Journal of Productivity Analysis, 13, 49–78 (2000)

Statistical Inference in Nonparametric Frontier Models: The State of the Art

LÉOPOLD SIMAR* simar@stat.ucl.ac.be, http://www.stat.ucl.ac.be/ispersonnel/simar.html
Institut de Statistique, Université Catholique de Louvain, Louvain-La-Neuve, Belgium

PAUL W. WILSON wilson@eco.utexas.edu, http://www.eco.utexas.edu/faculty/wilson
Department of Economics, University of Texas, Austin, Texas 78712, USA

Abstract

Efficiency scores of firms are measured by their distance to an estimated production frontier. The economic literature proposes several nonparametric frontier estimators based on the idea of enveloping the data (FDH and DEA-type estimators). Many have claimed that FDH and DEA techniques are non-statistical, as opposed to econometric approaches where particular parametric expressions are posited to model the frontier. We can now define a statistical model allowing determination of the statistical properties of the nonparametric estimators in the multi-output and multi-input case. New results provide the asymptotic sampling distribution of the FDH estimator in a multivariate setting and of the DEA estimator in the bivariate case. Sampling distributions may also be approximated by bootstrap distributions in very general situations. Consequently, statistical inference based on DEA/FDH-type estimators is now possible. These techniques allow correction for the bias of the efficiency estimators and estimation of confidence intervals for the efficiency measures. This paper summarizes the results which are now available, and provides a brief guide to the existing literature. Emphasizing the role of hypotheses and inference, we show how the results can be used or adapted for practical purposes.

Keywords: DEA, FDH, Nonparametric estimation, Efficiency, Frontier models, Bootstrapping.

1. Frontier Analysis and The Statistical Paradigm

1.1. The Frontier Model: Economic Theory

The economic theory underlying efficiency analysis is based on the work of Koopmans (1951) and Debreu (1951) on activity analysis. Farrell (1957) is the first empirical work where the problem of measuring efficiencies for a set of observed production units is analyzed. Shephard (1970) provides a modern economic formulation of the problem.[1]

* Invited paper presented at the Third Biennial Georgia Productivity Workshop, University of Georgia, Athens, Georgia USA (October 1998). A previous version of the paper was presented at the EURO Summer Institute *DEA 20 Years On*, University of Warwick, UK (August 1998). The authors are grateful for financial support from "Projet d'Actions de Recherche Concertées" (No. 98/03–217) du Gouvernement Belge and from the Management Science Group, US Department of Veterans Affairs.

The basic concepts and notation can be fixed as follows. The activity of production units is constrained by the production set Ψ of physically attainable points (x, y):

$$\Psi = \{(x, y) \in \mathbb{R}_+^{p+q} \mid x \text{ can produce } y\}. \tag{1.1}$$

where $x \in \mathbb{R}_+^p$ is an input vector and $y \in \mathbb{R}_+^q$ is an output vector.

This set can be described mathematically by its sections. For example, in the input space we have the input requirement sets for all $y \in \Psi$:

$$X(y) = \{x \in \mathbb{R}_+^p \mid (x, y) \in \Psi\}. \tag{1.2}$$

The radial (input-oriented) efficiency boundary (“efficient frontier”) $\partial X(y)$ is then defined by:

$$\partial X(y) = \{x \mid x \in X(y), \theta x \notin X(y) \; \forall \; 0 < \theta < 1\}. \tag{1.3}$$

$\partial X(y)$ defines the “radially efficient” pairs (x, y) and the Farrell input measure of efficiency[2] for a given point (x, y) is now defined as:

$$\theta(x, y) = \inf\{\theta \mid \theta x \in X(y)\}. \tag{1.4}$$

$\theta(x, y)$ gives the radial, proportionate reduction of inputs a firm located at (x, y) should perform in order to be considered as “input technically efficient”. Given an output level y and an input-direction determined by x, the efficient level of input is given by

$$x^{\partial}(y) = \theta(x, y)x. \tag{1.5}$$

Since radial distances are considered, the polar coordinates of x are useful for describing the position of (x, y):

$$\begin{aligned} \text{modulus:} \quad & \omega = \omega(x) \in \mathbb{R}_+^1 \\ \text{angle:} \quad & \eta = \eta(x) \in \left[0, \frac{\pi}{2}\right]^{p-1}. \end{aligned} \tag{1.6}$$

In general practical situations, when analyzing a particular sector of activity, the attainable set Ψ is ***unknown***, and so are $X(y)$, $\partial X(y)$, and $\theta(x, y)$. Typically, only a sample of production units is observed:

$$\mathcal{X}_n = \{(x_i, y_i), i = 1, \ldots, n\}. \tag{1.7}$$

The so-called “deterministic frontier” approach assumes that $(x_i, y_i) \in \Psi$, $i = 1, \ldots, n$. In other words, all the observed units are attainable (*i.e.*, there are no measurement errors, no noise, *etc.*).[3]

The problem is thus to ***estimate*** the unknown quantities listed above using $\mathcal{X}_n$. Suppose we are able to obtain, *by some technique*, $\widehat{\Psi}$, $\widehat{X}(y)$ and $\widehat{\theta}(x, y)$ (which are ***estimators*** of the corresponding quantities Ψ, $X(y)$, and $\theta(x, y)$ defined above), based on the sample $\mathcal{X}_n$.

The natural questions which are then raised include:

- What are the properties of these estimators? Are they consistent, unbiased?
- What can these estimators tell us about the corresponding true values? For example, what are the confidence intervals for $\theta(x, y)$? How can they be estimated?
- Can hypotheses about the production process be tested (e.g., comparison of groups, or testing returns to scale)?

The most popular nonparametric estimators of Ψ are defined as minimal sets containing the observed data, $\mathcal{X}_n$. The Free Disposal Hull (FDH) estimator is based on free disposability assumptions on Ψ, while the Data Envelopment Analysis (DEA) estimator relies on the additional assumption of convexity. In cases where one is willing to assume constant returns to scale, the DEA estimator (of Ψ) comprising the union of the conical and free disposal hulls of $\mathcal{X}_n$ is often used.[4]

The statistical aspects of the DEA and FDH methods have gained an increasing importance in the literature; see Grosskopf (1996) for a nice survey (see also Simar (1996)). In this paper we summarize the results which are now available and provide a guide to the existing literature as well as some new Monte Carlo evidence. Stressing the roles of hypotheses and inference-making, we show how the results can be used or adapted for practical applications.

It is clear that the properties of frontier and efficiency estimators will depend on the properties of the process which generates the data $\mathcal{X}_n$. Assumptions regarding this data-generating process (DGP) comprise the statistical model. Such a model is presented in the next subsection.

Next, section 2 briefly reviews the basic definitions of FDH and DEA estimators and summarizes their basic properties. Section 3 develops more details on the asymptotic sampling distributions, and discusses their usefulness as well as their drawbacks. Section 4 shows how, in practice, the bootstrap can be implemented within this framework to estimate confidence intervals, *etc.* Section 5 debunks some claims about a recently proposed, alternative method for bootstrapping in DEA models, and section 6 discusses evidence from Monte Carlo experiments. Section 7 concludes with some comments and open questions for future research.

1.2. A Data Generating Process

Assumptions on the DGP comprise the statistical model, which defines how the observations in $\mathcal{X}_n$ are generated. Many models can be considered, but since DEA and FDH are nonparametric estimation methods, there seems to be no need for parametric assumptions about the DGP, or indeed, any assumptions other than those required to establish consistency, convergence rates, *etc.* of the estimators.[5] Here, we present perhaps the simplest statistical model, consisting of assumptions on the DGP $\mathcal{P}$, a density $f(x, y)$ on Ψ generating independent, identically distributed (iid) observations (x_i, y_i), $i = 1, \ldots, n$.[6]

ASSUMPTION A1 *free disposability.*

Inputs and outputs contained in Ψ *are freely disposable;* i.e., *if* $(x, y) \in \Psi$, *then* $(\widetilde{x}, \widetilde{y}) \in \Psi$ *for all* $(\widetilde{x}, \widetilde{y})$ *such that* $\widetilde{x} \geq x$ *and* $\widetilde{y} \leq y$.[7]

This assumption will sometimes be reinforced by a convexity assumption. In the latter case, we will replace assumption A1 with

ASSUMPTION A1′ *free disposability, and convexity of* Ψ.
Inputs and outputs contained in Ψ *are freely disposable, and* Ψ *is convex.*

A DGP must define how a sample is generated. The simplest case is to consider random samples of size n:

ASSUMPTION A2 *iid sampling.*
The sample observations in $\mathcal{X}_n$ *are realizations of iid random variables on* Ψ *with probability density function* (pdf) $f(x, y)$.

The density $f(x, y)$ can be described through many decompositions. In our case, the easiest way is in terms of cylindrical coordinates: $(x, y) \Leftrightarrow (\omega, \eta, y)$. Then, we can decompose the joint pdf of (ω, η, y) as follows:

$$f(\omega, \eta, y) = f(\omega \mid \eta, y)\, f(\eta \mid y)\, f(y) \tag{1.8}$$

where all the conditional densities exist. In particular, $f(y)$ is defined on $\mathbb{R}^q_+$, $f(\eta \mid y)$ is defined on $[0, \frac{\pi}{2}]^{p-1}$, and conditional on (y, η), the modulus ω has density $f(\omega \mid y, \eta)$ defined on $\mathbb{R}^1_+$. This completely defines a pdf $f(x, y)$ for (x, y) on Ψ.

For a given (y, η), the ***frontier*** point $x^\partial(y)$ has modulus

$$\omega(x^\partial(y)) = \inf\left\{\omega \in \mathbb{R}^1_+ \mid f(\omega \mid y, \eta) > 0\right\}. \tag{1.9}$$

It is interesting to note that the Farrell efficiency score of a point (x, y) is a simple function of the corresponding modulus:

$$0 \leq \theta(x, y) = \frac{\omega(x^\partial(y))}{\omega(x)} \leq 1. \tag{1.10}$$

Remark 1. The density $f(\omega \mid y, \eta)$ on $[\omega(x^\partial(y)), \infty]$ induces a density $f(\theta \mid y, \eta)$ on $[0, 1]$.

In order to prove consistency, we must assume that efficient units can be observed with probability approaching unity as the sample size increases:

ASSUMPTION A3 *probability mass in a neighborhood of the true frontier.*
For all $y \geq 0$ *and all* $\eta \in [0, \frac{\pi}{2}]^{p-1}$, *there exist constants* $\epsilon_1 > 0$ *and* $\epsilon_2 > 0$ *such that* $\forall\, \omega \in [\omega(x^\partial(y)), \omega(x^\partial(y)) + \epsilon_2]$, $f(\omega \mid y, \eta) \geq \epsilon_1$.

Finally, the rates of convergence of the obtained estimator will depend on the smoothness of the frontier. This smoothness can be expressed through many characteristics of the frontier; here, we describe smoothness of the frontier in terms of the Farrell efficiency scores.[8]

ASSUMPTION A4 *smoothness of the true frontier.*
For all (x, y) *in the interior of* Ψ, $\theta(x, y)$ *is differentiable in both its arguments.*

The above assumptions define the DGP $\mathcal{P}$ which generates $\mathcal{X}_n$ on Ψ. $\mathcal{P}$ may be characterized several ways:

$$\mathcal{P} = \mathcal{P}(\Psi,\ f(x, y)) = \mathcal{P}(\Psi,\ f(\omega, \eta, y)) = \mathcal{P}(\Psi,\ f(\theta, \eta, y)). \tag{1.11}$$

Remark 2. The problem of estimating the DGP $\mathcal{P}$ is thus highly related to the problem of estimating the unknown support of a multivariate density. Regarding (1.11), we note that Ψ is implicit in (as the support of) $f(x, y)$ or $f(\omega, \eta, y)$, but is ***not implicit*** in $f(\theta, \eta, y)$ because $0 \leq \theta \leq 1$. We need to know Ψ to translate $f(\theta, \eta, y)$ to $f(\omega, \eta, y)$ or $f(x, y)$.

