h}\right)\right] \\
&\times \frac{1}{nh^{(p+q)}} \sum_{j=1}^n \left[K_1\left(\frac{z-z_j}{h}\right) + K_2\left(\frac{z-z_{Rj}}{h}\right)\right] dz \\
=&\frac{1}{n^2h^{2(p+q)}} \sum_{i=1}^n\sum_{j=1}^n \int_{z\in\mathcal{A}} \\
&\left[K_1\left(\frac{z-z_i}{h}\right) K_1\left(\frac{z-z_j}{h}\right) + K_1\left(\frac{z-z_i}{h}\right) K_2\left(\frac{z-z_{Rj}}{h}\right)\right. \\
&\left. + K_2\left(\frac{z-z_{Ri}}{h}\right) K_1\left(\frac{z-z_j}{h}\right) + K_2\left(\frac{z-z_{Ri}}{h}\right) K_2\left(\frac{z-z_{Rj}}{h}\right)\right] dz \\
=&\frac{1}{n^2h^{2(p+q)-1}} \sum_{i=1}^n\sum_{j=1}^n \int_{hu\in\mathcal{A}} \Big[K_1(h^{-1}z_i - u)K_1(u - h^{-1}z_j) \\
&+ K_1(h^{-1}z_i - u)K_2(u - h^{-1}z_{Rj}) + K_2(h^{-1}z_{Ri} - u)K_1(u - h^{-1}z_j) \\
&+ K_2(h^{-1}z_{Ri} - u)K_2(u - h^{-1}z_{Rj})\Big]\,du \\
=&\frac{1}{n^2h^{2(p+q)-1}} \sum_{i=1}^n\sum_{j=1}^n \left[K_{11}\left(\frac{z_i-z_j}{h}\right)\right. \\
&\left. + 2K_{12}\left(\frac{z_i-z_{Rj}}{h}\right) + K_{22}\left(\frac{z_{Ri}-z_{Rj}}{h}\right)\right].
\end{aligned} \tag{A.6}$$

where $u = h^{-1}z$ (implying $dz = h\,du$), $K_{11}(\cdot)$ is the convolution of $K_1(\cdot)$ with itself and hence is $N\left(0, 2\widehat{\Sigma}_1\right)$, $K_{12}(\cdot)$ is the convolution of $K_1(\cdot)$ with $K_2(\cdot)$ and hence is $N\left(0, \widehat{\Sigma}_1 + \widehat{\Sigma}_2\right)$ and $K_{22}(\cdot)$ is the convolution of $K_2(\cdot)$ with itself and hence is $N\left(0, 2\widehat{\Sigma}_2\right)$.

Consider now the **second term** on the r.h.s. of (A.5). Let $\beta = (y, \eta)$, $\mathcal{B} = \mathbb{R}_+^q \times \left[0, \frac{\pi}{2}\right]^{p-1}$, and $\mathcal{C} = [1, 1+\epsilon]$. Then

$$\begin{aligned} \int_{z \in \mathcal{A}_\epsilon} \widehat{f}_h^2(z)\, dz &= \int_{\delta \in \mathcal{C}} \int_{\beta \in \mathcal{B}} \widehat{f}_h^2(\beta|\delta) \widehat{f}_h^2(\delta)\, d\beta\, d\delta \\ &= \int_{\delta \in \mathcal{C}} \left[\int_{\beta \in \mathcal{B}} \widehat{f}_h^2(\beta|\delta)\, d\beta \right] \widehat{f}_h^2(\delta)\, d\delta \end{aligned} \tag{A.7}$$

Now consider the brackened term in the last part of (A.7). Partition z_i and z_{Ri} so that $z_i = (\beta_i, \widehat{\delta}_i)$ and $z_{Ri} = (\beta_i, 2 - \widehat{\delta}_i)$. Then conditionning on δ in our definition of $\widehat{f}_h(z)$ in (4.14) yields

$$\widehat{f}_h(\beta|\delta) = \frac{1}{nh^{(p+q)}} \sum_{i=1}^n \left[K_1\left(\frac{\beta - \beta_i}{h} \Big| \frac{\delta - \widehat{\delta}_i}{h} \right) + K_2\left(\frac{\beta - \beta_i}{h} \Big| \frac{\delta - (2 - \widehat{\delta}_i)}{h} \right) \right]. \tag{A.8}$$

