Performance of the Bootstrap for DEA Estimators and Iterating the Principle

Léopold Simar*

and

Paul W. Wilson**

December 1999

ABSTRACT

This paper further examines the bootstrap method proposed by Simar and Wilson (1998) for DEA efficiency estimators. Some simplifications are provided, and we provide Monte Carlo evidence on the coverage probabilities of confidence intervals estimated by the method. In addition, we provide similar evidence for confidence intervals estimated with the so-called naive bootstrap, which is known to be inconsistent in the DEA setting. Finally, we propose an interated version of the bootstrap which may be used to improve bootstrap estimates of confidence intervals.

Keywords: data envelopment analysis, bootstrap, distance function, efficiency, frontier models.

---

*Institut de Statistique, Université Catholique de Louvain, Voie du Roman Pays 20, Louvain-la-Neuve, Belgium. Research support from the contract "Projet d'Actions de Recherche Concertées" (PARC No. 98/03–217) of the Belgian Government is gratefully acknowledged.

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

1. INTRODUCTION

A large literature concerning the measurement of efficiency in production has developed since Debreu (1951) and Farrell (1957) provided basic definitions for technical and allocative efficiency in production. Approaches to efficiency estimation typically fall into one of two categories: (i) parametric approaches, including models with a noise term along the lines suggested by Aigner *et al.* (1977), Meeusen and van den Broeck (1977), and Jondrow *et al.* (1982), as well as models without a noise term such as those suggested by Winsten (1957), Greene (1980a,b), and Stevenson(1980); and (ii) nonparametric approaches such as those proposed by Charnes *et al.* (1978, 1979), Färe and Lovell (1978), Burley (1980), Zieschang (1984), Deprins *et al.* (1984), Banker *et al.* (1984), and Färe *et al.* (1985). The nonparametric approaches which rely on convexity assumptions have been termed Data Envelopment Analysis (DEA), and typically rely on linear programming (LP) techniques for solution. The Free Disposal Hull (FDH) estimator introduced by Deprins *et al.* (1984) does not impose convexity and may be written as an LP, although LP methods are typically not used in computing FDH efficiency estimates.

The parametric approaches have been widely applied to examine various notions of efficiency in a variety of industries. These approaches are frequently referred to as *econometric* approaches, and researchers using them have at their disposal a large array of statistical tools which provide aid in diagnosing estimation results, making statistical inferences, *etc.* Many of these tools are based on results originally obtained by Karl Gauss in the 19th century and by Ronald Fisher in the early 20th century—parametric regression has been used for almost 200 years, and so there has been time for many different statistical tools to be developed for use with these methods.

The nonparametric approaches to efficiency estimation have been perhaps even more widely applied (see Seiford, 1997, for an extensive bibliography; see also the survey by Lovell, 1993), but are comparatively young. The nonparametric approaches have been characterized as *deterministic* (as opposed to *econometric*), as if to suggest that the methods lack any statistical underpinnings. Studies which use these methods have typically presented point estimates of inefficiency, with no measure or even discussion of uncertainty surrounding these estimates. Indeed, many papers contain statements where efficiency is described as being *computed* as opposed to being *estimated,* and results are freqently referred to as *efficiencies* rather than *efficiency estimates.*

The choice of terminology in describing the nonparametric efficiency approaches and their results is understandable given the (until very recently) lack of a "tool box" with aids for diagnostics, inference, *etc.* such as the one available to researchers using the parametric approaches. But the terminology is also unfortunate, because it has served to cloud important issues in efficiency estimation, and has contributed to misunderstandings between researchers using the parametric approaches and those using the nonparametric approaches. Many readers will recall that the first venues where researchers from the parametric and nonparametric "camps" were brought together (the early productivity workshops organized in the 1980s by Ali Dogramaçi at Rutgers University and the 1988 NSF conference at the University of North Carlolina organized by Knox Lovell) were sometimes quite rancorous affairs.

Today, there is much more civility and good humor. In addition, differences between the two approaches which once seemed vast now seem much less so due to recent developments in the literature. Researchers using the nonparametric approaches now have access to a tool box, into which a number of tools based on bootstrap methods have recently been added. This paper summarizes some of those tools, and shows how the bootstrap principle may be iterated to improve confidence interval estimates. Section 2 briefly reviews the microeconomic theory of the firm, and establishes our notation and an economic model. Section 3 discusses estimation of efficiency, while section 4 establishes a statistical model by augmenting the economics assumptions in section 2 with some assumptions on the data generating process (DGP). Section 5 briefly reviews existing asymptotic results, which have implications for the bootstrap. The DEA bootstrap is discussed generally in section 6, and more specifically in section 7, where we give details on its implementation. Section 8 details our Monte Carlo experiments and their results. Section 9 offers an iterated version of the bootstrap which can be used to improve estimates of confidence intervals. The final section offers some concluding remarks.

2. EFFICIENCY AND THE THEORY OF THE FIRM

Standard microeconomics texts develop the theory of the firm by positing a production set which describes how a set of inputs may be somehow converted into outputs. To illustrate, let $\boldsymbol{x} \in \mathbb{R}_+^p$ denote a vector of p inputs and $\boldsymbol{y} \in \mathbb{R}_+^q$ denote a vector of q outputs. Then the production set may be defined as

$$\mathcal{P} \equiv \{(\boldsymbol{x}, \boldsymbol{y}) | \boldsymbol{x} \text{ can produce } \boldsymbol{y}\}, \tag{2.1}$$

which is merely the set of feasible combinations of $\boldsymbol{x}$ and $\boldsymbol{y}$. The production set $\mathcal{P}$ is sometimes described in terms of its sections

$$\mathcal{Y}(\boldsymbol{x}) \equiv \{\boldsymbol{y} | (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}\} \tag{2.2}$$

and

$$\mathcal{X}(\boldsymbol{y}) \equiv \{\boldsymbol{x} | (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}\}, \tag{2.3}$$

which form the output feasibility and input requirement sets, respectively. Knowledge of either $\mathcal{Y}(\boldsymbol{x})$ for all $\boldsymbol{x}$ or $\mathcal{X}(\boldsymbol{y})$ for all $\boldsymbol{y}$ is equivalent to knowledge of $\mathcal{P}$; $\mathcal{P}$ implies (and is implied by) both $\mathcal{Y}(\boldsymbol{x})$ and $\mathcal{X}(\boldsymbol{y})$. Thus, both $\mathcal{Y}(\boldsymbol{x})$ and $\mathcal{X}(\boldsymbol{y})$ inherit the properties of $\mathcal{P}$. Various assumptions regarding $\mathcal{P}$ are possible; we adopt those of Shephard (1970) and Färe (1988):

Assumption A1: *$\mathcal{P}$ is convex, $\mathcal{Y}(\boldsymbol{x})$ is convex, bounded, and closed for all $\boldsymbol{x} \in \mathbb{R}_+^p$, and $\mathcal{X}(\boldsymbol{y})$ is convex, bounded, and closed for all $\boldsymbol{y} \in \mathbb{R}_+^q$.*

Assumption A2: *$(\boldsymbol{x}, \boldsymbol{y}) \notin \mathcal{P}$ if $\boldsymbol{y} \geq 0, \boldsymbol{y} \neq 0$, i.e., all production requires use of some inputs.*

Assumption A2 merely says that there are no free lunches.[1]

[1] Throughout, inequalities involving vectors are defined on an element-by-element basis; *e.g.*, for $\widetilde{\boldsymbol{x}}$, $\boldsymbol{x} \in \mathbb{R}_+^p$, $\widetilde{\boldsymbol{x}} \geq \boldsymbol{x}$ means that some, but perhaps not all or none, of the corresponding elements of $\widetilde{\boldsymbol{x}}$ and $\boldsymbol{x}$ may be equal, while some (but perhaps not all or none) of the elements of $\widetilde{\boldsymbol{x}}$ may be greater than corresponding elements of $\boldsymbol{x}$.

Assumption A3: *for* $\widetilde{\boldsymbol{x}} \geq \boldsymbol{x}$, $\widetilde{\boldsymbol{y}} \leq \boldsymbol{y}$, *if* $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$ *then* $(\widetilde{\boldsymbol{x}}, \boldsymbol{y}) \in \mathcal{P}$ *and* $(\boldsymbol{x}, \widetilde{\boldsymbol{y}}) \in \mathcal{P}$, *i.e., both inputs and outputs are strongly disposable.*

Assumption A3 is sometimes called free disposability. The presentation in this section does not depend on these particular assumptions, and the estimators discussed in Section 3 can be adapted to cases where alternative assumptions might be warranted. For example, if pollution is an inadvertent byproduct of the production process, then it might not be reasonable to assume that this particular output is strongly (freely) disposable.