2. The Non-Parametric Envelopment Estimators

2.1. The FDH Estimator

Deprins et al. (1984) proposed measuring efficiency for a given unit (x, y) relative to the boundary of the free disposal hull of the sample $\mathcal{X}_n$:

$$\widehat{\Psi}_{FDH}(\mathcal{X}_n) = \left\{(x, y) \in \mathbb{R}_+^{p+q} \mid y \leq y_i, x \geq x_i, (x_i, y_i) \in \mathcal{X}_n\right\}. \tag{2.1}$$

The FDH estimator of input efficiency for a given point $(x, y) \in \Psi$ is then

$$\widehat{\theta}_{FDH}(x, y) = \inf\left\{\theta \mid (\theta x, y) \in \widehat{\Psi}_{FDH}(\mathcal{X}_n)\right\}. \tag{2.2}$$

To compute $\widehat{\theta}(x_0, y_0)$ at an arbitrary point $(x_0, y_0) \in \mathbb{R}_+^{p+q}$, one typically first determines the set

$$D(x_i, y_0 \mid \mathcal{X}_n) = \{i \mid (x_i, y_i) \in \mathcal{X}_n, y_i \geq y_0\}. \tag{2.3}$$

Then the (Farrell) input efficiency of an arbitrary point (x_0, y_0) may be estimated by

$$\widehat{\theta}_{FDH}(x_0, y_0) = \min_{i \in D(\mathcal{X}_n)} \ \max_{j=1,\ldots,p} \left(\frac{x_i^j}{x_0^j}\right), \tag{2.4}$$

where x_0^j is the jth component of x_0. [9] The efficient level of input for the unit (x_0, y_0) is then defined by

$$\widehat{x}^{\partial}(y_0) = \widehat{\theta}_{FDH}(x_0, y_0)\ x_0. \tag{2.5}$$

By construction, we know that $\widehat{\Psi}_{FDH} \subseteq \Psi$, and hence

$$\widehat{\partial X}_{FDH}(y_0) = \left\{x \mid x \in \mathbb{R}_+^p,\ (x, y_0) \in \widehat{\Psi}_{FDH}(\mathcal{X}_n),\ (\theta x, y_0) \notin \widehat{\Psi}_{FDH}(\mathcal{X}_n)\ \forall\ 0 < \theta < 1\right\} \tag{2.6}$$

is an inward-biased estimator of $\partial X(y_0)$ and $\widehat{\theta}_{FDH}(x_0, y_0)$ is an upward-biased estimator of $\theta(x_0, y_0)$; *i.e.*, $\widehat{\partial X}_{FDH}(y_0) \subseteq \partial X(y_0)$ and $\widehat{\theta}_{FDH}(x_0, y_0) \geq \theta(x_0, y_0)$.[10]

2.2. *The DEA Estimator*

The DEA approach was proposed by Farrell (1957) and operationalized by Charnes et al. (1978) in terms of linear programming. This approach assumes that Ψ is convex, and employs convex estimators of Ψ. The attainable set may be estimated by the ***convex*** hull[11] of $\widehat{\Psi}_{FDH}$:

$$\widehat{\Psi}_{DEA}(\mathcal{X}_n) = \left\{ (x, y) \in \mathbb{R}_+^{p+q} \mid y \leq \sum_{i=1}^{n} \gamma_i y_i,\, x \geq \sum_{i=1}^{n} \gamma_i x_i,\; \sum_{i=1}^{n} \gamma_i = 1,\, \gamma_i \geq 0\ \forall\, i = 1, \ldots, n \right\}. \tag{2.7}$$

In the DEA approach, efficiency corresponding to a given point (x_0, y_0) is estimated relative to the boundary of $\widehat{\Psi}_{DEA}$:

$$\widehat{\theta}_{DEA}(x_0, y_0) = \inf \left\{ \theta \mid (\theta x_0, y_0) \in \widehat{\Psi}_{DEA}(\mathcal{X}_n) \right\}. \tag{2.8}$$

As before, the estimate of the efficient level of inputs corresponding to the point (x_0, y_0) is given by

$$\widehat{x}^{\partial}(y_0) = \widehat{\theta}_{DEA}(x_0, y_0) x_0. \tag{2.9}$$

Again, by construction, we know that $\widehat{\Psi}_{FDH} \subseteq \widehat{\Psi}_{DEA} \subseteq \Psi$, and so

$$\widehat{\partial X}_{DEA}(y_0) = \left\{ x \mid x \in \mathbb{R}_+^p,\, (x, y_0) \in \widehat{\Psi}_{DEA}(\mathcal{X}_n),\, (\theta x, y_0) \notin \widehat{\Psi}_{DEA}(\mathcal{X}_n)\ \forall\, 0 < \theta < 1 \right\} \tag{2.10}$$

is an inward-biased estimator of $\partial X(y_0)$, and $\widehat{\theta}_{DEA}(x_0, y_0)$ is an upward-biased estimator of $\theta(x_0, y_0)$. Moreover, $\theta(x_0, y_0) \leq \widehat{\theta}_{DEA}(x_0, y_0) \leq \widehat{\theta}_{FDH}(x_0, y_0) \leq 1$ for $(x_0, y_0) \in \widehat{\Psi}_{FDH}$.

2.3. *Properties of FDH/DEA Estimators: What Do We Know?*

In this section, we briefly summarize the main results obtained to date regarding properties of the above estimators. Consider a fixed point $(x, y) \in \mathbb{R}^p \times \mathbb{R}^q$ and let the DGP be defined by Assumptions A1 (or A1′) through A4. For the DEA estimator we need assumption A1′, whereas all the properties discussed below for the FDH estimators hold with either A1 or A1′.

2.3.1. *What About Consistency?*

The first results on consistency of DEA and FDH estimators were obtained only for the univariate case (one input in the input-oriented case or one output for the output-oriented case).

(i) The univariate case: $p = 1, q \geq 1$.

Banker (1993) proves weak consistency of the DEA input efficiency estimator for this case, *i.e.*,

$$\widehat{\theta}_{DEA}(x, y) \xrightarrow{P} \theta(x, y), \tag{2.11}$$

but gives no indication of the rate of convergence.

Korostelev et al. (1995a and 1995b) analyze the convergence of the FDH and DEA estimators of the production set Ψ itself, and obtain

$$\widehat{\Psi}_{FDH}, \widehat{\Psi}_{DEA} \xrightarrow{P} \Psi. \tag{2.12}$$

Korostelev et al. also provide rates of convergence, which reflect the ***curse of dimensionality*** present in many nonparametric statistical techniques. In particular,

$$d_{\Delta}\left(\widehat{\Psi}_{FDH}, \Psi\right) = O_p\left(n^{-\frac{1}{q+1}}\right) \tag{2.13}$$

and

$$d_{\Delta}\left(\widehat{\Psi}_{DEA}, \Psi\right) = O_p\left(n^{-\frac{2}{q+2}}\right), \tag{2.14}$$

where $d_{\Delta}(\widehat{\Psi}, \Psi)$ is the Lebesgue measure of the difference between the two sets. The rates of convergence are not very different; unless q is very small, the differences are negligible. Korostelev et al. show that under general conditions, the FDH estimator of Ψ is efficient in a statistical sense (in terms of mini-max risk) when Ψ is not convex, and the DEA estimator of Ψ is efficient when Ψ is convex.

Unfortunately, the curse of dimensionality means that if q is large, the estimators exhibit very low rates of convergence, and the applied researcher will need a large quantity of data to get sensible estimates, *i.e.*, to avoid large variances and very wide confidence interval estimates. In parametric estimation, the typical rate of convergence is $n^{-1/2}$; (2.13) indicates that $\widehat{\Psi}_{FDH}$ will have a worse rate of convergence if $q > 1$, while $\widehat{\Psi}_{DEA}$ will have a worse rate of convergence if $q > 2$. As the number of outputs is increased, the number of observations must increase at an exponential rate to maintain a given mean-square error with the nonparametric estimators of Ψ.[12]

The curse of dimensionality results from the fact as a given set of n observations are projected in an increasing number of orthogonal directions, the Euclidean distance between the observations necessarily must increase. For a given sample size, as the number of dimensions is increased, there will be increasingly fewer observations that support the nonparametric estimators $\widehat{\Psi}_{FDH}$ and $\widehat{\Psi}_{DEA}$. Parametric estimators suffer little from this phenomenon, as the parametric structure incorporates information from all of the observations, regardless of the dimensionality of Ψ. This also explains why parametric estimators are usually (but not always) more efficient in a statistical sense than their nonparametric counterparts—the parametric estimators use more information from the data, assuming of course, that the parametric assumptions that are made are correct. This is frequently a big assumption.[13]

(ii) The multivariate case: $p, q \geq 1$.

The case where efficiencies are measured radially in a multivariate framework is much more complicated, but recent papers establish some analytic results. Kneip et al. (1998) analyze the DEA case and obtain

$$\widehat{\theta}_{DEA}(x_0, y_0) - \theta(x_0, y_0) = O_p\left(n^{-\frac{2}{p+q+1}}\right), \tag{2.15}$$

while Park et al. (1999) derive similar results for the FDH estimator:

$$\widehat{\theta}_{FDH}(x_0, y_0) - \theta(x_0, y_0) = O_p\left(n^{-\frac{1}{p+q}}\right). \tag{2.16}$$

These results again reflect the curse of dimensionality, which is worse for the multivariate case since the convergence rates are affected by $p+q$ rather than merely by q as in (2.13)–(2.14).

Note also from those results that the rates of convergence for the FDH estimators are smaller than those obtained for the DEA case. If Ψ is in fact nonconvex, then there is no choice—one must use the FDH estimator, as the DEA estimator will be inconsistent. However, it Ψ is convex, the DEA estimator is preferred over the FDH estimator since the DEA estimator offers a faster rate of convergence.

2.3.2. *What About Sampling Distributions?*

Consistency is an essential property for any estimator. Indeed, it would be rather meaningless to use an estimator that does not satisfy consistency, since even with an infinite amount of data, an inconsistent estimator cannot be expected to give an accurate estimate of the quantity of interest.[14] Consistency is, however, a minimal theoretical property; from the preceding discussion, we know that DEA or FDH efficiency estimators converge as the sample size increases (although at perhaps a slow rate), but by themselves, these results have little practical use other than to confirm that the DEA or FDH estimators are possibly reasonable to use for efficiency estimation.

For applied research, more is needed—in particular, the applied researcher must have some knowledge of the sampling distributions in order to make inferences about the true levels of efficiency or inefficiency. In the case of DEA estimators, analytic results on sampling distributions have been obtained only in the univariate situation; but for FDH estimators, fully general asymptotic results have recently become available.[15]

(i) DEA case: $p = q = 1$

Gijbels et al. (1999) obtain the asymptotic result

$$\widehat{\theta}_{DEA}(x, y) - \theta(x, y) \overset{\text{asy.}}{\sim} F(\cdot, \cdot) \tag{2.17}$$

where F is a regular distribution function known up to some unknown constants. These constants depend on the the DGP $\mathcal{P}$ and are related to the curvature of the frontier, and the value of $f(x, y)$ at the true frontier point. Precise details are summarized in Theorem 1 in the next section, with additional discussion given in Gijbels *et al.*, including tabulation of values for $F(\cdot,\cdot)$).

(ii) FDH case: $p, q \geq 1$

For the general multivariate case, Park et al. (1999) derive the following result:

$$\widehat{\theta}_{FDH}(x, y) - \theta(x, y) \overset{\text{asy.}}{\sim} \text{Weibull}(\cdot,\cdot) \tag{2.18}$$

where again, the limiting Weibull depends on some unknown parameters depending on the DGP $\mathcal{P}$ (here, the slope of the frontier and the value of $f(x, y)$ near the frontier).