Squaring both sides of this yields

$$\begin{aligned} \widehat{f}_h^2(\beta|\delta) = & \frac{1}{n^2 h^{2(p+q)}} \sum_{i=1}^n \sum_{j=1}^n \left[K_1\left(\frac{\beta - \beta_i}{h} \Big| \frac{\delta - \widehat{\delta}_i}{h} \right) K_1\left(\frac{\beta - \beta_j}{h} \Big| \frac{\delta - \widehat{\delta}_j}{h} \right) \right. \\ & + 2K_1\left(\frac{\beta - \beta_i}{h} \Big| \frac{\delta - \widehat{\delta}_i}{h} \right) K_2\left(\frac{\beta - \beta_j}{h} \Big| \frac{\delta - (2 - \widehat{\delta}_j)}{h} \right) \\ & \left. + K_2\left(\frac{\beta - \beta_i}{h} \Big| \frac{\delta - (2 - \widehat{\delta}_i)}{h} \right) K_2\left(\frac{\beta - \beta_j}{h} \Big| \frac{\delta - (2 - \widehat{\delta}_j)}{h} \right) \right]. \end{aligned} \tag{A.9}$$

We must integrate both sides of this expression to obtain the brackened term in (A.7). Since $K_1(\cdot)$ is multivariate $N\left(0, \widehat{\Sigma}_1\right)$, $K_1\left(\frac{\beta - \beta_i}{h} | \frac{\delta - \widehat{\delta}_i}{h}\right)$ must be

$$N\left(S_{12} S_{22}^{-1} \frac{\delta - \widehat{\delta}_i}{h}, S_{11} - S_{12} S_{22}^{-1} S_{21} \right).$$

Similarly, $K_2(\cdot)$ is multivariate $N\left(0, \widehat{\Sigma}_2\right)$, and hence $K_2\left(\frac{\beta - \beta_i}{h} | \frac{\delta - (2 - \widehat{\delta}_i)}{h}\right)$ must be $N\left(-S_{12} S_{22}^{-1} \frac{\delta - (2 - \widehat{\delta}_i)}{h}, S_{11} - S_{12} S_{22}^{-1} S_{21}\right)$.

Let $S_{11.2} = S_{11} - S_{12}S_{22}^{-1}S_{21}$, $\mu_{1i} = S_{12}S_{22}^{-1}\frac{\delta-\widehat{\delta}_i}{h}$ and $\mu_{2i} = -S_{12}S_{22}^{-1}\frac{\delta-(2-\widehat{\delta}_i)}{h}$. Then we can rewrite $K_1\left(\frac{\beta-\beta_i}{h}|\frac{\delta-\widehat{\delta}_i}{h}\right)$ as $\phi_{1i}\left(\frac{\beta-\beta_i}{h}\right)$, where $\phi_{1i}(\cdot)$ is $N\left(\mu_{1i}, S_{11.2}\right)$. Similarly, $K_2\left(\frac{\beta-\beta_i}{h}|\frac{\delta-(2-\widehat{\delta}_i)}{h}\right)$ can be rewritten as $\phi_{2i}\left(\frac{\beta-\beta_i}{h}\right)$, where $\phi_{2i}(\cdot)$ is $N\left(\mu_{2i}, S_{11.2}\right)$. Then, integrating (A.9), we have

$$\begin{aligned}\int_{\beta\in\mathcal{B}} \widehat{f}_h^2(\beta|\delta)\,d\beta =& \frac{1}{n^2h^{2(p+q)}}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_{\beta\in\mathcal{B}}\left[\phi_{1i}\left(\frac{\beta-\beta_i}{h}\right)\phi_{1j}\left(\frac{\beta-\beta_j}{h}\right)\right.\\ &+ 2\phi_{1i}\left(\frac{\beta-\beta_i}{h}\right)\phi_{2j}\left(\frac{\beta-\beta_j}{h}\right)\\ &\left.+ \phi_{2i}\left(\frac{\beta-\beta_i}{h}\right)\phi_{2j}\left(\frac{\beta-\beta_j}{h}\right)\right]d\beta.\end{aligned} \tag{A.10}$$

Using the same convolution argument as in (A.6), we obtain the brackened term of (A.7):

$$\begin{aligned}&\int_{\beta\in\mathcal{B}} \widehat{f}_h^2(\beta|\delta)d\beta = \\ &\qquad\frac{1}{n^2h^{2(p+q)-1}}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\phi_{11ij}\left(\frac{\beta_i-\beta_j}{h}\right) + 2\phi_{12ij}\left(\frac{\beta_i-\beta_j}{h}\right) + \phi_{22ij}\left(\frac{\beta_i-\beta_j}{h}\right)\right],\end{aligned} \tag{A.11}$$

where for $k, \ell = 1, 2$, $\phi_{k\ell ij}(\cdot)$ is the convolution of $\phi_{ki}(\cdot)$ and $\phi_{\ell j}(\cdot)$. Hence ϕ_{11ij} is $N\left(\mu_{1i}+\mu_{1j}, 2S_{11.2}\right)$, ϕ_{12ij} is $N\left(\mu_{1i}+\mu_{2j}, 2S_{11.2}\right)$, and ϕ_{22ij} is $N\left(\mu_{2i}+\mu_{2j}, 2S_{11.2}\right)$.