The boundary of $\mathcal{P}$ is sometimes referred to as the *technology* or *production frontier,* and is given by the intersection of $\mathcal{P}$ and the closure of its compliment, $\mathcal{P}^{\partial}$. Similarly, isoquants are defined by the intersection of $\mathcal{X}(\boldsymbol{y})$ and the closure of its complienent, or

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

In addition, the intersection of $\mathcal{Y}(\boldsymbol{x})$ and the closure of its compliment gives the iso-output curve,

$$\mathcal{Y}^{\partial}(\boldsymbol{x}) = \{\boldsymbol{y} \mid \boldsymbol{y} \in \mathcal{Y}(\boldsymbol{x}),\ \lambda\boldsymbol{y} \notin \mathcal{Y}(\boldsymbol{x})\ \forall\, \lambda > 1\}. \tag{2.5}$$

Firms which are *technically inefficient* operate at points in the interior of $\mathcal{P}$, while those that are technically *efficient* operate somewhere along the technology defined by $\mathcal{P}^{\partial}$. Various measures of technical efficiency are possible. The Shephard (1970) output distance function provides a normalized measure of Euclidean distance from a point $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$ to $\mathcal{P}^{\partial}$ in a direction orthogonal to $\boldsymbol{x}$, and may be defined as

$$D(\boldsymbol{x}, \boldsymbol{y}) \equiv \inf\{\theta > 0 | (\boldsymbol{x}, \theta^{-1}\boldsymbol{y}) \in \mathcal{P}\}. \tag{2.6}$$

Clearly, $D(\boldsymbol{x}, \boldsymbol{y}) \leq 1$ for all $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$. If $D(\boldsymbol{x}, \boldsymbol{y}) = 1$ then $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^{\partial}$; *i.e.,* that the point $(\boldsymbol{x}, \boldsymbol{y})$ lies on the boundary of $\mathcal{P}$, and the firm is technically efficient. One can similarly define the Shephard (1970) input distance function, which provides a normalized measure of Euclidean distance from a point $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$ to $\mathcal{P}^{\partial}$ in a direction orthogonal to $\boldsymbol{y}$. Alternatively, one could employ measures of hyperbolic graph efficiency defined by Färe *et* *al.* (1985), directional distance functions as defined by Färe *et al.* (1996), or perhaps other measures. From a purely technical viewpoint, any of these can be used to measure technical efficiency; the only real difference is in the direction in which distance to the technology is measured. However, if one wishes to take account of behavioral implications, then this might influence the choice between the various possibilities. We present our methodology only in terms of the output orientation to conserve space, but our discussion can be adapted to the other cases by straightforward changes in our notation. The Shephard output distance function is merely the reciprocal of the Farrell (1957) output efficiency measure, while the Shephard input distance funciton is the reciprocal of the Farrell input efficeincy measure.

In addition to technical efficiency, one may consider whether a particular firm is allocatively efficient. Assuming the firm is technically efficient, its input-allocative efficiency depends on whether it operates at the optimal location along one of the isoquants $\mathcal{X}^{\partial}(\boldsymbol{y})$ as determined by prevailing input prices. Alternatively, output-allocative efficiency depends on whether the firm operates at the optimal location along one of the iso-output curves $\mathcal{Y}^{\partial}(\boldsymbol{x})$ as determined by prevailing output prices. Färe *et al.* (1985) combine these notions of allocative efficiency with technical efficiency to obtain measures of "overall" efficiency. Here, we focus only on output technical efficiency, but it is straightforward to extend our discussion to the other notions of inefficiency.

Standard microeconomic theory suggests that with perfectly comptetitive input and output markets, firms which are either technically or allocatively inefficient will be driven from the market. However, in the real world, even where markets may be highly competitive, there is no reason to believe that this must happen instantaneously. Indeed, due to various frictions and imperfections in real markets for both inputs and outputs, this process might take many years, and 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 (1996, 1999) provide support for this view through empirical evidence for banks operating in the US.

3. ESTIMATION

Unfortunately, none of the theoretical items defined in the previous section are observed, including the productions set $\mathcal{P}$ and the output distance function $D(\boldsymbol{x}, \boldsymbol{y})$. Consequently, these must be *estimated.*

In the typical situation, all that is observed are inputs and outputs for a set of n firms; together, these comprise the observed sample:

$$\mathcal{S}_n = \{(\boldsymbol{x}_i, \boldsymbol{y}_i) \mid i = 1, \ \ldots, n\}. \tag{3.1}$$

These may be used to construct various estimators of $\mathcal{P}$, which in turn may be used to construct estimators of $D(\boldsymbol{x}, \boldsymbol{y})$. Charnes *et al.* (1978) used the conical hull of the free disposal hull of $\mathcal{S}_n$ to construct an estimator $\widehat{\mathcal{P}}_{CCR}$ of $\mathcal{P}$ in what has been called the "CCR model." This approach implicitly imposes an assumption of constant returns to scale on the production technology, $\mathcal{P}^\partial$. Banker *et al.* (1984) used the convex hull of the free disposal hull of $\mathcal{S}_n$ as an estimator $\widehat{\mathcal{P}}_{BCC}$ of $\mathcal{P}$. This approach, which has been termed the "BCC model," allows $\mathcal{P}^\partial$ to exhibit increasing, constant, or decreasing returns at different locations along the frontier; hence $\widehat{\mathcal{P}}$ is said to allow variable returns to scale. Other estimators of $\mathcal{P}$ are possible, such as the free disposal hull of $\mathcal{S}_n$ by itself (denoted $\widehat{\mathcal{P}}_{DST}$) as in Deprins *et al.* (1984), which in effect relaxes the assumption of convexity on $\mathcal{P}$.

Terminology used in the DEA literature is sometimes rather confusing and misleading, and the preceeding discussion offers an example. The terms "CCR model" and "BCC model" are misnomers; the conical and convex hulls of the free disposal hull of $\mathcal{S}_n$ are merely *different estimators* of $\mathcal{P}$—not different models. Thus, we will refer to the CCR *estimator* and the BCC *estimator.*

Given any one of these estimators of $\mathcal{P}$, estimators of $D(\boldsymbol{x}, \boldsymbol{y})$ can be constructed. For example, using $\widehat{\mathcal{P}}_{BCC}$, we may write

$$\widehat{D}_{BCC}(\boldsymbol{x}, \boldsymbol{y}) \equiv \inf\{\theta > 0 | (\boldsymbol{x}, \theta^{-1}\boldsymbol{y}) \in \widehat{\mathcal{P}}_{BCC}\}, \tag{3.2}$$

where $\widehat{\mathcal{P}}_{BCC}$ has replaced $\mathcal{P}$ on the right-hand side of (2.6). Alternative estimators of

$D(\boldsymbol{x}, \boldsymbol{y})$ can be obtained by defining

$$\widehat{D}_{CCR}(\boldsymbol{x}, \boldsymbol{y}) \equiv \inf\{\theta > 0 | (\boldsymbol{x}, \theta^{-1}\boldsymbol{y}) \in \widehat{\mathcal{P}}_{CCR}\} \tag{3.3}$$

and

$$\widehat{D}_{DST}(\boldsymbol{x}, \boldsymbol{y}) \equiv \inf\{\theta > 0 | (\boldsymbol{x}, \theta^{-1}\boldsymbol{y}) \in \widehat{\mathcal{P}}_{DST}\}. \tag{3.4}$$

