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

Léopold SIMAR**

Paul W. WILSON***

March 1999

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 determinion 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 situation. Consequently, statistical inference in DEA/FDH-type models is now available. These techniques allow correction for the bias of the efficiency estimators and construction 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. Stressing the role of hypotheses and inference, we show how the results can be used or adapted for practical purposes.

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

---

*Invited paper presented at the Third Biennal 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).

**CORE and Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
Supported by "Projet d'Actions de Recherche Concertées" (No. 98/03–217)

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

1. 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 unit is analyzed. Shephard (1970) provides a modern economic formulation of the problem.

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[1] for a given point (x, y) is now defined as:

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

[1] 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.

$\theta(x, y)$ 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 productions 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.*).[2]

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?

[2]The "stochastic frontier" approach allows some 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.

- 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.[3]

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 on 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

[3] One can test whether the assumption of constant returns to scale is appropriate; see Simar and Wilson (1998d).

for future research.

1.2 A Data Generating Process $\mathcal{P}$

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.[4] Here, we present perhaps the simplest statistical model: the idea is to build, by assumptions on $\mathcal{P}$, a density $f(x, y)$ on Ψ generating independent, identically distributed (iid) observations $(x_i, y_i),\ i = 1, \ldots, n$.[5]

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* $\widetilde{y} \leq y$.[6]

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

---

[4] 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.

[5] The presentation here is based on Kneip *et al.* (1998).

[6] 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.

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

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

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.[7]

[7] 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.

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. Looking to (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 for going from $f(\theta, \eta, y)$ 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 $\mathcal{X}_n$:

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

Therefore, the estimator of input efficiency for a given point $(x, y) \in \Psi$ is:

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

From a computational viewpoint, we typically first define the set D of observed points ***dominating*** (x, y):

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

Then the input efficiency score of an arbitrary point[8] (x_0, y_0) is computed as

$$\widehat{\theta}(x_0, y_0) = \min_{i \in D} \max_{j=1,\dots,p} \left(\frac{x_i^j}{x_0^j}\right), \tag{2.4}$$

[8]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 $\mathcal{X}_n$, $\widehat{\theta}(x_0, y_0)$ will not be defined.

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

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

From a statistical viewpoint, we know by construction that $\widehat{\Psi}_{FDH} \subseteq \Psi$, and so $\widehat{\partial X(y)}$ is an upward-biased estimator of $\partial X(y)$ and $\widehat{\theta}(x_0, y_0)$ is an upward-biased estimator of $\theta(x_0, y_0)$.

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[9] of $\widehat{\Psi}_{FDH}$:

$$\begin{aligned}\widehat{\Psi}_{DEA}(\mathcal{X}_n) = \Big\{(x, y) \in \mathbb{R}^{p+q} \;|y \leq \sum_{i=1}^{n} \gamma_i y_i,\; x \geq \sum_{i=1}^{n} \gamma_i x_i, \\ \sum_{i=1}^{n} \gamma_i = 1,\; \gamma_i \geq 0\; \forall\, i = 1, \,\dots\,, n\Big\}.\end{aligned} \tag{2.6}$$

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

$$\widehat{\theta}(x_0, y_0) = \min\Big\{\theta > 0 \mid (\theta x_0, y_0) \in \widehat{\Psi}_{DEA}(\mathcal{X}_n)\Big\}. \tag{2.7}$$

As before, the estimation of the efficient level of inputs for the point (x_0, y_0) is given by:

$$\widehat{x}^\partial(y_0) = \widehat{\theta}(x_0, y_0) x_0. \tag{2.8}$$

Again, by construction, we know that $\widehat{\Psi}_{FDH} \subseteq \widehat{\Psi}_{DEA} \subseteq \Psi$, and so $\widehat{\partial X(y)}$ is an upward-biased estimator of $\partial X(y)$ and $\widehat{\theta}(x_0, y_0)$ is an upward-biased estimator of $\theta(x_0, y_0)$.

[9] 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 (1998d), 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.

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 the above estimators. Consider a fixed point $x \in \mathbb{R}^p,\ y \in \mathbb{R}^q$ and suppose the DGP is defined by $\mathcal{P}$ with Assumptions A1 (or A1′) through A4. For the DEA estimator we need assumption A1′, whereas all the properties below for the FDH estimators hold with either A1 or A1′.