More details are given in the next section, but it is worthwhile at this point to mention that these results are potentially very useful. For instance, (2.17) and (2.18) allow one to compute the asymptotic value of $E(\widehat{\theta}_{DEA}(x, y))$ for the case where $p = q = 1$ or $E(\widehat{\theta}_{FDH}(x, y))$ for the general case where $p, q \geq 1$. This provides a way to estimate and perhaps correct the bias of the estimators $\widehat{\theta}_{DEA}(x, y)$ and $\widehat{\theta}_{FDH}(x, y)$. The asymptotic variance $\text{VAR}(\widehat{\theta}(x, y))$ can also be computed, and providing an indication of the sampling variability of the estimators. Finally, asymptotic confidence intervals for $\theta(x, y)$ may be estimated using the results in (2.17)–(2.18).

Unfortunately, however, (2.17) and (2.18) give only asymptotic results; approximations based on these results will only be reasonable for large n, and the curse of dimensionality is still a plague. Moreover, these results depend on unknown quantities which must be estimated, and this necessarily introduces additional noise into approximations of sampling distributions based on (2.17)–(2.18). Finite-sample properties of the approximations have been investigated via Monte-Carlo simulations in Gijbels et al. (1999) and Park et al. (1999).

2.3.3. An Alternative: The Bootstrap

The above results are important: we know the FDH and DEA estimators are consistent and we have asymptotic sampling distributions. The drawbacks, however, are also important: (i) we have ***only*** asymptotic results, which may be misleading when used in conjunction with small samples; (ii) additional noise is introduced when ***estimates*** of the unknown parameters of the limiting distributions are used in constructing estimates of confidence intervals; and (iii) in the DEA case, we have only the univariate result, whereas most applications of the DEA estimator have involved multivariate frameworks.

As in other situations where analytic results are not comforting, the bootstrap (Efron, 1979, 1982) might offer an attractive alternative for making inferences. Simar and Wilson (1998a) propose an algorithm to implement the bootstrap in the context of frontier estimation. The algorithm provides an approximation of the sampling distribution of $\widehat{\theta}_{DEA}(x, y) - \theta(x, y)$ in the multivariate case where $p, q \geq 1$; in principle, it is straightforward to extend the algorithm to approximate sampling distributions of FDH estimators. Once these approximations are obtained, confidence intervals for $\theta(x, y)$ or testing procedures can easily be implemented.

In a later section, the basic ideas behind bootstrapping in frontier models will be summarized. Note, however, as discussed below, that due to the frontier nature of the problem, implementation of the bootstrap requires some modifications of typical bootstrap methods to avoid ***inconsistency*** problems.

3. More about Asymptotic Sampling Distributions

3.1. The DEA Estimator ($p = q = 1$):

With only one input and one output, under assumption A1′, the frontier can be described by a monotone, concave production function $y = g(x)$. We will proceed with a discussion of results for the output-oriented case, but similar remarks for the input-oriented case can be made after straightforward changes in the notation below.

The Farrell output efficiency measure corresponding to a fixed point $(x_0, y_0) \in \mathbb{R}^2_+$ is defined by

$$\lambda(x_0, y_0) = \sup\{\lambda \mid (x_0, \lambda y_0) \in \Psi\}, \tag{3.1}$$

where

$$\Psi = \{(x, y) \mid f(x, y) \geq 0\} = \{(x, y) \mid y \leq g(x)\}. \tag{3.2}$$

The DEA estimators for $g(x_0)$ and $\lambda(x_0, y_0)$ are given by

$$\widehat{g}(x_0) = \sup\{y \mid (x_0, y) \in \widehat{\Psi}_{DEA}\} \tag{3.3}$$

and

$$\widehat{\lambda}(x_0, y_0) = \frac{\widehat{g}(x_0)}{y_0}, \tag{3.4}$$

respectively. Kneip et al. (1998) establishes $\widehat{g}(x_0) \xrightarrow{P} g(x_0)$ at the rate $n^{-2/3}$, which is faster than the usual parametric rate of $n^{-1/2}$; In Gijbels et al. (1999), the frontier function g is assumed to be twice continuously differentiable at x_0 and $g''(x_0) < 0$. The density function f is also assumed to be bounded away from zero and continuous in a neighborhood of the frontier. These assumptions yield

THEOREM 1 *With the above assumptions, for all* $z < 0$,

$$Pr\left\{n^{2/3}(b_0^2/b_2)^{1/3}[\widehat{g}(x_0) - g(x_0)] \leq z\right\} = \int_0^\infty \varphi(u, z)\, du + o(1), \tag{3.5}$$

where

$$\begin{aligned}
\varphi(u, z) &= (1/2)(-z)^{3/2}(1 + u^2)\exp\{-(1/6)(-z)^{3/2}(u + u^{-1})^3\},\\
b_0 &= f(x_0, g(x_0)),\\
b_2 &= -g''(x_0)/2.
\end{aligned}$$

The leading term in the right-hand side of (3.5) is known. The asymptotic bias and standard deviation of $\widehat{g}(x_0)$ can also be easily derived:

$$\text{asymp. bias of } \widehat{g(x_0)} = -n^{-2/3}(b_2/b_0^2)^{1/3}c_1, \tag{3.6}$$

and

$$\text{asymp. std. deviation of } \widehat{g(x_0)} = n^{-2/3}(b_2/b_0^2)^{1/3}c_2, \tag{3.7}$$

where

$$c_1 = \int_0^\infty \int_0^\infty \varphi(u, -z)\, du\, dz \approx 0.99360$$

and

$$c_2^2 = 2\int_0^\infty \int_0^\infty z\varphi(u, -z)\, du\, dz - c_1^2 \approx (0.55757)^2.$$

The expressions in (3.6) and (3.7) illustrate the role of the sample size n, as well as that of the characteristics of the DGP described by the unknown constants b_2 and b_0, which characterize the curvature of the frontier and the magnitude of the density at the frontier point $(x_0, g(x_0))$. In particular, it appears that the bias is much more important than the standard deviation in determining mean square error.

For practical purposes, the parameters b_0 and b_2 must be estimated from the data. Consistent nonparametric estimators of these are provided in Gijbels et al. (1999). The estimates of b_0 and b_2 can then be used to construct a bias-corrected estimator of the frontier, or to estimate asymptotic confidence intervals for $g(x_0)$ using the percentiles of the right-hand side of (3.5). Finite sample properties of the bias-corrected estimator of $g(x_0)$ are investigated in Gijbels et al. (1999). Even for moderate sample sizes, e.g. $n = 100$, the bias-corrected estimator and the coverage probabilities of the estimated confidence intervals behave reasonably well.

Of course, the above results can also be translated into terms of the efficiency estimator $\widehat{\lambda}(x_0, y_0)$ by a simple transformation of variables using (3.4), where y_0 is fixed.

3.2. *FDH Estimator* $p, q \geq 1$:

Maintaining the output-orientation and considering once again a fixed point (x_0, y_0), Park et al. (1999) obtain the following theorem, which is written in terms of Farrell output-efficiency scores:

THEOREM 2 *Under the assumptions A1–A4,*

$$\lambda(x_0, y_0) - \widehat{\lambda}(x_0, y_0) \overset{\text{asy.}}{\sim} \textit{Weibull}(n\mu^{p+q}, p+q). \tag{3.8}$$

The constant μ appearing in the limiting Weibull depends on the DGP $\mathcal{P}$. It can be interpreted as follows:

$$P(x_i \leq x_0, y_i \geq y_0^\partial(x_0) - zy_0) = (\mu z)^{p+q} + O(z^{p+q+1}), \tag{3.9}$$

where $y_0^\partial(x_0) = \lambda(x_0, y_0)y_0$ and $z = o(1)$. In other words, μ is proportional, for small z, to the $(p+q)$th root of the probability of observing a firm dominating the point $(x_0, (\lambda(x_0, y_0) - z)y_0)$, which is a point near the frontier above (x_0, y_0). Therefore, μ will be larger (smaller) if $f(x, y)$ provides more (less) mass in a neighborhood of the frontier.

The asymptotic bias and standard deviation of $\widehat{\lambda}(x_0, y_0)$ are given by:

$$\text{asymp. bias of } \widehat{\lambda}(x_0, y_0) = -n^{-1/(p+q)}\mu^{-1}\gamma_1 \tag{3.10}$$

and

$$\text{asymp. std. deviation of } \widehat{\lambda}(x_0, y_0) = n^{-1/(p+q)}\mu^{-1}\left(\gamma_2 - \gamma_1^2\right)^{1/2}, \tag{3.11}$$

where $\gamma_j = \Gamma\left(\frac{p+q+j}{p+q}\right)$ for $j = 1, 2$. Again, (3.10) and (3.11) illustrate the roles of the sample size n and of the parameter μ. The bias remains more important than standard deviation in determining mean square error (and still more so if $p+q$ is large).

In applied work, the parameter μ must be estimated. Park et al. (1999) propose a nonparametric, consistent estimator of μ. This allows one to construct asymptotic, bias-corrected estimators of the efficiency measure $\lambda(x_0, y_0)$, and to estimate confidence intervals for $\lambda(x_0, y_0)$ using percentiles of the Weibull distribution.

Finite-sample properties of these estimators are analyzed in Park et al. (1999). For small values of $p+q$ (e.g., 2 or 3), they behave rather well with $n = 100$ or 200, but performance deteriorates as $p+q$ increases due to the still-present curse of dimensionality. For example, with $p+q = 5$, at least 500 observations are needed to get meaningful estimates.

4. More on the Bootstrap

The bootstrap is an attractive alternative to the theoretical results discussed in the previous section. For the multivariate DEA case, at least so far, the bootstrap seems to offer the only approach for estimating the sampling variation of efficiency estimators. The essence of the bootstrap idea (Efron, 1979, 1982; Efron and Tibshirani, 1993) is to approximate the sampling distributions of interest by simulating, or mimicking, the DGP. The first use of the bootstrap in frontier models dates to Simar (1992). Its development for nonparametric envelopment estimators was introduced by Simar and Wilson (1998a).

4.1. General Principles

We will make the presentation in the input-oriented efficiency measure in order to use the notation introduced earlier in sections 1 and 2. We also focus on the DEA estimators; adaptation for output measures or for FDH estimators is straightforward.

4.1.1. The Basic Idea

The original data $\mathcal{X}_n$ are generated from $\mathcal{P}$, where

$$\mathcal{P} = \mathcal{P}(\Psi, f(x, y)). \tag{4.1}$$

Let $\widehat{\mathcal{P}}(\mathcal{X}_n)$ be a consistent estimator of the DGP $\mathcal{P}$:

$$\widehat{\mathcal{P}}(\mathcal{X}_n) = \mathcal{P}\left(\widehat{\Psi},\ \widehat{f}(x, y)\right). \tag{4.2}$$

In the ***true world***, $\mathcal{P}$, Ψ and $\theta(x_0, y_0)$ are unknown, but we observe the data $\mathcal{X}_n$ which are used to construct corresponding estimates $\widehat{\mathcal{P}}$, $\widehat{\Psi}$ and $\widehat{\theta}_{DEA}(x_0, y_0)$.