It is similarly straightforward (with the benefit of hindsight and the pioneering work of Charnes *et al.,* 1978) to compute estimates of $D(\boldsymbol{x}, \boldsymbol{y})$ using (3.2)–(3.4) and the data in $\mathcal{S}_n$. In particular, (3.2)–(3.3) may be rewritten as linear programs:

$$\left[\widehat{D}_{BCC}(\boldsymbol{x}, \boldsymbol{y})\right]^{-1} = \max\left\{\theta \mid \theta\boldsymbol{y} \le \boldsymbol{Y\tau},\ \boldsymbol{x} \ge \boldsymbol{X\tau},\ \boldsymbol{i\tau} = 1,\ \boldsymbol{\tau} \in \mathbb{R}^n_+\right\}, \tag{3.5}$$

and

$$\left[\widehat{D}_{CCR}(\boldsymbol{x}, \boldsymbol{y})\right]^{-1} = \max\left\{\theta \mid \theta\boldsymbol{y} \le \boldsymbol{Y\tau},\ \boldsymbol{x} \ge \boldsymbol{X\tau},\ \boldsymbol{\tau} \in \mathbb{R}^n_+\right\}, \tag{3.6}$$

where $\boldsymbol{Y} = [\boldsymbol{y}_1\ \ldots\ \boldsymbol{y}_n]$, $\boldsymbol{X} = [\boldsymbol{x}_1\ \ldots\ \boldsymbol{x}_n]$, with each $\boldsymbol{x}_i,\ \boldsymbol{y}_i\ i = 1, \ldots, n$ denoting the $(p \times 1)$ and $(q \times 1)$ vectors of observed inputs and outputs (respectively), $\boldsymbol{i}$ is a $(1 \times n)$ vector of ones, and $\boldsymbol{\tau} = [\tau_1\ \ldots\ \tau_n]'$ is a $(n \times 1)$ vector of intensity variables. One can also rewrite (3.4) as a linear program, *e.g.,*

$$\left[\widehat{D}_{DST}(\boldsymbol{x}, \boldsymbol{y})\right]^{-1} = \max\left\{\theta \mid \theta\boldsymbol{y} \le \boldsymbol{Y\tau},\ \boldsymbol{x} \ge \boldsymbol{X\tau},\ \boldsymbol{\tau} \in \{0, 1\}\right\}, \tag{3.7}$$

although linear programming is typically not used to compute estimates for the estimator based on the free disposal hull.

Given our assumption that $\mathcal{P}$ is convex, by construction,

$$\widehat{\mathcal{P}}_{DST} \subseteq \widehat{\mathcal{P}}_{BCC} \subseteq \begin{cases} \widehat{\mathcal{P}}_{CCR} \\ \mathcal{P}. \end{cases} \tag{3.8}$$

If the technology $\mathcal{P}^\partial$ exhibits constant returns to scale everywhere, then it is also the case that $\widehat{\mathcal{P}}_{CCR} \subseteq \mathcal{P}$, but otherwise, $\widehat{\mathcal{P}}_{CCR} \not\subseteq \mathcal{P}$.

The differences between $\mathcal{P}$ and any of the estimators $\widehat{\mathcal{P}}_{BCC}$, $\widehat{\mathcal{P}}_{CCR}$, and $\widehat{\mathcal{P}}_{DST}$ are of utmost importance, for these differences determine the difference between $D(\boldsymbol{x}, \boldsymbol{y})$ and any of the corresponding estimators $\widehat{D}_{BCC}(\boldsymbol{x}, \boldsymbol{y})$, $\widehat{D}_{CCR}(\boldsymbol{x}, \boldsymbol{y})$, and $\widehat{D}_{DST}(\boldsymbol{x}, \boldsymbol{y})$. Note that $\mathcal{P}$ and especially $D(\boldsymbol{x}, \boldsymbol{y})$ are the *true* quantities of interest. By contrast, $\widehat{\mathcal{P}}_{BCC}$, $\widehat{\mathcal{P}}_{CCR}$, and $\widehat{\mathcal{P}}_{DST}$ are *estimators* of $\mathcal{P}$, while $\widehat{D}_{BCC}(\boldsymbol{x}, \boldsymbol{y})$, $\widehat{D}_{CCR}(\boldsymbol{x}, \boldsymbol{y})$, and $\widehat{D}_{DST}(\boldsymbol{x}, \boldsymbol{y})$ are each *estimators* of $D(\boldsymbol{x}, \boldsymbol{y})$. The true things that we are interested in, namely $\mathcal{P}$ and $D(\boldsymbol{x}, \boldsymbol{y})$, are fixed, but unknown. Estimators, on the other hand, are necessarily *random variables.* When the data in $\mathcal{S}_n$ are used together with (3.5)–(3.7) to compute numerical values for the estimators $\widehat{D}_{BCC}(\boldsymbol{x}, \boldsymbol{y})$, $\widehat{D}_{CCR}(\boldsymbol{x}, \boldsymbol{y})$, and $\widehat{D}_{DST}(\boldsymbol{x}, \boldsymbol{y})$, the resulting real numbers are merely specific realizations of different random variables; these numbers (which are typically those reported in efficiency studies) are *estimates.*

In our view, anything we might compute from data is an estimate. The formula used for such a computation defines an estimator. The DEA/FDH setting is not an exception to this principle.

The skeptical reader may well ask, "what if I have observations on *all* the firms in the industry I am examining?" This is a perfectly reasonable question, but it begs for additional questions. What happens if an entrepreneuer establishes a new firm in this industry? How well might this new firm perform? Is there any reason why, at least in principle, it cannot perform better than the original firms in the industry? In other words, is it reasonable to assume that the new firm cannot operate somewhere above the convex hull of the existing firms represented by input/output vectors in $\mathbb{R}_+^{p+q}$?

We can see no reason why the general answer to this last question must be anything other than a resounding "no!" If one accepts this answer, it seems natural to then think in terms of infinite populations, and to view one's sample as having resulted from a draw from an infinite population. Anyone who does not accept our answer to the question posed at the end of the previous paragraph, or who wishes to think in terms of finite populations, must be thinking of a highly specialized case with which we are unfamiliar.

4. A STATISTICAL MODEL

Continuing with the reasoning at the end of the previous section, once one accepts that the world is an uncertain place and all we can hope for are reliable estimators of unobservables such as $D(\boldsymbol{x}, \boldsymbol{y})$, some additional questions are raised:

- What are the properties of our estimators? Are they at least consistent?
- How much can we learn about the true values of interest? Can we make inferences?
- Can we test hypotheses regarding $\mathcal{P}^{\partial}$, even though it is unobservable?
- Will our inferences and hypothesis test be reliable, meaningful?

Before these questions can be answered, a statistical model must be defined. A statistical model is merely a set of assumptions on the data-generating process (DGP). While many assumptions are possible, prudence dictates that the assumptions be chosen to provide enough structure so that desirable properties of estimators can be derived, but without imposing unneccesary restrictions. In addition to Assumptions A1–A3 adopted in Section 2, we now adopt assumptions based on those of Kneip *et al.* (1998).

Assumption A4: *the sample observations in $\mathcal{S}_n$ are realizations of identically, independently distributed (iid) random variables with probability density function $f(\boldsymbol{x}, \boldsymbol{y})$ with support over $\mathcal{P}$.*

Note that a point $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{p+q}$ represented by Cartesian coordinates can also be represented by cylindrical coordinates $(\boldsymbol{x}, \omega, \boldsymbol{\eta})$ where $(\omega, \boldsymbol{\eta})$ are the polar coordinates of $\boldsymbol{y}$ in $\mathbb{R}_+^q$. Thus, the modulus $\omega = \omega(\boldsymbol{y}) \in \mathbb{R}_+^1$ of $\boldsymbol{y}$ is given by the square root of the sum of the squared elements of $\boldsymbol{y}$, and the jth element of the corresponding angle $\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{y}) \in \left[0, \frac{\pi}{2}\right]^{q-1}$ of $\boldsymbol{y}$ is given by $\arctan(y^{j+1}/y^1)$ for $y^1 \neq 0$ (where y^j represents the jth element of $\boldsymbol{y}$); if $y^1 = 0$, then all elements of $\boldsymbol{\eta}(\boldsymbol{y})$ equal zero.

Writing $f(\boldsymbol{x}, \boldsymbol{y})$ in terms of the cylindrical coordinates, we can decompose the density by writing

$$f(\boldsymbol{x}, \omega, \boldsymbol{\eta}) = f(\omega \mid \boldsymbol{x}, \boldsymbol{\eta}) f(\boldsymbol{\eta} \mid \boldsymbol{x}) f(\boldsymbol{x}) \tag{4.1}$$

wehre all the conditional densities exist. In particular, $f(\boldsymbol{x})$ is defined on $\mathbb{R}_+^p$, $f(\boldsymbol{\eta} \mid \boldsymbol{x})$ is defined on $\left[0, \frac{\pi}{2}\right]^{q-1}$, and $f(\omega \mid \boldsymbol{x}, \boldsymbol{\eta})$ is defined on $\mathbb{R}_+^1$.