2.3.1 What about consistency?

The first results on consitency 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) shows consistency of DEA input efficiency scores for this case, *i.e.,*

$$\widehat{\theta}_{DEA}(x, y) \rightarrow \theta(x, y) \text{ as } n \rightarrow \infty, \tag{2.9}$$

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 Ψ itself, and obtain

$$\widehat{\Psi}_{FDH},\ \widehat{\Psi}_{DEA} \rightarrow \Psi \text{ as } n \rightarrow \infty. \tag{2.10}$$

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

$$d_{\Delta}(\widehat{\Psi}_{FDH}, \Psi) = O_p\left(n^{-\frac{1}{q+1}}\right) \tag{2.11}$$

and

$$d_{\Delta}(\widehat{\Psi}_{DEA}, \Psi) = O_p\left(n^{-\frac{2}{q+2}}\right), \tag{2.12}$$

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, FDH and DEA estimators share some optimality properties in terms of their rates of convergence.

Unfortunately, the curse of dimensionality means here 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.

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

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

$$\widehat{\theta}_{DEA}(x_0, y_0) - \theta(x_0, y_0) = O_p\left(n^{-\frac{2}{p+q+1}}\right), \tag{2.13}$$

while Park *et al.* (1997) derive similar results for the FDH estimator:

$$\widehat{\theta}_{FDH}(x_0, y_0) - \theta(x_0, y_0) = O_p\left(n^{-\frac{1}{p+q}}\right). \tag{2.14}$$

These results again reflect the curse of dimensionality problem, which is worse for the multivariate case since the convergence rates are affected by $p + q$ rather than merely q as in (2.11)–(2.12).

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 to be made—one must use the FDH estimator, as the DEA estimator will be inconsistent. However, it Ψ is convex, the DEA estimator is prefered 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. Consistency is, however, a minimal theoretical property; from the preceeding discussion, we know that DEA or FDH efficiency estimators converge as the sample size increases (although at perhaps a slow rate), but by itself, this result has little practical use other than to confirm that the DEA or FDH estimators are perhaps reasonable to use for efficiency estimation.

For applied research, more is needed—in particular, the applied researcher must have 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.

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

Gijbels *et al.* (1999) obtain the following asymptotic result:

$$\widehat{\theta}_{DEA}(x,y) - \theta(x,y) \overset{\text{asy.}}{\sim} F(\cdot,\cdot) \tag{2.15}$$

where F is a known distribution function depending on unknown quantities given by the DGP $\mathcal{P}$ (*e.g.,* curvature of the frontier and the value of $f(x,y)$ at the true frontier point; see Gijbels *et al.* for additional details, including tabulation of values for $F(\cdot,\cdot)$).

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

For the general multivariate case, Park *et al.* (1997) derive the following result:

$$\widehat{\theta}_{FDH}(x,y) - \theta(x,y) \overset{\text{asy.}}{\sim} \text{Weibull}(\cdot,\cdot) \tag{2.16}$$

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, (2.15) and (2.16) allow one to compute the asymptotic value of $E(\widehat{\theta}(x,y))$, which provides a way to estimate and perhaps correct the bias of the estimator $\widehat{\theta}(x,y)$. The asymptotic variance $VAR(\widehat{\theta}(x,y))$ can also be computed, and provides an indication of the sampling variability of the estimators. Finally, asymptoytic confidence intervals for $\theta(x,y)$ may be constructed.

Unfortunately, (2.15) and (2.16) 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. Finite-sample properties of the approximations have been investigated via Monte-Carlo simulations in Gijbels *et al.* (1999) and Park *et al.* (1997).

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: we have ***only*** asymptotic results, additional noise is introduced when estimating parameters of the limiting distributions, and in the DEA case we only have the univariate result.

As in other situations where analytic results are not comforting, the bootstrap (Efron, 1979, 1982) might offer an attractive alternative to making inferences. Simar and Wilson (1998a) propose an algorithm to correctly implement the bootstrap in the context of frontier estimation. The algorithm provides an approximation of the sampling distribution of $\widehat{\theta}(x,y)-\theta(x,y)$ in the very general situation where $p,q \geq 1$ and for both the DEA and the 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, that due to the frontier nature of the problem, implementation of the bootstrap requires some adjustments from more typical applications of bootstrap methods to obtain ***consistent*** bootstrap approximations. Unfortunately, as also discussed below, some recently published papers have proposed bootstrap procedures that provide inconsistent (and hence ***useless***) approximations of the desired distributions.