Now consider a virtual, simulated world we call the ***bootstrap world***. A new dataset $\mathcal{X}_n^* = \{(x_i^*, y_i^*),\ i = 1, \ldots, n\}$ can be drawn from $\widehat{\mathcal{P}}$. Within the bootstrap world, $\widehat{\Psi}$ is the true attainable set, and the union of the free disposal and the convex hulls of $\mathcal{X}_n^*$, gives an estimator of $\widehat{\Psi}$, namely

$$\widehat{\Psi}^*(\mathcal{X}_n^*) = \Big\{(x, y) \in \mathbb{R}^{p+q} \ \mid\ y \le \sum_{i=1}^n \gamma_i y_i^*; \quad x \ge \sum_{i=1}^n \gamma_i x_i^* \\ \sum_{i=1}^n \gamma_i = 1; \quad \gamma_i \ge 0\ \forall\, i = 1, \ldots, n\Big\}. \tag{4.3}$$

For a fixed point $(x_0, y_0) \in \mathbb{R}_+^p \times \mathbb{R}_+^q$,

$$\widehat{\theta}_{DEA}^*(x_0, y_0) = \inf\{\theta \mid (\theta x_0, y_0) \in \widehat{\Psi}^*\} \tag{4.4}$$

gives an estimator of $\widehat{\theta}_{DEA}(x_0, y_0)$, which is the true efficiency in the bootstrap world (but not the true world). $\widehat{\theta}_{DEA}^*(x_0, y_0)$ may be computed by solving the linear program

$$\widehat{\theta}_{DEA}^*(x_0, y_0) = \min\Big\{\theta > 0 \ \mid\ y_0 \le \sum_{i=1}^n \gamma_i y_i^*; \theta x_0 \ge \sum_{i=1}^n \gamma_i x_i^*; \\ \sum_{i=1}^n \gamma_i = 1; \gamma_i \ge 0\ \forall\, i = 1, \ldots, n\Big\}, \tag{4.5}$$

which is analogous to how one computes the original estimate $\widehat{\theta}_{DEA}(x_0, y_0)$ in the true world.

The ***key relation*** here is that ***within the true world***, $\widehat{\theta}_{DEA}(x_0, y_0)$ is an estimator of $\theta(x_0, y_0)$, based on the sample $\mathcal{X}_n$ generated from $\mathcal{P}$; whereas ***in the bootstrap world***, $\widehat{\theta}_{DEA}^*(x_0, y_0)$ is an estimator of $\widehat{\theta}_{DEA}(x_0, y_0)$, based on the pseudo-sample $\mathcal{X}_n^*$ generated from $\widehat{\mathcal{P}}(\mathcal{X}_n)$. If the bootstrap is ***consistent***, then

$$\left(\widehat{\theta}_{DEA}^*(x_0, y_0) - \widehat{\theta}_{DEA}(x_0, y_0)\right) \mid \widehat{\mathcal{P}}(\mathcal{X}_n) \overset{\text{approx}}{\sim} \left(\widehat{\theta}_{DEA}(x_0, y_0) - \theta(x_0, y_0)\right) \mid \mathcal{P}, \tag{4.6}$$

which is merely an application of the analogy principle (e.g., see Manski (1988)). Within the bootstrap world and conditional on the observed data $\mathcal{X}_n$, the sampling distribution of $\widehat{\theta}_{DEA}^*(x_0, y_0)$ is (in principle) completely known since $\widehat{\mathcal{P}}(\mathcal{X}_n)$ is known. However, in practice, it is impossible to compute this analytically. Hence Monte-Carlo simulations are necessary to approximate the left-hand side of (4.6).

Using $\widehat{\mathcal{P}}(\mathcal{X}_n)$ to generate B samples $\mathcal{X}_{nb}^*$ of size n, $b = 1, \ldots, B$, and applying the original estimator to these pseudo samples, yields a set of B pseudo estimates $\widehat{\theta}_{DEA,b}^*(x_0, y_0)$, $b = 1, \ldots, B$. The empirical distribution of these bootstrap values gives a Monte Carlo approximation of the sampling distribution of $\widehat{\theta}^*_{DEA}(x_0, y_0)$, conditional on $\widehat{\mathcal{P}}(\mathcal{X}_n)$, *i.e.*, the left-hand side of (4.6). The quality of the approximation relies in part on the value of B: by the law of large numbers, when $B \to \infty$, the error of this approximation due to the bootstrap resampling (*i.e.*, drawing from $\widehat{\mathcal{P}}$) tends to zero. The practical choice of B is limited by the speed of one's computer; for confidence intervals, values of 2000 or more may be needed to give a reasonable approximation.[16] In addition, the bootstrap is an asymptotic procedure, as indicated by the conditioning on the left-hand-side of (4.6). Thus the quality of the bootstrap approximation depends on both the number of replications B and and the sample size n. The approximation becomes exact as $B \to \infty$ and $n \to \infty$.

4.1.2. Bootstrap Confidence Intervals

The procedure described in Simar and Wilson (1998a) for constructing the confidence intervals depends on using bootstrap estimates of bias to correct for the bias of the DEA estimators; in addition, the procedure described there requires using these bias estimates to shift obtained bootstrap distribution appropriately. Use of bias estimates introduces additional noise into the procedure. In Simar and Wilson (1999c), we proposed the improved procedure outlined below which automatically corrects for bias without explicit use of a noisy bias estimator.

If we knew the distribution of $\left(\widehat{\theta}_{DEA}(x_0, y_0) - \theta(x_0, y_0)\right)$, then it would be trivial to find values a_α and b_α such that

$$\Pr\left(-b_\alpha \le \widehat{\theta}_{DEA}(x_0, y_0) - \theta(x_0, y_0) \le -a_\alpha\right) = 1 - \alpha. \tag{4.7}$$

Of course, a_α and b_α are unknown, but from the empirical bootstrap distribution of the pseudo estimates $\widehat{\theta}^*_{DEA,b}$, $b = 1, \ldots, B$, we can find values $\widehat{a}_\alpha$ and $\widehat{b}_\alpha$ such that

$$\Pr\left(-\widehat{b}_\alpha \le \widehat{\theta}^*_{DEA}(x_0, y_0) - \widehat{\theta}_{DEA}(x_0, y_0) \le -\widehat{a}_\alpha \mid \widehat{\mathcal{P}}(\mathcal{X}_n)\right) = 1 - \alpha. \tag{4.8}$$

Finding $\widehat{a}_\alpha$ and $\widehat{b}_\alpha$ involves sorting the values $\left(\widehat{\theta}^*_{DEA,b}(x_0, y_0) - \widehat{\theta}_{DEA}(x_0, y_0)\right)$, $b = 1, \ldots, B$ in increasing order and then deleting $(\frac{\alpha}{2} \times 100)$-percent of the elements at either end of the sorted list. Then set $-\widehat{b}_\alpha$ and $-\widehat{a}_\alpha$ equal to the endpoints of the truncated, sorted array, with $\widehat{a}_\alpha \le \widehat{b}_\alpha$.

The bootstrap approximation of (4.7) is then

$$\Pr\left(-\widehat{b}_\alpha \le \widehat{\theta}_{DEA}(x_0, y_0) - \theta(x_0, y_0) \le -\widehat{a}_\alpha\right) \approx 1 - \alpha. \tag{4.9}$$

The estimated $(1 - \alpha)$-percent confidence interval is then

$$\widehat{\theta}_{DEA}(x_0, y_0) + \widehat{a}_\alpha \le \theta(x_0, y_0) \le \widehat{\theta}_{DEA}(x_0, y_0) + \widehat{b}_\alpha. \tag{4.10}$$

This procedure can be used for any $(x_0, y_0) \in \mathbb{R}^{p+q}_+$ for which $\widehat{\theta}_{DEA}(x_0, y_0)$ exists. Typically, the applied researcher is interested in the efficiency scores of the observed units themselves; in this case, the above procedure can be repeated n times, with $(x_0, y_0) = (x_i, y_i)$, $i = 1, \ldots, n$, producing a set of n confidence intervals of the form (4.10), one for each firm.

4.1.3. Bootstrap bias corrections

We showed above that the DEA (and FDH) estimators are biased by construction. By definition,

$$\text{BIAS}\left(\widehat{\theta}_{DEA}(x_0, y_0)\right) = E\left(\widehat{\theta}_{DEA}(x_0, y_0)\right) - \theta(x_0, y_0). \tag{4.11}$$

The bootstrap bias estimate for the original estimator $\widehat{\theta}_{DEA}(x_0, y_0)$ is the empirical analog of (4.11):

$$\widehat{\text{BIAS}}_B\left(\widehat{\theta}_{DEA}(x_0, y_0)\right) = B^{-1}\sum_{b=1}^{B}\widehat{\theta}^*_{DEA,b}(x_0, y_0) - \widehat{\theta}_{DEA}(x_0, y_0). \tag{4.12}$$

It is tempting to construct a bias-corrected estimator of $\theta(x_0, y_0)$ by computing

$$\begin{aligned}\widehat{\widehat{\theta}}_{DEA}(x_0, y_0) &= \widehat{\theta}_{DEA}(x_0, y_0) - \widehat{\text{BIAS}}_B\left(\widehat{\theta}_{DEA}(x_0, y_0)\right) \\ &= 2\widehat{\theta}_{DEA}(x_0, y_0) - B^{-1}\sum_{b=1}^{B}\widehat{\theta}^*_{DEA,b}(x_0, y_0). \end{aligned} \tag{4.13}$$

It is well known (e.g., see Efron and Tibshirani (1993)), however, that this bias correction introduces additional noise; the mean-square error of $\widehat{\widehat{\theta}}_{DEA}(x_0, y_0)$ may be greater than the mean-square error of $\widehat{\theta}_{DEA}(x_0, y_0)$. The variance of the summation term on the right-hand side of (4.13) can be made arbitrarily small by increasing B. Yet, even if $B \to \infty$, the bias-corrected estimator $\widehat{\widehat{\theta}}_{DEA}(x_0, y_0)$ will have variance equal to four times that of the original, uncorrected estimator, $\widehat{\theta}_{DEA}(x_0, y_0)$, (again illustrating the fact that the bootstrap is an asymptotic procedure). The sample variance of the bootstrap values $\widehat{\theta}^*_{DEA,b}(x_0, y_0)$ gives an estimate $\widehat{\sigma}^2$ of the variance of $\widehat{\theta}_{DEA}(x_0, y_0)$:

$$\widehat{\sigma}^2 = B^{-1}\sum_{b=1}^{B}\left[\widehat{\theta}^*_{DEA,b}(x_0, y_0) - B^{-1}\sum_{b=1}^{B}\widehat{\theta}^*_{DEA,b}(x_0, y_0)\right]^2. \tag{4.14}$$

Hence the bias-correction should not be used unless

$$\widehat{\sigma}^2 < \frac{1}{3}\left[\widehat{\text{BIAS}}_B\left(\widehat{\theta}_{DEA}(x_0, y_0)\right)\right]^2. \tag{4.15}$$

4.2. Is The Bootstrap Consistent?

The crucial remaining question is how to generate pseudo samples $\mathcal{X}_n^*$ from $\widehat{\mathcal{P}}$ such that (4.6) holds. From (1.11) we have

$$\widehat{\mathcal{P}} = \mathcal{P}(\widehat{\Psi}, \widehat{f}(x, y)) = \mathcal{P}(\widehat{\Psi}, \widehat{f}(\omega, \eta, y)) = \mathcal{P}(\widehat{\Psi}, \widehat{f}(\theta, \eta, y)) \tag{4.16}$$

where $\widehat{\Psi}$ is a consistent estimator of Ψ.

Variants of the so-called ***naive bootstrap*** either (i) use the empirical distribution of $\{(x_i, y_i),\ i = 1, \ldots, n\}$ to estimate $f(x, y)$; or (ii) alternatively, the empirical distribution of $\{(\widehat{\theta}_i, \eta_i, y_i),\ i = 1, \ldots, n\})$ is used to estimate $f(\theta, \eta, y)$. In the first variation, pseudo samples $\{(x_i^*, y_i^*),\ i = 1, \ldots, n\}$ are drawn from the empirical density $\widehat{f}(x, y)$ to obtain $\mathcal{X}_n^*$. In the second variation, $\{(\theta_i^*, \eta_i^*, y_i^*),\ i = 1, \ldots, n\}$ is drawn from the empirical density $\widehat{f}(\theta, \eta, y)$, and the pseudo observations $(\theta_i^*, \eta_i^*, y_i^*)$ expressed in cylindrical coordinates are translated back to Cartesian coordinates (x_i^*, y_i^*), to obtain $\mathcal{X}_n^*$ (see Remark 2).