Now consider a point $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$, and the corresponding point $(\boldsymbol{x}, \boldsymbol{y}^{\partial}(\boldsymbol{x}))$ which is the projection of $(\boldsymbol{x}, \boldsymbol{y})$ onto $\mathcal{P}^{\partial}$ in the direction orthogonal to $\boldsymbol{x}$. Then the moduli of these points are related to the output distance function via

$$0 \leq D(\boldsymbol{x}, \boldsymbol{y}) = \frac{\omega(\boldsymbol{y})}{\omega(\boldsymbol{y}^{\partial}(\boldsymbol{x}))} \leq 1. \tag{4.2}$$

Then the density $f(\omega \mid \boldsymbol{x}, \boldsymbol{\eta})$ on $[0, \omega(\boldsymbol{y}^{\partial}(\boldsymbol{x}))]$ implies a density $f(D(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x}, \boldsymbol{\eta})$ on the interval [0.1].

In order for our estimators of $\mathcal{P}$ and $D(\boldsymbol{x}, \boldsymbol{y})$ to be consistent, the probability of observing firms on $\mathcal{P}^{\partial}$ must approach unity as the sample size increases:

Assumption A5: *for all $\boldsymbol{x} \geq \boldsymbol{0}$ and all $\boldsymbol{\eta} \in \left[0, \frac{\pi}{2}\right]^{q-1}$, there exist constants $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that for all $\omega \in \left[\omega(\boldsymbol{y}^{\partial}(\boldsymbol{x})),\ \omega(\boldsymbol{y}^{\partial}(\boldsymbol{x})) + \epsilon_2\right]$, $f(\omega \mid \boldsymbol{x}, \boldsymbol{\eta}) \geq \epsilon_1$.*

In addition, an assumption about the smoothness of the frontier is needed:[2]

Assumption A6: *for all $(\boldsymbol{x}, \boldsymbol{y})$ in the interior of $\mathcal{P}$, $D(\boldsymbol{x}, \boldsymbol{y})$ is differentiable in both its arguments.*

Assumptions A1–A6 define the DGP $\mathcal{F}$ which yields the data in $\mathcal{S}_n$. These assumptions are somewhat more detailed than the set of assumptions listed in Section 2, which were motivated by microeconomic theory.

5. SOME ASYMPTOTIC RESULTS

Recent papers have provided some asymptotic results for DEA/FDH estimators. Korostelev *et al.* (1995a, 1995b) examined the convergence of estimators of $\mathcal{P}$, and proved (for the special case where $p \geq 1$, $q = 1$) that both $\widehat{\mathcal{P}}_{DST}$ and $\widehat{\mathcal{P}}_{BCC}$ are consistent estimators of $\mathcal{P}$ in the sense that

$$d\left(\widehat{\mathcal{P}}_{DST}, \mathcal{P}\right) = O_p\left(n^{-\frac{1}{p+1}}\right) \tag{5.1}$$

and

$$d\left(\widehat{\mathcal{P}}_{BCC}, \mathcal{P}\right) = O_p\left(n^{-\frac{2}{p+2}}\right) \tag{5.2}$$

[2]Our characterization of the smoothness condition here is stronger than required; Kneip *et al.* (1998) require only Lipschitz continuity for $D(\boldsymbol{x}, \boldsymbol{y})$, which is implied by the simpler, but stronger requirement presented here.

where $d(\widehat{\mathcal{P}}, \mathcal{P})$ is the Lebesgue measure of the difference between two sets.[3]

The result for $\widehat{\mathcal{P}}_{DST}$ in (5.1) still holds if assumption A1 given above is dropped. However, the convexity assumption is necessary for (5.2). It would be seemingly straightforward to extend these results to obtain a convergence rate for $\widehat{\mathcal{P}}_{CCR}$ provided one adopted an additional assumption that $\mathcal{P}^{\partial}$ displays constant returns to scale everywhere.

These results also indicate that the *curse of dimensionality* which typically plagues nonparametric estimators is at work in the DEA/FDH setting; the convergence rates decrease as the number of inputs is increased. The practical implicaton of this is that researchers will need increasing amounts of data to get meaningful results as the number of inputs is increased. And, although these results were obtained with $q = 1$, presumably allowing the number of outputs to increase would have a similar effect on convergence rates and the resulting data requirements.

This intuition is confirmed by Park *et al.* (1997), and Park *et al.* (1998), who derive convergence rates for $\widehat{D}_{DST}(\boldsymbol{x}, \boldsymbol{y})$ and $\widehat{D}_{BCC}(\boldsymbol{x}, \boldsymbol{y})$ for the general case where $p \geq 1$, $q \geq 1$:

$$\widehat{D}_{DST}(\boldsymbol{x}, \boldsymbol{y}) - D(\boldsymbol{x}, \boldsymbol{y}) = O_p\left(n^{-\frac{1}{p+q}}\right) \tag{5.3}$$

and

$$\widehat{D}_{BCC}(\boldsymbol{x}, \boldsymbol{y}) - D(\boldsymbol{x}, \boldsymbol{y}) = O_p\left(n^{-\frac{2}{p+q+1}}\right). \tag{5.4}$$

In both cases, the convergence rates are affected by $p + q$. The convergence rate for the DST estimator is slower than for the BCC estimator, and thus the BCC estimator is preferred when $\mathcal{P}$ is convex. But if Assumption A1 does not hold, then there is no choice but to use the DST estimator—the BCC estimator as well as the CCR estimator would be inconsistent in this case since $\widehat{\mathcal{P}}_{BCC}$ and $\widehat{\mathcal{P}}_{CCR}$ are convex for all $1 \leq n \leq \infty$, and so cannot converge to a nonconvex set as $n \to \infty$.

For the special case where $p = q = 1$, Gijbels *et al.* (1999) derived results which indicate that

$$\widehat{D}_{BCC}(\boldsymbol{x}, \boldsymbol{y}) - D(\boldsymbol{x}, \boldsymbol{y}) \overset{\text{asy.}}{\sim} G(\cdot, \cdot) \tag{5.5}$$

[3]Banker (1993) showed that $\widehat{\mathcal{P}}_{BCC}$ is a consistent estimator of $\mathcal{P}$, but did not provide convergence rates.

where G is a known distribution function depending on unknown quantitites related to the DGP $\mathcal{F}$.[4]

For the general case where $p \geq 1$ and $q \geq 1$, Park *et al.* (1997) prove that

$$\widehat{D}_{DST}(\boldsymbol{x}, \boldsymbol{y}) - D(\boldsymbol{x}, \boldsymbol{y}) \overset{\text{asy.}}{\sim} \text{Weibull}(\cdot, \cdot), \tag{5.6}$$

where the unknonwn parameters of the Weibull distribution again must be estimated and depend on characteristics of the DGP.

These results are potentially very useful for applied researchers. In particular, (5.6) may be used to construct and then estimate confidence intervals for $D(\boldsymbol{x}, \boldsymbol{y})$ when the FDH estimator is used. It is unfortunate that a similar result for the DEA case to date exists only for the special case where $p = q = 1$. In any case, these are asymptotic results, and the quality of confindence intervals, *etc.* estimated using these results in small samples remains an open question.

6. BOOTSTRAPPING IN DEA/FDH MODELS

The bootstrap (Efron, 1979, 1982) offers an alternative approach to inference and hypothesis-testing in DEA/FDH models. In the case of DEA models with multiple inputs or outputs, the bootstrap currently offers the *only* approach to inference and hypothesis-testing.

Bootstrapping is based on the analogy principle (*e.g.,* see Manksi, 1988). The presentation here is based on our earlier work in Simar and Wilson (1998a, 1999c). In the *true world* we observe the data in $\mathcal{S}_n$ which are generated from

$$\mathcal{F} = \mathcal{F}(\mathcal{P}, f(\boldsymbol{x}, \boldsymbol{y})). \tag{6.1}$$

In the true world, $\mathcal{F}$, $\mathcal{P}$, and $D(\boldsymbol{x}, \boldsymbol{y})$ are unobserved, and must be estimated using $\mathcal{S}_n$.