Substituting (A.11) into (A.7), we have for the second term on the r.h.s. of (A.5),

$$\begin{aligned}\int_{z\in\mathcal{A}_\epsilon} \widehat{f}_h^2(z)\,dz =& \frac{1}{n^2h^{2(p+q)-1}}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_{\delta\in\mathcal{C}}\left[\phi_{11ij}\left(\frac{\beta_i-\beta_j}{h}\right)\right.\\ &\left.+ 2\phi_{12ij}\left(\frac{\beta_i-\beta_j}{h}\right) + \phi_{22ij}\left(\frac{\beta_i-\beta_j}{h}\right)\right]\widehat{f}_h^2(\delta)\,d\delta\end{aligned} \tag{A.12}$$

We now need $\widehat{f}_h^2(\delta)$. Note that from the definition of our multivariate density estimator (4.14), and the fact that the marginal of a multivariate normal is also normal, we have

$$\widehat{f}_h(\delta) = \begin{cases} \frac{1}{nh}\sum_{i=1}^{n}\left[K_*\left(\frac{\delta-\widehat{\delta}_i}{h}\right) + K_*\left(\frac{\delta-(2-\widehat{\delta}_i)}{h}\right)\right] & \text{if } \delta \geq 1 \\ 0 & \text{otherwise.}\end{cases} \tag{A.13}$$

where $K_*(\cdot)$ is the univariate normal density with mean 0 and variance S_{22}.

Squaring both sides of (A.13) yields

$$\widehat{f}_h^2(\delta) = \frac{1}{n^2h^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\left[K_*\left(\frac{\delta-\widehat{\delta}_k}{h}\right)K_*\left(\frac{\delta-\widehat{\delta}_l}{h}\right)+2K_*\left(\frac{\delta-\widehat{\delta}_k}{h}\right)K_*\left(\frac{\delta-(2-\widehat{\delta}_l)}{h}\right)\right.$$
$$\left.+K_*\left(\frac{\delta-(2-\widehat{\delta}_k)}{h}\right)K_*\left(\frac{\delta-(2-\widehat{\delta}_l)}{h}\right)\right]. \tag{A.14}$$

Finally, substituting (A.14) into (A.12) yields a form for the second term on the r.h.s. of (A.5) that we can compute:

$$\int_{z\in\mathcal{A}_\epsilon}\widehat{f}_h^2(z)\,dz = \frac{1}{n^4h^{2(p+q)+1}}\int_{\delta\in\mathcal{C}}$$
$$\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\phi_{11ij}\left(\frac{\beta_i-\beta_j}{h}\right)+2\phi_{12ij}\left(\frac{\beta_i-\beta_j}{h}\right)+\phi_{22ij}\left(\frac{\beta_i-\beta_j}{h}\right)\right]\right\}\times$$
$$\left\{\sum_{k=1}^{n}\sum_{l=1}^{n}\left[K_*\left(\frac{\delta-\widehat{\delta}_k}{h}\right)K_*\left(\frac{\delta-\widehat{\delta}_l}{h}\right)+2K_*\left(\frac{\delta-\widehat{\delta}_k}{h}\right)K_*\left(\frac{\delta-(2-\widehat{\delta}_l)}{h}\right)\right.\right.$$
$$\left.\left.+K_*\left(\frac{\delta-(2-\widehat{\delta}_k)}{h}\right)K_*\left(\frac{\delta-(2-\widehat{\delta}_l)}{h}\right)\right]\right\}d\delta. \tag{A.15}$$

This can be solved by any one of several one-dimensional numerical intergration techniques on δ over $\mathcal{C} = [1, 1+\epsilon]$. We use ten-point Gauss-Legendre quadrature in our empirical example, as discussed in section 5. The computation cost in solving (A.15) should not be too great, since the integral is over a short interval in one dimension. In addition, the integral must be evaluated only once each time $CV(h)$ is evaluated. Note that ϕ_{11ij}, ϕ_{12ij} and ϕ_{22ij} depend on δ through their mean terms, and so remain under the intergral sign.