The problem with the naive bootstrap is that ***both*** variations involve using empirical distributions to estimate distribution functions with bounded, unknown support on Ψ. It is well-known that bootstrapping in such a context may lead to serious problems (Efron and Tibshirani, 1993, and many others), although this point has been ignored by some in the efficiency estimation literature.

As shown below, the naive bootstrap is inconsistent regardless of which variation is used; in other words, with the naive bootstrap,

$$\left(\widehat{\theta}^*(x_0, y_0) - \widehat{\theta}(x_0, y_0)\right) \mid \widehat{\mathcal{P}}(\mathcal{X}_n) \overset{\text{approx}}{\not\sim} \left(\widehat{\theta}(x_0, y_0) - \theta(x_0, y_0)\right) \mid \mathcal{P}. \tag{4.17}$$

Consequently, the naive bootstrap is useless since it generates the left-hand side of (4.17), which does not approximate the desired distribution appearing on the right-hand side of (4.17). In the next section, we propose two approaches to solve this problem.

4.3. *A Smooth (Not Naive) Bootstrap*

Two approaches have been proposed to avoid the inconsistency of the naive bootstrap. Analogous with parametric regression problems where one sometimes bootstraps on the "residuals," here we can bootstrap on the radial inefficiencies; this approach is the homogeneous case proposed by Simar and Wilson (1998a). The approach is based on an homogeneity assumption for the structure of inefficiency:

$$f(\theta \mid \eta, y) = f(\theta). \tag{4.18}$$

The second approach is the heterogeneous case proposed by Simar and Wilson (2000), which relaxes the assumption in (4.18). In the heterogeneous approach, there are no restrictions on $f(\theta \mid \eta, y)$. Thus, the second approach allows for possible heterogeneity in the structure of inefficiency.

4.3.1. *The Homogeneous Bootstrap*

By bootstrapping on the estimated inefficiencies $\widehat{\theta}(x_i, y_i)$ corresponding to each observation in $\mathcal{X}_n$, we are bootstrapping on the radial distances from each (x_i, y_i), $i = 1, \ldots, n$ to the frontier of $\widehat{\Psi}(\mathcal{X}_n)$. This step provides θ_i^* which, together with (x_i, y_i) and $\widehat{\Psi}(\mathcal{X}_n)$, allows computation of (x_i^*, y_i). The pseudo sample is then defined by $\mathcal{X}_n^* = \{(x_i^*, y_i) \mid i =$

$1, \ldots, n\}$, where

$$x_i^* = \frac{\widehat{x^\partial}(y_i)}{\theta_i^*} = \frac{\widehat{\theta}(x_i, y_i)}{\theta_i^*} x_i. \tag{4.19}$$

In the homogeneous case, it seems natural to estimate $f(\theta \mid \eta, y) = f(\theta)$ from the set $\{\widehat{\theta}(x_i, y_i) \mid i = 1, \ldots, n\}$. This yields a density estimate $\widehat{f}(\theta)$, and the bootstrapped radial distances $(\theta_i^*, i = 1, \ldots, n)$ are then drawn from $\widehat{f}(\theta)$.

The crucial aspect of the procedure revolves around how the density estimate $\widehat{f}(\theta)$ is defined. The naive bootstrap proposed by Ferrier and Hirschberg (1997) estimates $f(\theta)$ via the empirical density function, placing a mass $1/n$ at each observed $\widehat{\theta}_i$; (*i.e.*, Ferrier and Hirschberg resample θ_i^* by drawing independently, with replacement, from $\{\widehat{\theta}_i \mid i = 1, \ldots, n\}$). As discussed in Simar and Wilson (1999a, 1999b), this naive bootstrap is inconsistent.

The problem in the naive bootstrap results from the discreteness of the empirical probability density function of $\widehat{\theta}(x_i, y_i)$, combined with the fact that at least one (and frequently, many) of the $\widehat{\theta}(x_i, y_i)$ is necessarily equal to 1. Thus, in the bootstrap pseudo sample $\mathcal{X}_n^*$ for any particular bootstrap replication, we have for any $i = 1, \ldots, n$

$$\Pr(\theta_i^* = 1 \mid \mathcal{X}_n) \geq 1 - (1 - n^{-1})^n > 0, \tag{4.20}$$

with equality holding only if only one of the $\widehat{\theta}(x_i, y_i)$ equals unity. Letting $n \to \infty$ reveals

$$\lim_{n \to \infty} \Pr(\theta_i^* = 1 \mid \mathcal{X}_n) \geq 1 - e^{-1} \approx 0.632. \tag{4.21}$$

Note that this result does not depend on the DGP generating the $\theta(x_i, y_i)$. Even if one allows a probability mass along the frontier, the naive bootstrap would still be inconsistent, even if this mass was $1 - e^{-1}$. Under our assumptions, $\Pr(\theta_i = 1) = 0$, which is far from the value in (4.21).

Similar reasoning suggests that with p inputs and q outputs, regardless of the DGP, the probability that the efficiency score estimated from the bootstrap pseudo sample $\mathcal{X}_n^*$ for a point (x_0, y_0) equals the efficiency score estimated from the original sample for the same point is given by

$$\Pr(\widehat{\theta}^*(x_0, y_0) = \widehat{\theta}(x_0, y_0) \mid \mathcal{X}_n) \geq \left(1 - (1 - n^{-1})^n\right)^{p+q} > 0. \tag{4.22}$$

Again, letting $n \to \infty$ reveals

$$lim_{n \to \infty} \Pr(\widehat{\theta}^*(x_0, y_0) = \widehat{\theta}(x_0, y_0) \mid \mathcal{X}_n) \geq \left(1 - e^{-1}\right)^{p+q} \approx 0.632^{p+q}. \tag{4.23}$$

However, under our assumptions $\Pr(\theta_i = 1) = 0$, and hence

$$\Pr(\widehat{\theta}(x_0, y_0) = \theta(x_0, y_0)) = 0. \tag{4.24}$$

Therefore the naive bootstrap remains inconsistent in higher dimensions. Even if $\Pr(\widehat{\theta}(x_0, y_0) = \theta(x_0, y_0)) > 0$, the naive bootstrap would still be inconsistent in general because there is not reason to assume that the implied probability mass is the same as implied by (4.23). One might suspect that for large values of $p+q$, the term on the right-hand-side of (4.23) will be reasonably close to zero, and hence the naive bootstrap might give reasonable results. Unfortunately, this is not the case; see Simar and Wilson (1999a, 1999b) for discussion. Moreover, (4.23) shows that these are not problems of finite samples. The naive bootstrap does not mimic the desired distribution, and so (4.6) is violated.

One solution to these problems is to draw θ_i^* from a smooth, consistent, nonparametric estimator of $f(\theta)$, taking account of the boundary conditions ($0 \leq \theta \leq 1$). An easy-to-implement algorithm for consistently generating the bootstrap values θ_i^* from a kernel density estimate is given in Simar and Wilson (1998a). The bandwidth parameter needed for the kernel density estimator can be chosen via cross-validation techniques as in Simar and Wilson (2000, 1998b, 1999c).

4.3.2. The Heterogeneous Bootstrap

Bootstrapping on the pairs $(x_i, y_i) \in \mathcal{X}_n$ generates (x_i^*, y_i^*) from a consistent estimator of $f(x, y)$. The naive bootstrap involves estimating the distribution of (x, y) by the empirical distribution of the observed sample $\mathcal{X}_n$. In other words, the naive bootstrap resamples by drawing the input-output pairs (x_i, y_i) identically, with replacement, from the original pairs in $\mathcal{X}_n$ to obtain $\mathcal{X}_n^* = \{x_i^*, y_i^*) \mid i = 1, \ldots, n\}$. This again yields inconsistent bootstrap approximations, for reasons similar to those in the homogeneous naive bootstrap. In particular, for a particular replication, the present variation of the naive bootstrap yields

$$\Pr\left(\widehat{\theta}^*(x_0, y_0) = \widehat{\theta}(x_0, y_0)\right) = \sum_{j=0}^{p+q} \binom{p+q}{j} (-1)^j \left(1 - \frac{j}{n}\right)^n > 0. \tag{4.25}$$

This is the probability that the bootstrap sample $\mathcal{X}_n^*$ contains the dominating facet of (x_0, y_0) in the original sample $\mathcal{X}_n$. This is also the probability that $\widehat{\theta}^*(x_0, y_0) = \widehat{\theta}(x_0, y_0)$ on a given bootstrap replication. Since this probability is nonzero, fixed, and does not depend on the DGP $\mathcal{P}$, the resulting bootstrap estimate of the sampling distribution of the original estimator $\widehat{\theta}(x_0, y_0)$ will have a probability mass at a point. The problem does not disappear in large samples since

$$\lim_{n\to\infty} \Pr\left(\widehat{\theta}^*(x_0, y_0) = \widehat{\theta}(x_0, y_0)\right) = \left(1 - e^{-1}\right)^{p+q} > 0. \tag{4.26}$$

Hence, even with $n = \infty$, the bootstrap estimate of the sampling distribution of the original estimator $\widehat{\theta}(x_0, y_0)$ will have a probability mass at a point. Since this mass depends exclusively on the dimension of the input/output space, $p+q$, rather than any feature of the true DGP, the heterogeneous naive bootstrap cannot be consistent.

As in the homogeneous case, the solution to this problem is (at least conceptually) simple: draw iid bootstrap samples $\mathcal{X}_n^* = \{(x_i^*, y_i^*) \mid i = 1, \ldots, n\}$ from a smooth, consistent estimator $\widehat{f}(x, y)$ of the joint density of (x, y) on Ψ. For practical purposes, it is easier to draw $(\theta_i^*, \eta_i^*, y_i^*)$ from a smooth, consistent estimator of $f(\theta, \eta, y)$; using $\widehat{\Psi}$, these points expressed in cylindrical coordinates can then be translated to Cartesian coordinates to obtain the desired bootstrap pseudo sample $\mathcal{X}_n^* = \{(x_i^*, y_i^*) \mid i = 1, \ldots, n\}$. The complete algorithm is provided in Simar and Wilson (2000).

5. Yet Another Bad Bootstrap Idea

While we have addressed the problems inherent in the naive bootstrap elsewhere, as noted in the previous section, another unfortunately bad idea has recently appeared. Löthgren and Tambour (1996, 1997) and Löthgren (1997, 1998) employ a peculiar variant of the homogeneous naive bootstrap described above in section (4.3.1). In their method, bootstrap values θ_i^* are drawn from the empirical distribution of the original efficiency estimates $\widehat{\theta}_i$ as in the ordinary homogeneous naive bootstrap. For the input-orientation, pseudo data are constructed by computing $x_i^* = x_i\widehat{\theta}(x_i, y_i)/\theta_i^*$ and setting $y_i^* = y_i$ for each $i = 1, \ldots, n$ to obtain a pseudo sample $\mathcal{X}_n^*$.

Computing $\widehat{\theta}^*(x_0, y_0)$ from (4.5) would yield the ordinary homogeneous naive bootstrap results described in section (4.3.1). Löthgren and Tambour deviate from the more ordinary naive bootstrap at this point, however, by computing

$$\widehat{\theta}_0^* = \widehat{\theta}^*(x_0^*, y_0^*) = \min\Bigg\{\theta > 0 \;\Big|\; y_0^* \le \sum_{i=1}^n \gamma_i y_i^*;\ \theta x_0^* \ge \sum_{i=1}^n \gamma_i x_i^*; \sum_{i=1}^n \gamma_i = 1;\ \gamma_i \ge 0\ \forall\ i = 1, \ldots, n\Bigg\} \tag{5.1}$$

where $y_0^* = y_0$ and $x_0^* = x_0\widehat{\theta}_0/\theta^*$, where θ^* is a random drawn from the empirical density function placing a mass $1/n$ at each observed $\widehat{\theta}_i$ (as in the naive bootstrap method). They then repeat the procedure B times, drawing a new pseudo sample $\mathcal{X}_n^*$ and solving (5.1) for each of the B replications, yielding a set $\{\widehat{\theta}_{0,b}^* \mid b = 1, \ldots, B\}$.