Let $\widehat{\mathcal{F}}(\mathcal{S}_n)$ be a consistent estimator of $\mathcal{F}$:

$$\widehat{\mathcal{F}}(\mathcal{S}_n) = \mathcal{F}(\widehat{\mathcal{P}}, \widehat{f}(\boldsymbol{x}, \boldsymbol{y})). \tag{6.2}$$

[4] In particular, the unknown quantities are determined by the curvature of $\mathcal{P}^{\partial}$ and the value of $f(\boldsymbol{x}, \boldsymbol{y})$ at the point where $(\boldsymbol{x}, \boldsymbol{y})$ is projected onto $\mathcal{P}^{\partial}$ in the direction orthogonal to $\boldsymbol{x}$. See Gijbels *et al.* (1999) for additional details.

One can easily simulate what happens in the true world by drawing a new dataset $\mathcal{S}_n^*$ from $\widehat{\mathcal{F}}(\mathcal{S}_n)$, and then applying the original estimator to these new data. If the original estimator for a point $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ (not necessarily contained in $\mathcal{S}_n$) was $\widehat{D}_{BCC}^*(\boldsymbol{x}_0, \boldsymbol{y}_0)$, for example, then this could be accomplished by solving

$$\left[\widehat{D}_{BCC}^*(\boldsymbol{x}_0, \boldsymbol{y}_0)\right]^{-1} = \max\left\{\theta \mid \theta\boldsymbol{y}_0 \le \boldsymbol{Y}^*\boldsymbol{\tau},\ \boldsymbol{x}_0 \ge \boldsymbol{X}^*\boldsymbol{\tau},\ \boldsymbol{i\tau} = 1,\ \boldsymbol{\tau} \in \mathbb{R}_+^n\right\} \tag{6.3}$$

where $\boldsymbol{Y}^* = [\boldsymbol{y}_1^* \ \ldots \ \boldsymbol{y}_n^*]$, $\boldsymbol{X}^* = [\boldsymbol{x}_1^* \ \ldots \ \boldsymbol{x}_n^*]$, with each $(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)$, $i = 1, \ldots, n$ denoting observations in the pseudo dataset $\mathcal{S}_n^*$. Repeating this process B times (where B is appropriately large) will result in a set of bootstrap values $\left\{\widehat{D}_{BCC,b}^*(\boldsymbol{x}_0, \boldsymbol{y}_0)\right\}_{b=1}^B$.

When the bootstrap is consistent, then

$$\begin{aligned}\left(\widehat{D}^*(\boldsymbol{x}_0, \boldsymbol{y}_0) - \widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)\right) &|\mathcal{F}(\widehat{\mathcal{P}}, \widehat{f}(\boldsymbol{x}, \boldsymbol{y})) \\ &\overset{\text{approx}}{\sim} \left(\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0) - D(\boldsymbol{x}_0, \boldsymbol{y}_0)\right) \mid \mathcal{F}(\mathcal{P}, f(\boldsymbol{x}, \boldsymbol{y})).\end{aligned} \tag{6.4}$$

Given the original estimate $\widehat{D}_{BCC}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ and the set of bootstrap values

$$\left\{\widehat{D}_{BCC,b}^*(\boldsymbol{x}_0, \boldsymbol{y}_0)\right\}_{b=1}^B,$$

the left-hand side of (6.4) is known with arbitrary precision (determined by the choice of number of bootstarp replications, B); the approximation improves as the sample size, n, increases.

Note that we could tell the story of the previous paragraph in terms of either the CCR or the DST estimator by merely changing the subscript on the notation for the distance function estimator. Since this is the case for everything that follows as well, we will drop the CCR, BCC, and DST subscripts from our notation for the distance function estimator in the remainder of the paper.

Equation (6.4) is the essence of the bootstrap. In principle, since $\mathcal{F}(\widehat{\mathcal{P}}, \widehat{f}(\boldsymbol{x}, \boldsymbol{y}))$ is known, it should be possible to determine the distribution on the left-hand side analytically. In practice, however, this is intractible in most problems, including the present one. Hence Monte Carlo simulations are used to approximate this distribution. In our presentation, after the B bootstrap replications are performed, the set of bootstrap values $\left\{\widehat{D}_b^*(\boldsymbol{x}_0, \boldsymbol{y}_0)\right\}_{b=1}^{B}$ gives an empirical approximation to this distribution.

Once the set of bootstrap values $\left\{\widehat{D}_b^*(\boldsymbol{x}_0, \boldsymbol{y}_0)\right\}_{b=1}^{B}$ has been obtained, it is straightforward to estimate confidence intervals for the true distance function value $D(\boldsymbol{x}_0, \boldsymbol{y}_0)$. This is accomplished by noting that if we knew the true distribution of $\left(\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0) - D(\boldsymbol{x}_0, \boldsymbol{y}_0)\right)$, then it would be trivial to find values a_α and b_α such that

$$\Pr\left(-b_\alpha \leq \widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0) - D(\boldsymbol{x}_0, \boldsymbol{y}_0) \leq -a_\alpha\right) = 1 - \alpha. \tag{6.5}$$

Of course, a_α and b_α are unknown, but from the empirical bootstrap distribution of the pseudo estimates $\widehat{D}_b^*(\boldsymbol{x}_0, \boldsymbol{y}_0)$, $b = 1, \ldots, B$, we can find values $\widehat{a}_\alpha$ and $\widehat{b}_\alpha$ such that

$$\Pr\left(-\widehat{b}_\alpha \leq \widehat{D}^*(\boldsymbol{x}_0, \boldsymbol{y}_0) - \widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0) \leq -\widehat{a}_\alpha \mid \widehat{\mathcal{F}}(\mathcal{S}_n)\right) = 1 - \alpha. \tag{6.6}$$

Finding $\widehat{a}_\alpha$ and $\widehat{b}_\alpha$ involves sorting the values $\left(\widehat{D}_b^*(\boldsymbol{x}_0, \boldsymbol{y}_0) - \widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)\right)$, $b = 1, \ldots, B$ 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 (6.5) is then

$$\Pr\left(-\widehat{b}_\alpha \leq \widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0) - D(\boldsymbol{x}_0, \boldsymbol{y}_0) \leq -\widehat{a}_\alpha\right) \approx 1 - \alpha. \tag{6.7}$$

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

$$\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0) + \widehat{a}_\alpha \;\leq\; D(\boldsymbol{x}_0, \boldsymbol{y}_0) \;\leq\; \widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0) + \widehat{b}_\alpha. \tag{6.8}$$

This procedure can be used for any $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathbb{R}_+^{p+q}$ for which $\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ 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 $(\boldsymbol{x}_0, \boldsymbol{y}_0) = (\boldsymbol{x}_i, \boldsymbol{y}_i)$, $i = 1, \ldots, n$, producing a set of n confidence intervals of the form (6.8), one for each firm.

The method described here for estimating confidence intervals from the set of bootstrap values $\left\{\widehat{D}_b^*(\boldsymbol{x}_0,\boldsymbol{y}_0)\right\}_{b=1}^{B}$ differs slightly from what we proposed in Simar and Wilson (1998a); here, we avoid explicit use of a bias estimate, which adds unnecessary noise to the estimated confidence intervals. The bias estimate itself, however remains interesting.