Finally, the second term in $CV(h)$ in (A.3) can be written

$$\frac{2}{n}\sum_{i=1}^{n}w(z_i)\widehat{f}_{h,-i}(z_i) = \frac{2}{n}\sum_{i=1}^{n}w(z_i)\frac{1}{(n-1)h^{(p+q)}}\sum_{\substack{j=1\\j\neq i}}^{n}\left[K_1\left(\frac{z_i-z_j}{h}\right)+K_2\left(\frac{z_i-z_{Rj}}{h}\right)\right]$$
$$= \frac{2}{n(n-1)h^{(p+q)}}\sum_{i=1}^{n}w(z_i)\sum_{\substack{j=1\\j\neq i}}^{n}\left[K_1\left(\frac{z_i-z_j}{h}\right)+K_2\left(\frac{z_i-z_{Rj}}{h}\right)\right], \tag{A.16}$$

which involves straight-forward computations.

Substituting (A.6) and (A.15) into (A.5), and then substituting the resulting expression together with (A.16) into (A.3) yields an expression for $CV(h)$ that can be computed numerically. Of course, the resulting expression of $CV(h)$ will be rather complicated; consequently, minimization with respect to h will typically require a grid search over a range of values $h \in [h_{\min}, h_{\max}]$. The range of values could be determined by taking factors of, say 0.25 and 2.0, of the bandwidth suggested by the normal reference rule in (4.15). Note that due to the nature of kernel functions, the expression comprising $CV(h)$ will be easier to compute for small values of h than for large values. In a similar problem, Fan and Gijbels (1996) suggest evaluating $CV(h = h_{\min})$, then successively increasing h by small increments, evaluating $CV(h)$ each time. The process terminated after some number of successive increases in $CV(h)$ have been observed, and the value of h that yielded the smallest value for $CV(h)$ is then chosen as the optimum.

REFERENCES

Banker, R.D. (1993), Maximum Likelihood, consistency and data envelopment analysis: a statistical foundation, *Management Science*, 39,10,1265-1273.

Beran, R., and G. Ducharme (1991), *Asymptotic Theory for Bootstrap Methods in Statistics,* Montreal: Centre de Reserches Mathematiques, University of Montreal.

Bickel, P. J., and D. A. Freedman (1981), Some asymptotic theory for the bootstrap, *Annals of Statistics* 9, 1196–1217.

Campbell, N.A. (1980), Robust procedures in multivariate analysis I: Robust covariance estimation, *Applied Statistics* 29, 231-237.

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

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

Charnes, A., W. W. Cooper, and E. Rhodes (1981), Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through, *Management Science* 27, 668–697.

Debreu, G. (1951), The coefficient of resource utilization, *Econometrica* 19, 273–292.

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.

Efron, B. (1979), Bootstrap methods: another look at the jackknife, *Annals of Statistics* 7, 1–16.

Efron, B. (1982), *The Jackknife, the Bootstrap, and Other Resampling Plans,* DBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38, Society for Industrial and Applied Mathematics, Philadelphia.

Efron, B., and R. J. Tibshirani (1993), *An Introduction to the Bootstrap,* New York: Chapman and Hall.

Fan, J. and I. Gijbels (1996), *Local Polynomial Modelling and its Applications,* New York: Chapman and Hall.

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

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

Gijbels, I., E. Mammen, B.U. Park and L. Simar (1996), "On Estimation of Monotone and Concave Frontier Functions," Discussion paper #9611, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Härdle, W. (1990), *Applied Nonparametric Regression,* Cambridge: Cambridge University Press.

Hampel, F.R. , E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel (1986), *Robust Statistics: The Approach Based on Influence Functions,* New York: John Wiley.

Kneip, A., B. Park and L. Simar (1998), A note on the convergence of nonparametric DEA efficiency measures, *Econometric Theory,* forthcoming.

Kneip, A. and L. Simar (1996), A general framework for frontier estimation with panel data, *Journal of Productivity Analysis*, 7, 2/3, 187-212.

Korostelev, A., Simar, L. and A. Tsybakov (1995a), Efficient estimation of monotone boundaries, *The Annals of Statistics* 23, 476-489.

Korostelev, A., Simar, L. and A. Tsybakov (1995b), On estimation of monotone and convex boundaries, *Publications de l'Institut de Statistique de l'Université de Paris XXXIX* 1, 3-18.