Note that on each of the B bootstrap replications, the Löthgren-Tambour (LT) method replaces θx_0 on the left-hand side of the second constraint in (4.5) with θx_0^*. Consequently, the LT bootstrap involves measuring the distance ***from a different, random*** (as opposed to fixed) point to the boundary of $\widehat{\Psi}^*(\mathcal{X}_n^*)$ ***on each replication*** of the bootstrap Monte Carlo exercise. It seems entirely unclear what this procedure estimates. Certainly, it does not estimate anything of interest: the resulting set $\{\widehat{\theta}_{0,b}^* \mid b = 1, \ldots, B\}$ has nothing to do with the original, unknown $\theta_0 = \theta(x_0, y_0)$; at best, it provides an approximation of $f(\theta)$ itself, as does the empirical density of the $\widehat{\theta}_i$. To be more explicit, θ_0 may or may not belong to a 95% probability interval constructed from the set $\widehat{\theta}_{0,b}^*$, $b = 1, \ldots, B$ even when $B \to \infty$: it depends on where the point of interest (x_0, y_0) is in located.

Moreover, in the Monte Carlo experiments Löthgren and Tambour provide, the presentation is misleading and incorrect. In their experiments, on each bootstrap replication, they compute the efficiency scores points which differ with each bootstrap replication. ***By construction***, the true values $\theta(x_0, y_0)$ in their experiments will lie within the 95 percent confidence intervals they estimate in roughly 95 percent of their Monte Carlo trials. However, the 95 percent confidence intervals they estimate are, for any $(x_0, y_0) = (x_i, y_i)$, roughly equivalent to the 2.5th and 97.5th percentiles of their approximation to $f(\theta)$. So, averaging over n firms and over the M Monte Carlo experiments will provide estimates of coverage probabilities similar to those reported in Löthgren (1998) which appear to be approximately correct (small differences are due to finite values of n, M and B). But, the estimated confidence intervals merely reflect the 2.5th and 97.5th percentiles of an approximation to $f(\theta)$, and thus are much wider than the true 95 percent confidence intervals corresponding to the true efficiency for any particular point. Moreover, if (x_0, y_0) is not chosen to be one of the original sample observations, there is no guarantee that the corresponding true efficiency will ever lie within the confidence intervals estimated via the LT bootstrap.

In any estimation problem where inference is to be performed, it is essential to remember what is known and what is unknown. In the typical DEA or FDH setting, the data in $\mathcal{X}_n$ are observed and hence must be known. If this were not the case, it would not be possible to estimate anything. On the other hand, Ψ and $\theta(x_0, y_0)$ are unknown, and hence must be ***estimated***. If this were not the case, *i.e.* if Ψ and $\theta(x_0, y_0)$ were known, there would be nothing to estimate, as the things that are typically of interest would already be known. We risk *la gaucherie* by stating the obvious, but we are compelled to do so because some have ignored these points. In efficiency estimation, since Ψ is unknown, we must rely on an ***estimator*** of this set, namely $\widehat{\Psi}(\mathcal{X}_n)$. Interest lies in the distance from a fixed point, perhaps corresponding to one of the original sample observations, to the true frontier. But the true frontier is unobserved; all that is available is a biased, but consistent, estimate of its location. Estimators are random, and the bootstrap mimics this by determining a new estimate of the production frontier on each bootstrap replication. The original data, however, are known, and it is inconceivable that one should throw away data in any estimation problem—but this is exactly what is done in the LT bootstrap. In other words, the LT method assumes not only that Ψ is unknown, but also (implicitly) that the point from which one wishes to measure distance to the frontier of Ψ is unknown. This is absurd.

Our reasoning in this section is confirmed by Monte Carlo experiments discussed in the next section.

6. Some Monte Carlo Evidence

6.1. The Experimental Framework

We conducted a series of Monte Carlo experiments to measure the performance of the bootstrap under both constant returns to scale (CRS) and variable returns to scale (VRS). Each Monte Carlo experiment involved $M = 1000$ trials; in each trial, we simulated data from a true, known model, estimated Shephard input distance functions, and then bootstrapped to estimate confidence intervals. In each application of the bootstrap, we set the number of bootstrap replications $B = 2000$.

To examine the performance of the smooth, Simar-Wilson (SW) bootstrap, on each Monte Carlo trial we estimated efficiency for a fixed point (x_0, y_0), not necessarily contained in the data. We then bootstrapped to obtain confidence intervals for the corresponding efficiency of the point (x_0, y_0). Since the true model is known within this experimental framework, the true efficiency corresponding to the fixed point (x_0, y_0) is also known. For confidence intervals of a given nominal size, we counted the number of trials in which the estimated confidence intervals included the true efficiency score. We then divided this count by $M = 1000$ to obtain Monte Carlo estimates of the coverage probabilities of our estimated confidence intervals. Hence, for example, at nominal size equal to .05, the estimated confidence intervals should include the true efficiency score in roughly 95 percent of the Monte Carlo trials.

To examine the behavior of the Löthgren/Tambour (LT) bootstrap discussed earlier in section 5, we used a slightly different protocol to remain consistent with the presentation in the various papers by Löthgren and Tambour. In experiments with the LT bootstrap, we use the same true models and the same mechanisms to simulate data as in our experiments with the SW bootstrap. However, instead of considering a single point (x_0, y_0), we apply the LT bootstrap to each of the n sample observations simulated for a particular Monte Carlo trial. Coverage probabilities for the LT bootstrap are then estimated as described before, but with the LT bootstrap we average across n sample observations as well as over M Monte Carlo trials.[17]

In order to minimize computational costs, we consider only production of a single output from a single input so that in terms of our earlier notation, $p = q = 1$. In all cases, we use the homogeneous bootstrap described in section 4.3.1 since we generate the data under the homogeneity condition described by (4.18).

For the case of CRS, we employed the ***true model***

$$y = xe^{-|v|}, \tag{6.1}$$

where $v \in N(0, 1)$ and $x \in$ Uniform(1,9). Draws from the uniform distribution were simulated using a multiplicative congruential pseudo random number generator with modulus $2^{31} - 1$ and multiplier 7^5 (see Lewis *et al.* (1969)). Pseudo random normal deviates were generated via the Box-Muller method (see Press *et al.* (1986, pp. 202–203)). In all of our experiments where the true model is given by (6.1), the fixed point is set at $x_0 = 5$, $y_0 = 2$ on each Monte Carlo trial. Given the true model in (6.1), the ***true*** input technical efficiency for this point, measured in terms of the Shephard input distance function (which is the reciprocal of the Farrell input measure of efficiency), is necessarily $\delta = 1/\theta = x_0/y_0 = 2.5$.

To consider cases involving variable returns to scale (VRS), we employed

$$y = (x - 2)^{2/3}e^{-|v|} \tag{6.2}$$

as the true model, where $v \sim N(0, 1)$. Data for experiments where (6.2) was used as the true model were simulated by drawing output values $y \sim$ Uniform$(0.5, 1.5)$, and then setting $x = y^{3/2}e^{\frac{3}{2}|v|} + 2$ after inverting (6.2) to obtain values for the inputs.[18] For these experiments, we chose $x_0 = 7.5$, $y_0 = 1.25$ as our fixed point in all Monte Carlo trials, so that the ***true*** efficiency measured by the Shephard input distance function is $\delta = 1/\theta = x_0/(y_0^{3/2} + 2) = 2.207$.

In all cases, confidence intervals are constructed here as described in section 4.1.2, and do not rely on the bias correction employed in Simar and Wilson (1998a). In our experiments where the true model exhibits CRS, we chose the bandwidth used in the kernel density estimator via the normal reference rule, which sets the bandwidth $h = 1.06\widehat{\sigma}n^{-1/5}$, where $\widehat{\sigma}$ is the square root of the estimated variance of the reflected efficiency scores from the initial estimation We use the normal reference rule here since with $p = q = 1$ and constant returns to scale, the (reflected about 1) estimated efficiency scores appear approximately normal in most cases; use of the normal reference rule, as opposed to cross-validation, reduces computational costs. In the experiments where the true model exhibits VRS, bandwidths for the kernel density estimator were chosen via weighted least-squares cross-validation as described in Simar and Wilson (1998b).[19]

In examining the VRS cases, we use the estimators defined above in section 2. For cases where the true technology exhibits CRS, we replace $\widehat{\Phi}_{DEA}(\mathcal{X}_n)$ with the conical hull of $\widehat{\Phi}_{FDH}(\mathcal{X})$, obtained by dropping the constraint $\sum_{i=1}^{n} \gamma_i = 1$ in (2.7).

6.2. *Results for the SW Bootstrap*

Results on coverage levels for the SW and LT bootstraps are given in Table 1; we discuss results for the SW bootstrap first. The first column of Table 1 gives the sample size in each of our Monte Carlo experiments. The second column gives 1 minus the nominal size of our confidence intervals. The third and fourth columns give the coverages for the SW bootstrap estimated from our Monte Carlo experiment as described above, for both the CRS and VRS cases.

The results indicate that for small sample sizes, the coverage of the confidence intervals estimated by the SW bootstrap is too small, but only slightly so, even with only 10 observations. With $n >= 100$, the coverage is almost exact; small deviations from the nominal sizes are to be expected due to sampling variation in the Monte Carlo exercises. Results for the VRS case are rather similar to those for the CRS case, although coverages for the VRS experiments are very slightly smaller than those obtained under CRS.

To further examine the performance of the SW bootstrap, we give information on the bootstrap estimates of bias and 95 percent confidence intervals in Table 2. For both the CRS and VRS cases, average bias estimates (column (1)) as well as the average difference between estimated bias and true bias (column (2)) shown to decrease as sample size increases. Similarly, the width of the estimated 95 percent confidence intervals (column (3)) decreases as sample size increases, becoming quite small (on average) at $n = 400$. For the CRS case, the average width of estimated 95 percent confidence intervals at $n = 400$ for the fixed point (5,2) is only 1.87 percent of the value of the true efficiency ($\delta = 2.5$) for this point. Similarly, for the VRS case, the average width of estimated 95 percent confidence intervals at $n = 400$ for the fixed point (7.5,1.25) is only 2.46 percent of the value of the true efficiency ($\delta = 2.207$) for this point.

Columns (4) and (5) in Table 2 give the ranges of the lower and upper bounds for estimated 95 percent confidence intervals, over the M Monte Carlo trials in our experiments. These ranges become smaller as sample size increases, consistent with our previous statements. Note that with $n = 10$, the estimated confidence intervals are highly variable, especially for the VRS case. In fact, under VRS, the upper and lower bounds of the estimated confidence intervals are sometimes less than unity, indicating that the fixed point (7.5,1.25) lies above $\widehat{\Psi}_{DEA}(\mathcal{X}_n)$ for some Monte Carlo trials. This merely confirms the slow convergence rate of the estimators; even with $p = q = 1$, $n = 10$ is a very small sample size, perhaps two small to obtain meaningful results in applied studies. Nonetheless, when n is increased

Table 1.