3. More about Asymptotic Sampling Distributions

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

Since $q = 1$, under assumption A1$'$, the frontier can be described by a monotone, concave production function $y = g(x)$. We will present the result in term of the output-oriented case where the Farrell output efficiency score of a point (x_0, y_0) is defined by

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

where

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

For the fixed point (x_0, y_0). The DEA estimators are:

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

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

We know that $\widehat{g}(x_0) \to g(x_0)$ 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 the following:

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

$$\varphi(u,z) = (1/2)(-z)^{3/2}(1+u^2)\exp\{-(1/6)(-z)^{3/2}(u+u^{-1})^3\},$$

$$b_0 = f(x_0, g(x_0)),$$

$$b_2 = -g''(x_0)/2.$$

The leading term in the right-hand side of (3.5) is known; see Gijbels *et al.* (1999). The asymptotic bias and standard deviation of $\widehat{g}(x_0)$ can also be easely 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 b_2 and b_0, *i.e.,* the curvature and the density at the frontier point $(x_0, g(x_0))$. In particular, it appears that the bias is much more important than 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). Then, we can construct a bias-corrected estimator of the frontier, or 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 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 confidence intervals behave reasonably well.

Of course, the above results can also be translated into terms of the efficiency score $\lambda(x_0, y_0)$ 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.* (1997) obtain the following theorem, which is written in terms of the Farrell ouput-efficiency scores:

Theorem 2: *Under the assumptions A1–A4,*

$$\lambda(x_0, y_0) - \widehat{\lambda}(x_0, y_0) \overset{\text{asy.}}{\sim} \textit{Weibull}(\mu n^{\frac{1}{p+q}}, p+q). \tag{3.8}$$

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:

$$\text{asymp. bias of } \widehat{\lambda}(x_0, y_0) = -n^{-1/(p+q)}\mu^{-1}c_1 \tag{3.10}$$

and

$$\text{asymp. std. deviation of } \widehat{\lambda}(x_0, y_0) = n^{-1/(p+q)}\mu^{-1}\left(c_2 - c_1^2\right)^{1/2}, \tag{3.11}$$

where $c_j = \Gamma\left(\frac{p+q+j}{p+q}\right)$, $j = 1, 2$. Again, (3.10) and (3.11) allow us to appreciate the role of the sample size n and of the parameter μ; in particular, it appears that the bias is more important than standard deviation in determining mean square error (and still more so if $p + q$ is large).

From a practical point of view, the parameter μ must be estimated. Park *et al.* (1997) propose a nonparametric, consistent estimator of μ. This allows one to construct asymptotic, bias-corrected estimators of the efficiency measure $\lambda(x_0, y_0)$ as well as confidence intervals for $\lambda(x_0, y_0)$, using the percentiles of the Weibull distribution.

Finite-sample properties of these estimators are analyzed in Park *et al.* (1997). For small values of $p+q$ (*e.g.,* 2 or 3), they behave rather well with $n = 100$ or 200, but performance deteriorates when 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 simulate and thereby approximate the sampling distributions of interest by 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) and Ferrier and Hirshberg (1997). Unfortunately, the bootstap procedure gives inconsistent estimates if the so-called "naive" bootstrap or one of its variants is used as in Ferrier and Hirshberg (1997) (see Simar and Wilson, 1999a and 1999b, for more details on this particular issue).

4.1 General Principles

We will make the presentation in the input-oriented efficiency measure in order to use the notation introduced above. We will also use the DEA estimators; the 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 $\widehat{\theta}(x_0, y_0)$.

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 boostrap world, $\widehat{\Psi}$ is the true attainable set, and the convex hull of $\mathcal{X}_n^*$, $\widehat{\Psi}^*(\mathcal{X}_n^*)$, gives an estimator of $\widehat{\Psi}$:

$$\begin{aligned}\widehat{\Psi}^* = \{(x, y) \in \mathbb{R}^{p+q}\ |y \leq \sum_{i=1}^{n} \gamma_i y_i^*; \quad x \geq \sum_{i=1}^{n} \gamma_i x_i^* \\ \sum_{i=1}^{n} \gamma_i = 1; \quad \gamma_i \geq 0\ \forall\ i = 1, \ldots, n\}\end{aligned} \tag{4.3}$$

For a fixed point (x_0, y_0),

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

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

$$\begin{aligned}\widehat{\theta}^*(x_0, y_0) = \min\Big\{\theta > 0\ |y_0 \leq \sum_{i=1}^{n} \gamma_i y_i^*;\ \theta x_0 \geq \sum_{i=1}^{n} \gamma_i x_i^*; \\ \sum_{i=1}^{n} \gamma_i = 1;\ \gamma_i \geq 0\ \forall\ i = 1,\ \ldots, n\Big\}.\end{aligned} \tag{4.5}$$

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

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

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 $\widehat{\theta}^*(x_0, y_0)$ 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 $\widehat{\theta}_b^*(x_0, y_0)$, $b = 1, \ldots, B$. The empirical distribution of these bootstrap values gives a Monte Carlo approximation of the sampling distribution of $\widehat{\theta}^*(x_0, y_0)$, 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.[10] 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 $n \to \infty$.