By definition,

$$\text{BIAS}\left(\widehat{D}(\boldsymbol{x}_0,\boldsymbol{y}_0)\right) = E\left(\widehat{D}(\boldsymbol{x}_0,\boldsymbol{y}_0)\right) - D(\boldsymbol{x}_0,\boldsymbol{y}_0). \tag{6.9}$$

The bootstrap bias estimate for the original estimator $\widehat{D}(\boldsymbol{x}_0,\boldsymbol{y}_0)$ is the empirical analog of (6.9):

$$\widehat{\text{BIAS}}_B\left(\widehat{D}(\boldsymbol{x}_0,\boldsymbol{y}_0)\right) = B^{-1}\sum_{b=1}^{B}\widehat{D}_b^*(\boldsymbol{x}_0,\boldsymbol{y}_0) - \widehat{D}(\boldsymbol{x}_0,\boldsymbol{y}_0). \tag{6.10}$$

It is tempting to construct a bias-corrected estimator of $D(\boldsymbol{x}_0,\boldsymbol{y}_0)$ by computing

$$\begin{aligned}\widehat{\widehat{D}}(\boldsymbol{x}_0,\boldsymbol{y}_0) &= \widehat{D}(\boldsymbol{x}_0,\boldsymbol{y}_0) - \widehat{\text{BIAS}}_B\left(\widehat{D}(\boldsymbol{x}_0,\boldsymbol{y}_0)\right)\\ &= 2\widehat{D}(\boldsymbol{x}_0,\boldsymbol{y}_0) - B^{-1}\sum_{b=1}^{B}\widehat{D}_b^*(\boldsymbol{x}_0,\boldsymbol{y}_0).\end{aligned} \tag{6.11}$$

It is well known (see e.g. Efron and Tibshirani, 1993), however, that this bias correction introduces additional noise; the mean-square error of $\widehat{\widehat{D}}(\boldsymbol{x}_0,\boldsymbol{y}_0)$ may be greater than the mean-square error of $\widehat{D}(\boldsymbol{x}_0,\boldsymbol{y}_0)$. The variance of the summation term on the right-hand side of (6.10) can be made arbitrarily small by increasing B. Yet, even if $B\to\infty$, the bias-corrected estimator $\widehat{\widehat{D}}(\boldsymbol{x}_0,\boldsymbol{y}_0)$ will have variance equal to four times that of the original, uncorrected estimator, $\widehat{D}(\boldsymbol{x}_0,\boldsymbol{y}_0)$, (again illustrating the fact that the bootstrap is an asymptotic procedure). The sample variance of the bootstrap values $\widehat{D}_b^*(\boldsymbol{x}_0,\boldsymbol{y}_0)$ gives an estimate $\widehat{\sigma}^2$ of the variance of $\widehat{D}(\boldsymbol{x}_0,\boldsymbol{y}_0)$:

$$\widehat{\sigma}^2 = B^{-1}\sum_{b=1}^{B}\left[\widehat{D}_b^*(\boldsymbol{x}_0,\boldsymbol{y}_0) - B^{-1}\sum_{b=1}^{B}\widehat{D}_b^*(\boldsymbol{x}_0,\boldsymbol{y}_0)\right]^2. \tag{6.12}$$

Hence the bias-correction should not be used unless

$$\widehat{\sigma}^2 < \frac{1}{3}\left[\widehat{\text{BIAS}}_B\left(\widehat{D}(\boldsymbol{x}_0,\boldsymbol{y}_0)\right)\right]^2. \tag{6.13}$$

7. IMPLEMENTING THE BOOTSTRAP

We have shown elsewhere (Simar and Wilson 1998a, 1998c, 1999a, 1999b, 2000) that the generation of the bootstrap pseudo data $\mathcal{S}_n^*$ is crucial in determining whether the bootstrap gives consistent estimates of confidence intervals, bias, *etc.* In the classical linear regression model, one may resample from the estimated residuals, or alternatively resample the original sample observations to construct the pseudo dataset $\mathcal{S}_n^*$; in either case, the bootstrap will give consistent estimates. Both these approaches are variants of what has been called the *naive* bootstrap. The analog of these methods in frontier estimation would be to resample from the original distance function estimates, or alternatively from the $(\boldsymbol{x}, \boldsymbol{y})$ pairs in $\mathcal{S}_n$.

In the present setting, however, there is a crucial difference from the classical linear regression model: here, the DGP $\mathcal{F}$ has bounded support over $\mathcal{P}$, while in the classical linear regression model, the DGP has unbounded support. A related problem is that under our assumptions, the conditional density $f(D(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x}, \boldsymbol{\eta})$ has bounded support over the interval $(0, 1]$, and is right-discontinuous at 1. It is widely known that problems such as these can cause the naive bootstrap to give *inconsistent* estimates (*e.g.,* see Bickel and Friedman, 1981; Swanepoel, 1986; Beran and Ducharme, 1991; and Efron and Tibshirani, 1993), and this is true in the present setting as we have shown in Simar and Wilson (1999a, 1999b, 2000).[5]

To address the the boundary problems which doom the naive bootstrap in the present situation, we draw pseudo datasets from a smooth, consistent, nonparametric estimate of the DGP $\mathcal{F}$, as represented by $f(\boldsymbol{x}, \omega, \boldsymbol{\eta})$ in (4.1). In Simar and Wilson (1998a), we drew values D^* from a kernel estimate $\widehat{f}(D)$ of the marginal density of the original estimates $\widehat{D}(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $i = 1, \ldots, n$; given (4.2), this is tantamount to assuming $f(\omega \mid \boldsymbol{x}, \boldsymbol{\eta}) = f(\omega)$ in (4.1), or in other words, that the distribution of inefficiencies is homogeneous and does

[5] Explicit descriptions of why either variation of the naive bootstrap results in inconsistent estimates are given in Simar and Wilson (1999a, 2000). Löthgren and Tambour (1996, 1997) and Löthgren (1997, 1998) employ a bizarre, illogical variant of the naive bootstrap different from the more typical variations we have mentioned. This approach also leads to an inconsistency problem, as discussed and confirmed with Monte Carlo experiments in Simar and Wilson (2000).

not depend upon location within the produciton set $\mathcal{P}$. This assumption can be relaxed, as in Simar and Wilson (1998c), but at a cost of increased complexity and computational burden. In this paper, we will focus on the homogeneous case.

Kernel density estimation is rather easy to apply and has been widely studied. Given values z_i, $i = 1, \ldots, n$ on the real number line, the kernel estimate of the density $g(z)$ is given by

$$\widehat{g}(z) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{z - z_i}{h}\right), \tag{7.1}$$

where $K(\cdot)$ is a *kernel function* and h is a *bandwidth parameter.* Both $K(\cdot)$ and h must be chosen, but there are well-established guidelines to aid with these choices. In particular, the kernel function $K(\cdot)$ must be piecewise continuous and must satisfy $\int_{-\infty}^{\infty} K(u)\, du = 1$ and $\int_{-\infty}^{\infty} uK(u)\, du = 0$. Thus any probability density function that is symmetric around zero is a valid kernel function; in addition, one could use even-ordered polynomials bounded on some interval with coefficients chosen so that the polynomial integrates to unity over this interval.[6] As Silverman (1986) shows, the choice of the kernel function is far less critical for obtaining a good estimate (in the sense of low mean integrated square error) of $f(z)$ than the choice of the bandwidth parameter h.

In order for the kernel density estimator to be consistent, the bandwidth must be chosen so that $h = O(n^{-1/5})$ for the univariate case considered here. If the data are approximately normally distributed, then one may employ the normal reference rule and set $h = 1.06\widehat{\sigma}n^{-1/5}$, where $\widehat{\sigma}$ is the sample standard deviation of the data whose density is being estimated (Silverman, 1986). In cases where the data are clearly nonnormal, as will be the case when estimating the density $f(D)$, one can plot the density estimate for various values of h and choose the value of h that gives a reasonable estimate, as in Silverman (1978) and Simar and Wilson (1998a). This approach, however, contains an element of subjectivity; a better approach is to employ least squares cross-validation, which involves choosing the value of h that minimizes an approximation to mean integrated square error;

[6]Appropriately chosen high-order kernels can reduce the order of the bias in the kernel density estimator, but run the risk of producing negative density estimates at some locations.

see Silverman (1986) for details. In many cases involving high dimensions (*i.e.,* large $p+q$), many of the distance function estimates will equal unity (this is especially the case when the BCC or FDH estimators are used), and this creates a discretization problem in the cross-validation procedure. Simar and Wilson (1998b) propose a weighted cross-validation procedure to avoid this problem in the univariate setting; Simar and Wilson (1998c) extend this idea to the multivariate setting.

Regardless of how the bandwidth parameter is chosen, it is important to note that ordinary kernel estimators (such as the one in (7.1) above) of densities with bounded support will be biased near the boundaries of the support. Necessarily, for $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$, we have $0 < D(\boldsymbol{x}, \boldsymbol{y}) \leq \widehat{D}(\boldsymbol{x}, \boldsymbol{y}) \leq 1$, and so this will be a problem in estimating $f(D)$. It is easy to see why this problem arises: in (7.1), when $K(\cdot)$ is symmetric about zero, the density estimate is determined at a particular point on the real number line by adding up the value of functions $K(\cdot)$ centered over the observed data along the real number line *on either side* of the point where the density estimate is being evaluated. If, for example, $K(\cdot)$ is chosen as a standard normal probability density function, then nearby observations contribute relatively more to the density estimate at this particular point than do observations that are farther away along the real number line.[7] When (7.1) is used to evaluate the density estimate at unity in our application, then if $K(\cdot)$ is symmetric, there will necessarily be no data on the right side of the boundary to contribute to the smoothing, and this causes the bias problem. An obvious approach would be to let the kernel function become increasingly asymmetric as the boundary is approached, but this is problematic for several reasons as discussed by Scott (1992). A much simpler solution is to use the reflection method as in Simar and Wilson (1998a).