Monte Carlo Estimates of Confidence
Interval Coverages
One Input, One Output ($p = 1, q = 1$)

	1 − Nominal	SW		LT	
n	size	CRS	VRS	CRS	VRS
10	.80	0.693	0.571	0.711	0.443
10	.90	0.814	0.695	0.782	0.455
10	.95	0.886	0.773	0.783	0.457
10	.975	0.919	0.832	0.783	0.458
10	.99	0.942	0.873	0.783	0.458
25	.80	0.772	0.685	0.786	0.662
25	.90	0.883	0.805	0.878	0.680
25	.95	0.935	0.868	0.905	0.685
25	.975	0.973	0.916	0.915	0.687
25	.99	0.983	0.943	0.915	0.687
50	.80	0.784	0.729	0.793	0.773
50	.90	0.894	0.833	0.893	0.801
50	.95	0.940	0.839	0.939	0.807
50	.975	0.970	0.941	0.953	0.809
50	.99	0.985	0.974	0.959	0.809
100	.80	0.794	0.751	0.796	0.821
100	.90	0.911	0.853	0.896	0.866
100	.95	0.946	0.911	0.946	0.877
100	.975	0.973	0.948	0.970	0.881
100	.99	0.988	0.980	0.979	0.880
200	.80	0.810	0.708	0.798	0.845
200	.90	0.899	0.818	0.897	0.903
200	.95	0.946	0.902	0.947	0.921
200	.975	0.970	0.943	0.972	0.926
200	.99	0.994	0.968	0.985	0.927
400	.80	0.807	0.723	0.798	0.855
400	.90	0.903	0.839	0.898	0.922
400	.95	0.953	0.907	0.948	0.946
400	.975	0.977	0.946	0.973	0.954
400	.99	0.995	0.977	0.988	0.956

from 10 to 25, the average width of the estimated confidence intervals is reduced by almost half, and both the lower and upper bounds of the estimated 95 percent confidence intervals exceed unity in every Monte Carlo trial. Thus, in our experiments, with $n \geq 25$, the fixed point (7.5,1.25) lies below the VRS DEA frontier estimate in every Monte Carlo trial.

Table 2.

Monte Carlo Estimates of Bias, Confidence Intervals
Smooth (SW) Bootstrap
One Input, One Output ($p = 1, q = 1$)

n	(1)	(2)	(3)	(4)	(5)
Constant Returns to Scale:					
10	−0.3490	−0.0876	0.9110	1.1612, 2.5319	1.4601, 4.8300
25	−0.1906	−0.0754	0.5859	1.8824, 2.5090	2.1988, 3.6863
50	−0.1054	−0.0453	0.3512	2.0257, 2.5041	2.3093, 3.2033
100	−0.0535	−0.0239	0.1866	2.2598, 2.5016	2.4419, 2.8291
200	−0.0266	−0.0115	0.0952	2.4078, 2.5008	2.4840, 2.6417
400	−0.0129	−0.0054	0.0467	2.4237, 2.5004	2.4677, 2.5747
Variable Returns to Scale:					
10	−0.2984	0.0282	0.6075	0.3725, 2.4143	0.5506, 3.7469
25	−0.1736	−0.0200	0.3742	1.5515, 2.3070	1.8611, 3.0136
50	−0.1069	−0.0215	0.2373	1.6978, 2.2558	1.8455, 2.7885
100	−0.0642	−0.0107	0.1433	1.9336, 2.2331	2.0433, 2.5570
200	−0.0390	−0.0060	0.0858	2.0898, 2.2231	2.1610, 2.3930
400	−0.0241	−0.0043	0.0524	2.1491, 2.2129	2.1857, 2.2966

NOTE: Columns contain the following information:

(1) Average of bootstrap bias estimates over M Monte Carlo trials;
(2) Average of bootstrap bias estimate minus true bias over M Monte Carlo trials;
(3) Average width of estimated 95 percent confidence intervals over M Monte Carlo trials;
(4) Range of lower limits of estimated 95 percent confidence intervals over M Monte Carlo trials;
(5) Range of upper limits of estimated 95 percent confidence intervals over M Monte Carlo trials.

The true efficiency, measured by the Shephard output distance function, is 2.5 in the CRS case and 2.207 for the VRS case.

6.3. Results for the LT Bootstrap

The last two columns of Table 1 show the estimated coverages for the LT bootstrap. These results are roughly similar to estimated coverages obtained with the SW bootstrap shown in the same table and discussed above. Indeed, these results are similar to those reported by Löthgren (1998). These results tell only part of the story, however, and are not inconsistent with our earlier observations made in section 5.

Table 3 shows information on bias estimates and 95 percent confidence interval estimates obtained with the LT bootstrap, analogous to the information shown in Table 2 for the SW bootstrap discussed above. The ***average*** of bias estimates obtained with the LT bootstrap appear comparable to those from the SW bootstrap. However, the variances of the bias estimates obtained from the LT bootstrap over the M Monte Carlo trails is quite large; in numerous cases, the bias estimates have the wrong sign. Moreover, the average widths of the estimated confidence intervals obtained with the LT bootstrap are very large, and do not decrease as sample size increases.

Table 3.

Monte Carlo Estimates of Bias, Confidence Intervals
LT Bootstrap
One Input, One Output ($p = 1, q = 1$)

n	(1)	(2)	(3)	(4)	(5)	(6)
Constant Returns to Scale:						
10	−0.1485	0.1630	5.9144	−44.9008, 46.9008	1.0000, 92.8017	46.9008
25	−0.0709	0.0622	9.0854	−67.2309, 71.8270	1.0000, 137.4618	69.2309
50	−0.0368	0.0315	7.9999	−89.9321, 127.6011	0.9491, 187.0135	94.0068
100	−0.0196	0.0144	8.4093	−32.9001, 182.1128	0.9315, 191.8975	96.4531
200	−0.0097	0.0077	8.4106	−20.1716, 207.0097	0.9234, 215.4471	108.2308
400	−0.0051	0.0035	8.3667	−17.6605, 201.7171	0.9205, 212.7049	106.8705
Variable Returns to Scale:						
10	−0.2664	0.4315	2.3865	− 69.9353, 87.4752	1.0000, 173.9504	87.4752
25	−0.1865	0.2371	3.1187	− 98.4237, 107.0897	1.0000, 213.1749	107.0897
50	−0.1257	0.1315	3.2834	−173.3275, 184.1700	1.0000, 367.3401	184.1700
100	−0.0794	0.0829	3.2592	−326.3818, 341.3882	0.0319, 681.7764	341.3882
200	−0.0478	0.0468	3.0461	−332.0420, 353.3978	−9.9218, 705.7955	353.3978
400	−0.0285	0.0286	2.7247	−295.2114, 385.9263	−7.7177, 747.2589	385.9263

NOTE: Columns contain the following information:

(1) Average of bootstrap bias estimates over M Monte Carlo trials and n observations per trial;
(2) Average of bootstrap bias estimate minus true bias over M Monte Carlo trials and n observations per trial;
(3) Average width of estimated 95 percent confidence intervals over M Monte Carlo trials and n observations per trial;
(4) Range of lower limits of estimated 95 percent confidence intervals over M Monte Carlo trials and n observations per trial;
(5) Range of upper limits of estimated 95 percent confidence intervals over M Monte Carlo trials and n observations per trial;
(6) Maximum distance function estimate over M Monte Carlo trials and n observations per trial (minimum estimate is necessarily 1.0000 in all cases).

In these experiments, the true efficiency varies over observations as well as over Monte Carlo trials.

Even more revealing, the lower limits of the 95 percent confidence interval estimates obtained from the LT bootstrap range over very large intervals as indicated by column (4) of Table 3, and are frequently negative, even though the distance function estimates are (weakly) greater than one in every case. For the VRS case, the same result occurs for $n = 200$ and $n = 400$, but not for the smaller sample sizes. The upper limits of the 95 percent confidence interval estimates similarly range over very large intervals as indicted by column (5) in Table 3. These ranges include values far larger than the maximum distance function estimates (given in column (6)) over n observations and M Monte Carlo trials in each experiment.

When we look beyond the results reported in Löthgren (1998), we see that the LT bootstrap yields results that are nonsense, quite different from results obtained with the SW bootstrap. Confidence interval estimates from the LT bootstrap are extraordinarily wide, and so it is not surprising that they frequently include the true efficiency score. However, due to their excessive widths, the estimates are not informative.

7. Conclusions

Statistical inference is now possible when using DEA or FDH-type estimators in nonparametric frontier models. Those estimators are consistent under mild, general conditions, but the curse of dimensionality means that the obtained estimates should be taken with caution whenever estimates are obtained from moderate sample sizes.

Asymptotic sampling distributions for the efficiency estimators exist, but using these for statistical inference introduces additional noise to the presence of unknown parameters which must be estimated. Therefore, bootstrap methods seem to be an attractive alternative. Recently developed bootstrap algorithms are simple, but care should be used to avoid those that give inconsistent, useless approximations. The basic bootstrap methodology for DEA and FDH efficiency estimators has been extended to Malmquist productivity indices and their decompositions (Simar and Wilson, 1999c), and also to testing problems (Simar and Wilson, 1998b). Experimental evidence presented in this paper as well as our other papers indicates that when the sample size is appropriate to the dimensionality of the problem, our bootstrapping procedures give sensible, useful results.

Open issues for future research include (i) the search for asymptotic results for the DEA estimator when $p, q \geq 1$, (ii) reduction of statistical noise in the asymptotic results, and (iii) comparisons of the performances of the bootstrap and asymptotic theory. A solution to the first of these is necessary to allow conventional estimation of confidence intervals, *etc.*, but appears to be a very difficult problem. A quick reading of Gijbels et al. (1999) reveals the complexity of the analysis for the "simple" case where $p = q = 1$, and we can only expect the required level of complexity to increase for the more general multivariate setting where $p, q \geq 1$.

The asymptotic results that have obtained require estimation of unknown constants in order to use the results for making inferences. Existing estimation techniques that do not impose arbitrary structure on the problem yield noisy estimates, diminishing the usefulness of the asymptotic results for making inferences. Finally, comparisons of the performances of the bootstrap and asymptotic theory in inference-making would provide guidance on which method to use where either is possible. In a number of situations, bootstrap approximations are known to yield estimates with lower mean-square error than approximations base on asymptotic theory, but it remains unknown whether that is the case in the context of efficiency estimation.

We note that so far, we lack a rigorous proof of consistency for the SW bootstrap. Beran and Ducharme (1991), Hall (1992), and others have stated general conditions under which bootstrap estimators are consistent, but establishing these conditions for the bootstrap version of the DEA estimator is difficult and so far, at least, elusive. Nonetheless, our bootstrap does not suffer from the obvious sources of inconsistency that doom the various versions of the naive bootstrap that have been proposed for these problems. Moreover, Monte Carlo experiments we have conducted for this paper as well as others have, in every instance, supported the notion that our bootstrap is consistent. It is straightforward to extend our bootstrap for DEA estimators to FDH estimators, at least in principle, but we note here that the problem of bootstrapping in FDH is different, due to the differences between $\widehat{\partial X}_{DEA}$ and $\widehat{\partial X}_{FDH}$. Thus, a separate proof is required to confirm consistency of the bootstrap for FDH estimators.

While bootstrap methods offer a tractable approach to statistical inference in DEA or FDH models, a larger, remaining challenge is to find a way for allowing stochastic noise in the data in these models. We know that identification problems may exist if a purely nonparametric approach is maintained. Kneip and Simar (1996) propose one version of a stochastic DEA use with panel data, but this approach requires some unpleasant assumptions; in particular, inefficiency must be assumed constant over long time periods, which seems counter to implications of functioning, competitive markets. Apparently, there are ample opportunities for contributions to important problems in frontier estimation by careful researchers.