[10]The issue of computational constaints 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.

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 (1998b), 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 $\left(\widehat{\theta}(x_0, y_0) - \theta(x_0, y_0)\right)$, then it would be trivial to find values a_α and b_α such that

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

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

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

Finding $\widehat{a}_\alpha$ and $\widehat{b}_\alpha$ involves sorting the values $\left(\widehat{\theta}_b^*(x_0, y_0) - \widehat{\theta}(x_0, y_0)\right)$, $b = 1, \ldots, B$ 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 $\widehat{a}_\alpha \leq \widehat{b}_\alpha$.

The bootstrap approximation of (4.7) is then

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

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

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

This procedure can be used for any $(x_0, y_0) \in \mathbb{R}_+^{p+q}$ for which $\widehat{\theta}(x_0, y_0)$ 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,

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

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

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

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

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

It is well known (see e.g. Efron and Tibshirani, 1993), however, that this bias correction introduces additional noise; the mean-square error of $\widehat{\widehat{\theta}}(x_0, y_0)$ may be greater than the mean-square error of $\widehat{\theta}(x_0, y_0)$. 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 $\widehat{\widehat{\theta}}(x_0, y_0)$ will have variance equal to four times that of the original, uncorrected estimator, $\widehat{\theta}(x_0, y_0)$, (again illustrating the fact that the bootstrap is an asymptotic procedure). The sample variance of the bootstrap values $\widehat{\theta}_b^*(x_0, y_0)$ gives an estimate $\widehat{\sigma}^2$ of the variance of $\widehat{\theta}(x_0, y_0)$:

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

Hence the bias-correction should not be used unless

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

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,

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

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 Consistent Bootstrap

Two approaches have been proposed. 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 (1998c), 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 $\widehat{\theta}_i$, 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 compution 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

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

In the homogeneous case, it seems natural to estimate $f(\theta \mid \eta, y) = f(\theta)$ from the set $\{\widehat{\theta}_i \mid i = 1, \ldots, n\}$. 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 $\widehat{\theta}_i$, combined with the fact that at least one of the $\widehat{\theta}_i$ 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 $\widehat{\theta}_i$ 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 θ_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 impied 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 resample 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 (1998b, 1998c, 1998d).

4.3.2 The Heterogeneous Bootstrap

Bootstrapping on the pairs 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 $\mathcal{X}_n$. 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 probabilty 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 probabilty 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 homogenous 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 (1998c).

5. Yet Another Inconsistent 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 $x_i^* = x_i\widehat{\theta}_i/\theta_i^*$ 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

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

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 $\theta(x_0, y_0)$ for $(x_0, y_0) \in \Psi$ 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

We conducted a series of Monte Carlo experiments to measure the performance of the bootstrap under both constant and variable returns to scale. Each Monte Carlo experiment involved M trials; in each trial, we simulated data from a true, known model. To examine performance of the Simar/Wilson (SW) bootstrap, on each trial we estimated efficiency for a fixed point (x_0, y_0) not necessarily contained in the data, and then bootstrapped to obtain confidence intervals for the corresponding efficiency. 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 count the number of trials in which the true efficiency falls within the estimated confidence intervals. We then divide this count by M and report these values in Table 1. Hence, for example, at nominal size equal to .05, estimated confidence intervals should include the true efficiency score in 95 percent of the Monte Carlo trials.

In order to minimize computational costs, we only consider 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). The number of Monte Carlo trials, M, was 1000 in each experiment. The number of bootstrap replications on each Monte Carlo trial was set at 2000 in all cases.

6.1 Case I: Constant Returns to Scale

For the case of constant returns to scale (CRS), we adopt 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 each Monte Carlo trial, the fixed point is given by $x_0 = 5$, $y_0 = 2$. 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 then $\delta = x_0/y_0 = 2.5 = 1/\theta$.