The reflection method involves reflecting each of the n original estimates $\widehat{D}(\boldsymbol{x}_i, \boldsymbol{y}_i)$ about the boundary at unity (by computing $1 - \widehat{D}(\boldsymbol{x}_i, \boldsymbol{y}_i)$ for each $\widehat{D}(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $i = 1, \ldots, n$) to

[7] This is the sense in which the kernel density estimator is a *smoother,* since it is, in effect, smoothing the empirical density function which places probability mass $1/n$ at each observed datum. Setting $h = 0$ in (7.1) yields the empirical density function, while letting $h \to \infty$ yields a flat density estimate. The requirement that $h = O(n^{-1/5})$ to ensure consistency of the kernel density estimate results from the fact that as n increases, h must become smaller, but not too quickly.

obtain $2n$ points along the real number line. The output distance function estimates are also bounded on the left at 0, but typically there will be no values near zero, suggesting that the density is near zero at this boundary. Hence we ignore the effect of the left boundary, and concentrate on the boundary at unity. Viewing the reflected data (with $2n$ observations) as unbounded, we can estimate the density of these data using the estimator in (7.1) with no special problems. Then, this unbounded density estimate can be truncated on the right at unity to obtain an estimate of the density of $\widehat{D}$ with support on the interval $(0, 1]$.

Once the kernel function $K(\cdot)$ and the bandwidth h have been chosen, it is not necessary to evaluate the kernel density estimate in (7.1) in order to draw random values. Rather, a computational shortcut is afforded by Silverman (1986) and for cases where the kernel function $K(\cdot)$ is a regular probability density function.

We need to draw n values D^* from the estimated density of the original distance function estimates. Let $\{\epsilon_i\}_{i=1}^n$ be a set of n iid draws from the probability density function used to define the kernel function; let $\{d_i\}_{i=1}^n$ be a set of values drawn independently, uniformly, and with replacement from the set of reflected distance function estimates $\mathcal{R} = \left\{\widehat{D}(\boldsymbol{x}_i, \boldsymbol{y}_i),\ 2 - \widehat{D}(\boldsymbol{x}_i, \boldsymbol{y}_i)\right\}$; and let $\overline{d} = n^{-1}\sum_{i=1}^n d_i$. Then compute

$$d_i^* = \overline{d} + \left(1 + h^2/s^2\right)^{1/2} \left(d_i + h\epsilon_i - \overline{d}\right), \tag{7.2}$$

where s^2 is the sample variance of the values $\nu_i = d_i + h\epsilon_i$. Using the convolution theorem, it can be shown that $\nu_i \sim \widehat{g}(\cdot)$, where $\widehat{g}(\cdot)$ is the kernel estimate of the density of the original distance function estimates and their reflections in $\mathcal{R}$. As is typical with kernel density estimates, the variance of the ν_i must be scaled upward as in (7.2). Straightforward manipulations reveal that

$$E(d_i^* \mid \mathcal{R}) = 1 \tag{7.3}$$

and

$$\mathrm{VAR}(d_i^* \mid \mathcal{R}) = s^2 \left(1 + \frac{h^2}{n(s^2 + h^2)}\right) \tag{7.4}$$

so that the variance of the d_i^* is asymptotically correct. All that remains is to reflect the d_i^* about unity by computing, for each $i = 1, \ldots, n$,

$$D_i^* = \begin{cases} d_i^* & \text{if } d_i^* \leq 1; \text{ or} \\ 2 - d_i^* & \text{otherwise.} \end{cases} \tag{7.5}$$

The last step is equivalent to "folding" the right half of the symmetric (about 1) estimate $\widehat{g}(\cdot)$ to the left around 1, and ensures that $d_i^* \leq 1$ for all i.

Putting all of this together, bootstrap estimates of confidence intervals, bias, *etc.* for the output distance function $D(\boldsymbol{x}_0, \boldsymbol{y}_0)$ evaluated for a particular, arbitrary point $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathbb{R}_+^{p+q}$ (such that the corresponding estimate $\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ exists) can be obtained via the following algorithm:

****

Algorithm #1:

[1] For each $(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathcal{S}_n$, apply one of the distance function estimators in (3.5)–(3.7) to obtain estimates $\widehat{D}(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $i = 1, \ldots, n$.

[2] If $(\boldsymbol{x}_0, \boldsymbol{y}_0) \notin \mathcal{S}_n$, repeat step [1] for $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ to obtain $\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)$.

[3] Reflect the n estimates $\widehat{D}(\boldsymbol{x}_i, \boldsymbol{y}_i)$ about unity, and determine the bandwidth parameter h via least-squares cross-validation.

[4] Use the computational shortcut in (7.2) to draw n bootstrap values D_i^*, $i = 1, \ldots, n$, from the kernel density estimate of $f(\widehat{D})$.

[5] Construct a pseudo dataset $\mathcal{S}_n^*$ with elements $(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)$ given by

$$\boldsymbol{y}_i^* = D_i^* \boldsymbol{y}_i / \widehat{D}(\boldsymbol{x}_i, \boldsymbol{y}_i)$$

and $\boldsymbol{x}_i^* = \boldsymbol{x}$.

[6] Use (6.3) (or an analog of (6.3) if the CCR or FDH estimators were used in step [1]) to compute the bootstrap estimate $\widehat{D}^*(\boldsymbol{x}_0, \boldsymbol{y}_0)$.

[7] Repeat steps [4]–[6] B times to obtain a set of B bootstrap estimates

$$\left\{ \widehat{D}_b^*(\boldsymbol{x}_0, \boldsymbol{y}_0) \right\}_{b=1}^{B}.$$

[8] Use (6.6) to determine $\widehat{a}_\alpha$ and $\widehat{b}_\alpha$, and then use these in (6.8) together with the original estimate $\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ obtained in step [2] (or step [1] if $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{S}_n$) to obtain an estimated confidence interval for $D(\boldsymbol{x}_0, \boldsymbol{y}_0)$. In addition, the bootstrap estimates can be used in (6.10) to obtain an estimate of the bias of $\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)$, and in (6.11) to obtain a bias-corrected estimator if the condition in (6.13) is satisfied.

****

Note that the arbitrary point $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ might or might not be in $\mathcal{P}$. Typically, however, researchers are interested in the efficiency of firms represented in the observed sample, $\mathcal{S}_n$. In that case, $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ will correspond in turn with each of the $(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathcal{S}_n$. Rather than repeatedly apply Algorithm #1 n times, considerable computational cost can be saved by noting that step [2] is not applicable in this case, and performing step [6] n times (for each element of $\mathcal{S}_n$) inside each of the bootstrap loops in step [7]. This will result in n distinct sets of bootstrap estimates when step [8] is reached; each can then be used to estimate a confidence interval for its corresponding true distance function.

It is rather straightforward to extend Algorithm #1 to estimation of confidence intervals for Malmquist indices (Simar and Wilson, 1999c) or to the problem of testing hypotheses regarding returns to scale as in Simar and Wilson (1998b). In choosing between the BCC and CCR estimators defined in (3.5)–(3.6), it is important to have some idea of whether the true technology $\mathcal{P}^\partial$ exhibits constant returns to scale everywhere; if it does not, the CCR estimator in (3.6) will necessarily be inconsistent. It would be trivial to extend the results in Simar and Wilson (1998b) to test for concavity of the production set $\mathcal{P}$ using the BCC estimator and the FDH estimator defined in (3.7). Tests of other hypotheses regarding the shape of the technology should also be possible.