Lovell, C. A. K. (1993), "Production Frontiers and Productive Efficiency," in *The Measurement of Productive Efficiency: Techniques and Applications,* ed. by Hal Fried, C. A. Knox Lovell, and Shelton S. Schmidt, Oxford University Press, Inc., Oxford, pp. 3–67.

Ray, S. and E. Desli (1997), Productivity growth, technical progress, and efficiency change in industrialized countries: Comment, *American Economic Review* 87, 1033–1039.

Rousseeuw, P.J. (1985), "Multivariate Estimation with High Breakdown Point," in *Mathematical Statistics and Applications, Vol B* (ed. by W. Grossman, G. Plug, I. Vincze, and W. Wertz), Dordrecht: Reidel Publishing, pp. 283-297.

Scott, D.W. (1992), *Multivariate Density Estimation.* John Wiley and Sons, Inc., New-York.

Seiford, L.M. (1996), Data envelopment analysis: The evolution of the state-of-the-art (1978–1995),*Journal of Productivity Analysis*, 7, 2/3, 99-138 .

Seiford, L. M. (1997), A bibliography for data envelopment analysis (1978–1996), *Annals of Operations Research* 73, 393–438.

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

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

Simar, L. (1996), Aspects of statistical analysis in DEA-type frontier models, *Journal of Productivity Analysis* 7, 177-185.

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

Simar,L. and P. Wilson (1998b), Estimating and bootstrapping Malmquist indices, *European Journal of Operational Research,* forthcoming.

Wilson, P. W. (1993), Detecting outliers in deterministic nonparametric frontier models with multiple outputs, *Journal of Business and Economic Statistics* 11, 319–323.

TABLE 1

Estimates from Charnes *et al.* (1981) Data

i	$\widehat{\delta}(x_i,y_i)$	$\widehat{bias}_B\left(\widehat{\delta}(x_i,y_i)\right)$	$\widehat{\sigma}$	r_i	$\widehat{\widehat{\delta}}(x_i,y_i)$	95% conf. int.	
1	1.0000	−0.3682	0.08596	6.1160	1.3682	1.2959	1.5947
2	1.0571	−0.0972	0.03040	3.4030	1.1543	1.0992	1.2176
3	1.0000	−0.1331	0.01476	27.1100	1.1331	1.1109	1.1686
4	1.0719	−0.0485	0.01613	3.0170	1.1204	1.0910	1.1541
5	1.0000	−0.4667	0.04273	39.7600	1.4667	1.3980	1.5637
6	1.0464	−0.1882	0.02865	14.3800	1.2346	1.1827	1.2916
7	1.0793	−0.0422	0.01345	3.2800	1.1215	1.0971	1.1495
8	1.0000	−0.1023	0.09845	0.3601	1.1023	0.9647	1.3512
9	1.1019	−0.0244	0.01512	0.8649	1.1263	1.0999	1.1585
10	1.0000	−0.2249	0.05197	6.2400	1.2249	1.1770	1.3648
11	1.0000	−0.2281	0.04076	10.4400	1.2281	1.1617	1.3183
12	1.0000	−0.2573	0.05528	7.2200	1.2573	1.1577	1.3833
13	1.1160	−0.0278	0.01331	1.4560	1.1438	1.1212	1.1737
14	1.0000	−0.3458	0.03345	35.6200	1.3458	1.2865	1.4176
15	1.0000	−0.5260	0.06330	23.0200	1.5260	1.3946	1.6391
16	1.0000	−0.1326	0.13340	0.3290	1.1326	0.9844	1.4770
17	1.0000	−0.5113	0.01693	303.9000	1.5113	1.5011	1.5347
18	1.0000	−0.2028	0.03513	11.1100	1.2028	1.1377	1.2728
19	1.0000	−0.0900	0.06258	0.6902	1.0900	1.0146	1.2496
20	1.0000	−0.3042	0.05086	11.9200	1.3042	1.2033	1.3986
21	1.0000	−0.2982	0.07396	5.4190	1.2982	1.1885	1.4811
22	1.0000	−0.1832	0.02495	17.9700	1.1832	1.1323	1.2317
23	1.0000	−0.0692	0.08493	0.2210	1.0692	1.0000	1.3186
24	1.0000	−0.3326	0.03063	39.3200	1.3326	1.2740	1.3927
25	1.0123	−0.0541	0.01088	8.2470	1.0664	1.0472	1.0888
26	1.0567	−0.0730	0.02136	3.8940	1.1297	1.0968	1.1744
27	1.0000	−0.2359	0.03695	13.5900	1.2359	1.1664	1.3097
28	1.0000	−0.2847	0.02152	58.3400	1.2847	1.2534	1.3342
29	1.0718	−0.3338	0.06020	10.2500	1.4056	1.2969	1.5373
30	1.0801	−0.1013	0.03264	3.2080	1.1814	1.1269	1.2564
31	1.1560	−0.0803	0.02326	3.9730	1.2363	1.1938	1.2871
32	1.0000	−0.3986	0.06210	13.7300	1.3986	1.2748	1.5274
33	1.0440	−0.1211	0.09716	0.5178	1.1651	1.0776	1.4244
34	1.1570	−0.0683	0.01981	3.9680	1.2253	1.1952	1.2707
35	1.0000	−0.2047	0.07134	2.7450	1.2047	1.1304	1.3863