Notes

1. Standard microeconomic theory texts suggest that with competetitive input and output markets, firms which are either technically or allocatively inefficient will be driven from the market. These same texts, however, make no statements about how long this process might take. Indeed, in the real world, even where markets are highly competitive, there is no reason to believe that the market will eliminate inefficient firms instantaneously. Due to various frictions and imperfections in real markets for both inputs and outputs, the process whereby inefficient firms are eliminated might take many years. Moreover, firms that are initially inefficient in one or more respects may recover and begin to operate efficiently before they are driven from the market. Wheelock and Wilson (1995, 2000) provide support for this view through empirical evidence for commercial banks operating in the US.
2. In this paper our presentation is primarily in terms of inputs. The same could be done in the output-orientation where the output requirement sets would be defined for all $x \in \Psi$ as $Y(x) = \{y \in \mathbb{R}^q_+ \mid (x, y) \in \Psi\}$. Then the Farrell output measures of efficiency would be defined as

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

   Shephard input and output distance functions are merely the reciprocals of the corresponding Farrell measures. The results discussed below may also be extended to hyperbolic graph efficiency measures which consider simultaneous, proportionate reduction of inputs and expansion of outputs.
3. The "stochastic frontier" approach allows some observed points to be outside the attainable set, but there, only parametric restrictions on the shape of the frontier allows identification of the frontier from the sample. In addition, the stochastic frontier approach requires parametric assumptions regarding the distribution of both the inefficiency process as well as the noise process to recover estimates of firm-specific inefficiencies using cross-sectional data. Perhaps even more problematic, estimates of inefficiency from the stochastic frontier approach are inconsistent when cross sectional data are used, due to the conditioning on individual, composite error terms. This problem is avoided when there is no noise process, as in our statistical model which we define in the next subsection. With panel data, the stochastic frontier approach gives consistent estimates of inefficiency, but consistency only as the number of time periods approaches infinity. This has the unpleasant implication that inefficiency for individual firms remains constant over an infinite time horizon.
4. One can test whether the assumption of constant returns to scale is appropriate; see Simar and Wilson (1998b).
5. If one is willing to make parametric assumptions regarding the DGP, then statistical efficiency of the estimators could presumably be enhanced by incorporating this information into the estimation methods. But then the estimation method would be something other than DEA or FDH.
6. The presentation here is based on Kneip et al. (1998).

7. Throughout, inequalities involving vectors are defined on an element-by-element basis; e.g., for $\widetilde{x}, x \in \mathbb{R}^p$, $\widetilde{x} \geq x$ means that some, but perhaps not all or none, of the corresponding elements of $\widetilde{x}$ and x may be equal, while some (but perhaps not all or none) of the elements of $\widetilde{x}$ may be greater than corresponding elements of x.
8. Our characterization of the smoothness condition here is stronger than required; Kneip et al. (1998) require only Lipschitz continuity for $\theta(x, y)$, which is implied by the simpler, but stronger requirement presented here.
9. The point (x_0, y_0) might be chosen to correspond to one of the sample observations, or any other point in $\mathbb{R}_+^{p+q}$. However, if any element of y_0 exceeds the maximum of the corresponding elements of the y_i in $\mathcal{X}_n$, $\widehat{\theta}_{FDH}(x_0, y_0)$ will not be defined. Also, for purposes of (2.4), 0/0 is defined to equal 1. If $x_0^j = 0$ for all $j = 1, \ldots, p$ then $\widehat{\theta}_{FDH}(x_0, y_0)$ is undefined. If $x_0^j = 0$ for at least one $j = 1, \ldots, p$ and $x_i^j \neq 0$ for the same j and for all $i \in D(\mathcal{X}_n)$, then $\widehat{\theta}_{FDH}(x_0, y_0)$ is undefined. If $\widehat{\theta}_{FDH}(x_0, y_0)$ is undefined, then this means that points $(\theta x_0, y_0) \notin \widehat{\Psi}_{FDH}(\mathcal{X}_n)$ for all $\theta \in [0, \infty]$. This can occur, for instance, when production requires strictly positive quantities of all inputs.
10. Recall that the DGP defined by Assumptions (A1)–(A4) does not include a two-sided noise process. This feature, together with Assumption A2, determines the direction of the biases of $\widehat{\partial X}_{FDH}(y_0)$ and $\widehat{\theta}_{FDH}(x_0, y_0)$.
11. In order to simplify the presentation, we describe the varying returns to scale (VRS) version of the DEA estimator here. As observed in Simar and Wilson (1998b), this is the least restrictive version. All the arguments here can be adapted to any other convex estimator of Ψ which one might use depending on beliefs about returns to scale.
12. Exact, general statements on the number of observations required to achieve a given level of mean-square error are not possible since the exact convergence of the nonparametric estimators depends on unknown constants. Nonetheless, it is always true that for estimation purposes, more data are better than fewer data. In the case of nonparametric estimators such as $\widehat{\Psi}_{FDH}$ and $\widehat{\Psi}_{DEA}$, this statement is more than doubly true—it is exponentially true! To put the problem into economic terms, the law of diminishing marginal product applies to data used for estimation, but with nonparametric estimators, marginal product decreases at a rate that is exponentially slower than is the case for parametric estimation.
13. To put things yet another way, the points made here go to the heart of the tradeoff between nonparametric and parametric estimators. Parametric estimators incur the risk of misspecification, which typically results in inconsistency, but are almost always statistically more efficient than nonparametric estimators if properly specified. Nonparametric estimators avoid the risk of misspecification, but usually involve more noise than parametric estimators. Lunch is not free.
14. Recalling our remarks in footnote #3, we do not mean to suggest that it is meaningless to use "stochastic frontier" analysis (SFA). If the data contain noise, DEA and FDH estimators will be inconsistent, and there seems little choice but to rely on SFA, even though the SFA estimates of efficiency will be inconsistent in cross-sectional applications. Provided the parametric assumptions are made correctly, SFA will at least consistently estimate the frontier, unlike the DEA and FDH frontier estimators in the presence of noise.
15. Weiner (1998) cleverly noticed that the FDH efficiency estimator in the multivariate case can be cast in terms of a univariate problem, making analysis of its properties much easier than in the case of the DEA estimator. The ability to translate the multivariate FDH estimator to a univariate setting is due to the structure of the FDH frontier estimator; with the convexity added by the DEA estimator, this ability is lost. Analysis of the DEA estimator is more difficult because one must deal with facets of the estimated frontier which are convex combinations of several observations. Facets off the FDH frontier estimator, however, are always parallel to one or more axes in $\mathbb{R}_+^{p+q}$.
16. The issue of computational constraints continues to diminish in importance as computing technology advances. For many problems, however, a single desktop system may be sufficient. Given the independence of each of the B replications in the Monte Carlo approximation, the bootstrap algorithm is easily adaptable to parallel computing environments. Those who lack access to parallel supercomputers can efficiently and rather inexpensively mimic a parallel environment by using a series of networked personal computers; the price of these machines continues to decline while processor speeds increase.
17. In none of their papers that we have seen do Löthgren and Tambour discuss how one would use their bootstrap to examine efficiency for an arbitrary point (x_0, y_0) not contained in the data. Indeed, their presentations do not admit this possibility, although the logical extension would be to do as we describe in section 5. Our experiments with the LT bootstrap, however, are designed to remain consistent with what Löthgren and Tambour have written.

18. We invert (6.2) for purposes of simulating the data to ensure that our samples contain a wide range of values for y.
19. The weighting is necessary to avoid a discretization problem in the cross-validation procedure which results from numerous efficiency estimates at unity in the variable returns to scale case.

References

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

Banker, R. D. (1993). "Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation." *Management Science* 39(10), 1265–1273.

Charnes, A., W. W. Cooper, and E. Rhodes. (1978). "Measuring the Inefficiency of Decision Making Units." *European Journal of Operational Research* 2(6), 429–444.

Debreu, G. (1951). "The Coefficient of Resource Utilization." *Econometrica* 19(3), 273–292.

Deprins, D., L. Simar, and H. Tulkens. (1984). "Measuring Labor Inefficiency in Post Offices." In M. Marchand, P. Pestieau, and H. Tulkens (eds.), *The Performance of Public Enterprises: Concepts and Measurements*. 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*. CBMS-NSF Regional Conference Series in Applied Mathematics, #38. Philadelphia: SIAM.

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

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

Ferrier, G. D., and J. G. Hirschberg. (1997). "Bootstrapping Confidence Intervals for Linear Programming Efficiency Scores: With an Illustration Using Italian Bank Data." *Journal of Productivity Analysis* 8, 19–33.

Gijbels, I., E. Mammen, B. U. Park, and L. Simar. (1999). "On Estimation of Monotone and Concave Frontier Functions." *Journal of the American Statistical Association* 94, 220–228.

Grosskopf, S. (1996). "Statistical Inference and Nonparametric Efficiency: A Selective Survey." *Journal of Productivity Analysis* 7, 161–176.

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

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

Kneip, A., B. U. Park, and L. Simar. (1998). "A Note on the Convergence of Nonparametric DEA Estimators for Production Efficiency Scores." *Econometric Theory* 14, 783–793.

Koopmans, T. C. (1951). "An Analysis of Production as an Efficient Combination of Activities." In T. C. Koopmans (ed.), *Activity Analysis of Production and Allocation*. Cowles Commission for Research in Economics, Monograph 13. New York: John-Wiley and Sons, Inc.

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

Korostelev, A., L. Simar, and A. B. Tsybakov. (1995b). "On Estimation of Monotone and Convex Boundaries." *Publications de l'Institut de Statistique des Universités de Paris XXXIX* 1, 3–18.

Lewis, P. A., A. S. Goodman, and J. M. Miller. (1969). "A Pseudo-Random Number Generator for the System/360." *IBM Systems Journal* 8, 136–146.

Löthgren, M. (1997). "Bootstrapping the Malmquist Productivity Index: A Simulation Study." Working paper series in Economics and Finance #204, Department of Economic Statistics, Stockholm School of Economics, Sweden: forthcoming in *Applied Economics Letters*.

Löthgren, M. (1998). "How to Bootstrap DEA Estimators: A Monte Carlo Comparison." (contributed paper presented at the Third Biennial Georgia Productivity Workshop, University of Georgia, Athens, GA, October 1998), Working paper series in Economics and Finance #223, Department of Economic Statistics, Stockholm School of Economics, Sweden.

Löthgren, M., and M. Tambour. (1996). "Scale Efficiency and Scale Elasticity in DEA Models—A Bootstrapping Approach." Working paper series in Economics and Finance #91, Department of Economic Statistics, Stockholm School of Economics, Sweden: forthcoming in *Applied Economics*.

Löthgren, M., and M. Tambour. (1997). "Bootstrapping the DEA-based Malmquist Productivity Index." in

Magnus Tambour (Ph.D. dissertation), *Essays on Performance Measurement in Health Care*. Stockholm School of Economics, Stockholm, Sweden: forthcoming in *Applied Economics*.

Manski, C. F. (1988). *Analog Estimation Methods in Econometrics*. New York: Chapman and Hall.

Park, B., L. Simar, and C. Weiner. (1999). "The FDH Estimator for Productivity Efficiency Scores: Asymptotic Properties." *Econometric Theory*, forthcoming.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. (1986). *Numerical Recipes*. Cambridge: Cambridge University Press.

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

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.

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

Simar, L., and P. W. Wilson. (1998a). "Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models." *Management Science* 44(11), 49–61.

Simar, L., and P. W. Wilson. (1998b). "Nonparametric Tests of Returns to Scale, Discussion paper #9814." Institut de Statistique and CORE, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Simar, L., and P. W. Wilson. (2000). "A General Methodology for Bootstrapping in Nonparametric Frontier Models." *Journal of Applied Statistics*, forthcoming.

Simar, L., and P. W. Wilson. (1999a). "Some Problems with the Ferrier/Hirschberg Bootstrap Idea." *Journal of Productivity Analysis* 11, 67–80.

Simar, L., and P. W. Wilson. (1999b). "Of Course We Can Bootstrap DEA Scores! But Does It Mean Anything? Logic Trumps Wishful Thinking." *Journal of Productivity Analysis* 11, 93–97.

Simar, L., and P. W. Wilson. (1999c). "Estimating and Bootstrapping Malmquist Indices." *European Journal of Operations Research* 115, 459–471.

Weiner, H. C. M. (1998). *Nonparametric Statistical Analysis of Productivity and Efficiency with the Free Disposal Hull*. Doctoral Thesis, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Wheelock, D. C., and P. W. Wilson. (1995). "Explaining Bank Failures: Deposit Insurance, Regulation, and Efficiency." *Review of Economics and Statistics* 77, 689–700.

Wheelock, D. C., and P. W. Wilson. (2000). "Why Do Banks Disappear? The Determinants of US Bank Failures and Acquisitions." *Review of Economics and Statistics*, forthcoming.