The first column of Table 1 gives the sample size in each of our Monte Carlo experiments. The second column gives the nominal size of our confidence intervals. The third column gives the size estimated from our Monte Carlo experiment as described above for the SW bootstrap.[11] In these experiments, 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.[12] The results indicate that for small sample sizes, the coverage of the confidence intervals is too small, although 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.

Similar experiments using the LT bootstrap produced coverage probabilities similar to those reported in Löthgren (1998). In our experiments with the LT bootstrap, we let the fixed point (x_0, y_0) correspond in turn to each of the n sample observations, and then averaged results over firms as well as Monte Carlo trials. Even with $n = 400$, the LT bootstrap yielded bias estimates as defined in (4.12) ranging from -96.93 to 2.51, with a standard deviation of 2.56, and a positive estimate of bias in 69 percent of all cases even though the true bias is negative. The realized bias (given by the distance function estimate minus its true value), on the other hand, ranges from -1.1864 to 0.0, with a

[11] 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 (1998).

[12] 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.

standard deviation of 0.015. The SW bootstrap, by contrast and by design, can never produce a positive bias estimate for Shephard input distance function estimators. Our experiments using the SW bootstrap yield an average bias estimate of -0.0129; the SW bias estimates range from -0.0237 to -0.0097, with a standard deviation of 0.001196—an order of magnitude smaller than with the LT bootstrap! Moreover, differences between bias estimates from the SW bootstrap and the corresponding true values did not exceed 0.0645 in absolute value across the 1000 Monte Carlo trials in the experiment.

Similarly, even though the true Shephard input distance function is always grater than or equal to one, the left boundaries of the LT-estimated 95-percent confidence intervals are less than one in 89.8 percent of all cases, and less than zero in 87.4 percent of all cases. The widths of the LT-estimated 95-percent confidence intervals range from 4.794 to 18.862, with a median of 8.231 and a standard deviation of 1.28. Thus, even with $n = 400$, the LT bootstrap provides no useful information on the true efficiencies confirming our reasoning in Section 5, and gives improbable, (indeed, ***impossible***) estimates in more than half of all cases. This is not apparent from examining the tables in Löthgren (1998), where only the estimated coverages of the bootstrap confidence intervals are given.

The SW bootstrap, as revealed by our experiments, performs much better. For the fixed point considered there, with $n = 400$, the SW-estimated 95-percent confidence intervals have left bounds ranging from 2.4237 to 2.5004 (recall that the true efficiency score in this experiment is 2.5), while the right bounds range from 2.4677 to 2.5747. The widths of the SW-estimated 95 percent confidence intervals range from 0.0355 to 0.0819 across the 1000 Monte Carlo trials, with a standard deviation of 0.00448—3 orders of magnitude smaller than with the LT bootstrap! For the case with only 10 observations, the SW bootstrap yields estimated 95-percent confidence intervals with widths ranging from 0.2989 to 1.7786; the widths have a standard deviation of 0.2567 across the Monte Carlo trials with $n = 10$—still an order of magnitude smaller than the LT bootstrap with $n = 400$!

6.2 Case II: Variable Returns to Scale

We repeated the experiment outlined in the previous section while considering variable returns to scale. Here, the true model is

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

where $v \sim N(0, 1)$. The data were simulated by drawing output values $y \sim$ Uniform(0.5, 1.5) and setting $x = y^{3/2} e^{\frac{3}{2}|v|} + 2$ after inverting (6.2) to obtain values for the inputs.[13] We chose $x_0 = 7.5$, $y_0 = 1.25$ as our fixed point, so that the ***true*** efficiency measured by the Shephard input distance function is $\delta = x_0/(y_0^{3/2} + 2) = 2.207 = 1/\theta$. Bandwidths for the kernel density estimator were chosen via weighted least-squares cross-validation as described in Simar and Wilson (1998d).[14]

Results for the SW bootstrap in the variable returns to scale case are given in the fifth column of Table 1. The performance here is similar to what we saw with constant returns to scale in section 6.1.1 and in the third column of Table 1. Coverages are slightly smaller with variable returns to scale than with constant returns to scale for a given nominal size, reflecting the fact that here we are estimating a more complicated frontier. Even with $n = 400$, the coverages remain smaller than desired, although only slightly.