TABLE 1 (continued)

i	$\widehat{\delta}(x_i, y_i)$	$\widehat{bias}_B\left(\widehat{\delta}(x_i, y_i)\right)$	$\widehat{\sigma}$	r_i	$\widehat{\widehat{\delta}}(x_i, y_i)$	95% conf. int.	
36	1.2447	−0.2954	0.02241	57.9200	1.5401	1.5007	1.5869
37	1.1505	−0.2261	0.02341	31.1000	1.3766	1.3332	1.4243
38	1.0000	−0.5085	0.11970	6.0120	1.5085	1.5011	1.5303
39	1.0595	−0.1692	0.02942	11.0200	1.2287	1.1778	1.2932
40	1.0374	−0.1752	0.01736	33.9700	1.2126	1.1782	1.2444
41	1.0500	−0.0502	0.01008	8.2590	1.1002	1.0820	1.1219
42	1.0491	−0.3375	0.02402	65.7900	1.3866	1.3519	1.4424
43	1.1426	−0.1252	0.02976	5.9050	1.2678	1.2151	1.3329
44	1.0000	—	—	—	—	—	—
45	1.0000	−0.5108	0.00723	1662.0000	1.5108	1.5013	1.5287
46	1.0954	−0.0711	0.02836	2.0960	1.1665	1.1342	1.2257
47	1.0000	−0.2345	0.02478	29.8500	1.2345	1.1943	1.2918
48	1.0000	−0.5242	0.08405	12.9700	1.5242	1.3430	1.6700
49	1.0000	−0.2593	0.06459	5.3720	1.2593	1.1485	1.3968
50	1.0414	−0.2707	0.01963	63.3600	1.3121	1.2790	1.3555
51	1.0454	−0.2070	0.04253	7.8960	1.2524	1.1703	1.3419
52	1.0000	−0.3571	0.04373	22.2200	1.3571	1.3095	1.4628
53	1.1488	−0.1169	0.02165	9.7190	1.2657	1.2276	1.3119
54	1.0000	−0.4995	0.09610	9.0070	1.4995	1.4208	1.7678
55	1.0000	−0.2078	0.01653	52.6800	1.2078	1.1792	1.2450
56	1.0000	−0.4805	0.07156	15.0300	1.4805	1.3366	1.6082
57	1.0671	−0.2321	0.01516	78.1400	1.2992	1.2705	1.3287
58	1.0000	−0.4151	0.02163	122.8000	1.4151	1.3741	1.4588
59	1.0000	—	—	—	—	—	—
60	1.0199	−0.1316	0.02220	11.7200	1.1515	1.1102	1.1977
61	1.0486	−0.3641	0.05708	13.5600	1.4127	1.3125	1.5386
62	1.0000	−0.5648	0.02408	183.4000	1.5648	1.5312	1.6231
63	1.0179	−0.2247	0.02919	19.7400	1.2426	1.1899	1.3030
64	1.0749	−0.0753	0.01160	14.0500	1.1502	1.1302	1.1745
65	1.0037	−0.2360	0.02823	23.2900	1.2397	1.1915	1.3002
66	1.0687	−0.0747	0.01146	14.1500	1.1434	1.1236	1.1683
67	1.0568	−0.0760	0.01350	10.5700	1.1328	1.1093	1.1611
68	1.0000	−0.3042	0.07951	4.8780	1.3042	1.1651	1.4781
69	1.0000	−0.5739	0.03143	111.1000	1.5739	1.5226	1.6441
70	1.0373	−0.1647	0.03244	8.5870	1.2020	1.1453	1.2724