We performed similar experiments allowing for variable returns to scale using the LT bootstrap, but again letting the fixed point (x_0, y_0) correspond to each of the n firms in a given sample and then averaging over firms as well as Monte Carlo trials. As before, these experiments produced estimates of coverage probabilities similar to those reported by Löthgren (1998). But, as in the previous case, the coverage probabilities tell only a small part of the story. As discussed above, the Shephard input distance function estimator is necessarily biased downward. Yet, even for $n = 400$, the LT bootstrap yields bias estimates ranging from -294.8966 to 3.5526 with a standard deviation of 4.6097, while the distance

[13] In this DGP, we allow x to be determined by y and v, rather than having y determined by x and v as in the previous case, in order to ensure that our samples contain a wide range of values for y.

[14] 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.

function estimator minus the true distance function value ranges from -17.4882 to -0.0001, with a standard deviation of only 0.2329. Over 75 percent of the bias estimates from the LT bootstrap are positive—*i.e.,* in the wrong direction. By construction, this cannot happen, and does not happen, with the SW bootstrap. Estimated 95-percent confidence intervals from the LT bootstrap are similarly ridiculous: with $n = 400$, 99 percent of those estimated in our experiment have a width of 6.74 or more; 90 percent of those estimated have a width of 14.9 or more. By contrast, 95-percent confidence intervals estimated by the SW bootstrap have a ***maximum width*** of 0.0926 when $n = 400$. Aside from lacking any logical foundation, the LT bootstrap does not provide useful information about the true inefficiencies (or anything else).

7. Conclusions

Statistical inference is now possible when using DEA or FDH-type estimators in non-parametric frontier models. Those estimators are consitent 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 inconsitent, 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, 1998b), and also to testing problems (Simar and Wilson, 1998d).

Open issues for future research include the search for asymptotic results for the DEA estimator when $p, q \geq 1$, reduction of statistical noise in the asymptotic results, and comparisons of the performances of the bootstrap and asymptotic theory. 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.

While bootstrap methods offer a tractible 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. Apparently, there are many remaining problems to be solved by careful researchers in frontier estimation.

REFERENCES

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 ressource utilization, *Econometrica* 19(3), 273–292.

Deprins, D., L. Simar, and H. Tulkens (1984), "Measuring Labor Inefficiency in Post Offices," in *The Performance of Public Enterprises: Concepts and Measurements*, ed. by M. Marchand, P. Pestieau and H. Tulkens. 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*, forthcoming.

Grosskopf, S. (1996), Statistical inference and nonparametric efficiency: A selective survey, *Journal of Productivity Analysis* 7, 161–176.

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 *Activity Analysis of Production and Allocation*, ed. by T.C. Koopmans, Cowles Commision 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, Stockhold 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 Biennal Georgia Productivity Workshop, University of Georgia, Athens, GA, October 1998), Working paper series in economics and Finance #223, Department of Economic Statistics, Stockhold 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, Stockhold School of Economics, Sweden; forthcoming in *Applied Economics.*

Löthgren, M., and M. Tambour (1997), "Bootstrapping the DEA-based Malmquist Productivity Index," in *Essays on Performance Measurement in Health Care,* Ph.D. dissertation by Magnus Tambour, 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 (1997), "FDH Efficiency Scores from a Stochatic Point of View," Discussion paper #9715, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

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

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), Estimating and bootstrapping Malmquist indices, *European Journal of Operations Research*, forthcoming.

Simar, L., and P.W. Wilson (1998c), "A General Methodology for Bootstrapping in Nonparametric Frontier Models," Discussion paper #9811, Institut de Statistique and CORE, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Simar, L., and P.W. Wilson (1998d), 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 (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.

TABLE 1

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

n	1 − Nominal size	CRS	VRS
10	.80	0.693	0.571
10	.90	0.814	0.695
10	.95	0.886	0.773
10	.975	0.919	0.832
10	.99	0.942	0.873
25	.80	0.772	0.685
25	.90	0.883	0.805
25	.95	0.935	0.868
25	.975	0.973	0.916
25	.99	0.983	0.943
50	.80	0.784	0.729
50	.90	0.894	0.833
50	.95	0.940	0.839
50	.975	0.970	0.941
50	.99	0.985	0.974
100	.80	0.794	0.751
100	.90	0.911	0.853
100	.95	0.946	0.911
100	.975	0.973	0.948
100	.99	0.988	0.980
200	.80	0.810	0.708
200	.90	0.899	0.818
200	.95	0.946	0.902
200	.975	0.970	0.943
200	.99	0.994	0.968
400	.80	0.807	0.723
400	.90	0.903	0.839
400	.95	0.953	0.907
400	.975	0.977	0.946
400	.99	0.995	0.977