8. MONTE CARLO EVIDENCE

We conducted a series of Monte Carlo experiments to examine the coverage probabilitites of estimated confidence intervals for output distance functions. We simulated two different DGPs for these Monte Carlo experiments, but to minimize computational costs, we set

$p = q = 1$ in both cases so that we examine only a single output produced from a single input. To consider a true technology with constant returns to scale everywhere, we used the model

$$y = xe^{-|v|} \tag{8.1}$$

where $v \sim N(0, 1)$ and $x \sim$ Uniform on $(1, 9)$. To allow for variable returns to scale, we used the model

$$y = (x - 2)^{2/3} e^{-|v|}; \tag{8.2}$$

in experiments where this model was used, $v \sim N(0, 1)$ as in the previous case, but $x \sim$ Uniform on $(3, 12)$. Pseudo random uniform deviates were generated from 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 $x_0 = 7.5$, $y_0 = 2$.

Each Monte Carlo experiment involved 1000 Monte Carlo trials; on each trial, we performed $B = 2000$ bootstrap replications. For each trial, we estimated confidence intervals for $D(x_0, y_0)$ and then checked whether the estimated confidence intervals included the true value $D(x_0, y_0)$. After 1000 trials, we divided the number of cases where $D(x_0, y_0)$ was included in the estimated confidence intervals by 1000 to produce the results shown in Table 1.

Table 1 contains 6 numbered columns; the first 3 correspond to the constant returns to scale model in (8.1) where the CCR estimator defined in (3.6) was employed, while the last 3 correspond to the variable returns to scale model in (8.2) where the BCC estimator defined in (3.5) was used. The columns labelled "Smooth" correspond to Algorithm #1 given earlier, using the kernel smoothing described in section 7. Results in the columns labelled "(x, y)" were obtained using the naive bootstrap where resampling is from the elements of $\mathcal{S}_n$. In terms of Algorithm #1, this amounts to deleting step [3] and replacing steps [4]–[5] with a single step where we draw n observations from $\mathcal{S}_n$ uniformly, independently, and with replacement to form the pseudo sample $\mathcal{S}_n^*$. Results in the columns labelled "$\widehat{D}$" were obtained with the other variant of the naive bootstrap mentioned previously, i.e., where the original distance function estimates were resampled. Again in terms of Algorithm #1, this involved deleteing step [3] and replacing step [4] with a step where we draw n times uniformly, independently, and with replacement, from the set $\left\{\widehat{D}(\boldsymbol{x}_i, \boldsymbol{y}_i)\right\}_{i=1}^{n}$ to obtain the bootstrap values $D^*(\boldsymbol{x}_i, \boldsymbol{y}_i)$.

Under constant returns to scale, the smooth bootstrap performs reasonably well with only 10 observations in the single input, single output case. The minor fluctuations in the reported coverages in column (1) that are seen as the sample size increases beyond 25 are to be expected due to sampling variation in the Monte Carlo experiment, and due to the fact that a finite number of bootstrap replications are being used. Also, the two variants of the naive bootstrap appear to perform, for the particular model in (8.1), reasonably well in terms of coverages, although not as well as the smooth bootstrap. One should not conclude from this, however, that it is safe to use the naive bootstrap—since it is inconsistent, there is no reason to think that coverages would be reasonable in any other setting than the one considered here. Moreover, the case of $p = q = 1$ with constant return to scale, although not typical of actual applications, is the most favorable case for the naive bootstrap since typically the estimated frontier will intersect only one sample observation. In higher dimensions, or where the BCC or DST estimators are used, a far higher portion of the original observations will have corresponding efficiency estimates equal to unity, which will be problematic for either version of the naive bootstrap.

Differences between the smooth bootstrap and the variants of the naive bootstrap become more pronounced with variable returns to scale, as columns (4)–(6) in Table 1 reveal. The smooth bootstrap does not perform quite as well as in the contant returns to scale case, but this is to be expected given the greater degree of curvature of the frontier in (8.2) (see Gijbels *et al.,* 1996, for details on the relation between the curvature of the frontier and the convergence rate of the output distance funciton estimator used here). Also, the coverages obtained with the smooth bootstrap appear to fluctuate somewhat as the sample size increases, but eventually seem to be on a path toward the nominal significance levels.

This merely reflects that fact that consistency does not imply that convergence must be monotonic. With the two variants of the naive bootstrap, however, even with 6400 observations, coverages at all significance levels are worse than with the smooth bootstrap at any of the sample sizes we examined. Again, the problem is that with variable returns to scale, many sample observations will typically have efficiency estimates equal to unity. Between the two variants of the naive bootstrap, the one based on resampling from the original distance function estimates seems to give poorer coverage than the one based on resampling from $\mathcal{S}_n$; but this distinction is irrelevant, since both methods yield inconsistent estimates.

It is important to note that the performance of the bootstrap will vary not only as sample size increases, but also depending on features of the true model, particularly the curvature of the frontier and the shape of the density $f(\boldsymbol{x}, \boldsymbol{y})$, The data generating process represented by (8.2) is somewhat unfavorable to the smooth bootstrap, in that the range of the inputs is considerably larger than the range of the outputs. To illustrate this point, we performed additional Monte Carlo experiments for $n = 200$, generating data for each trial using (8.2), but varying the range of the input variable. Results from these experiments are shown in Table 2, where we report estimated confidence interval coverages as in Table 1. We considered four cases, where x is alternately distributed uniformly on the (6.5,8.5), (5.5,9.5), (4.5,10.5), and (3.5,11.5) intervals (other features of the experiments remained unchanged). The results show that as the distribution of the input becomes more disperse, the coverages of confidence intervals estimated with each method worsen. As before, however, the smooth bootstrap dominates both variants of the naive bootstrap in every instance.

Coverage probabilities of estimated confidence intervals tell only part of the story. Consequently, we also examined the widths of the confidence intervals estimated in the experiments reported in Table 1. Table 3 shows, for the case of constant returns to scale, the mean width of the estimated confidence intervals over 1000 Monte Carlo trials, the standard deviation of these widths, and the range of the widths of the estimated confidence intervals. Up to about 100 observations, the smooth bootstrap produces narrower confidence interval estimates than either of the naive bootstrap methods. Differences beyone $n = 100$ appear inconsequential. Table 4 shows similar results for the case of variable returns to scale. Here, the differences between the smooth bootstrap and the naive variants are more pronounced at large sample sizes. The smooth bootstrap yields narrower intervals than the naive methods at smaller sample sizes, but wider intervals at very large sample sizes. With either constant or variable returns to scale, the widths of the confidence interval estimates from the smooth bootstrap have smaller standard deviations than those from the naive methods.

In Tables 5–6, we give results on the bias estimates (based on (6.10)) for the constant returns to scale case (Table 5) and the variable returns to scale case (Table 6). In both tables, the first column gives the sample size, while the second column (labelled "True") gives the Monte Carlo estimate of the true bias, *i.e.,* the mean of the original efficiency estimates $\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ produced in step [1] of Algorithm #1 over the 1000 Monte Carlo Trials, minus the known, true value $D(\boldsymbol{x}_0, \boldsymbol{y}_0) = 0.6419$. The remaining columns in the tables give the mean and standard deviation of the bias estimates (6.10) over 1000 Monte Carlo trials, as well as the range of the bias estimates.

In both the constant and variable returns to scale cases, the naive methods yield, on average, smaller and less accurate bias estimates than the smooth bootstrap, with the differences becoming more pronounced in the variable returns to scale case. With variable returns to scale, the naive methods typically give bias estimates whose averages are equal to about half the average of the corresponding bias estimates from the smooth bootstrap. The naive methods also produce bias estimates with larger standard deviation than in the case of the smooth bootstrap for sample sizes up to 100, but these differences become minimal for larger sample sizes. The naive method based on resampling (x, y)-pairs yields negative bias estimates in some cases for $n = 10$, 25, and 50 with variable returns to scale—even though bias is necessarily positive for our simulated DGP.

9. EHANCING THE PERFORMANCE OF THE BOOTSTRAP

The results of the preceeding section show that in less favorable situations, even if the bootstrap is consistent, the coverage probabilities could be poorly approximated in finite samples.

Let $\mathcal{I}_\alpha$ denote the $(1-\alpha)$-level confidence interval for $D(\boldsymbol{x}_0, \boldsymbol{y}_0)$ given by (6.8). We have:

$$\mathcal{I}_\alpha = \left[\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0) + \hat{a}_\alpha, \widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0) + \hat{b}_\alpha\right]$$

where $\hat{a}_\alpha$ and $\hat{b}_\alpha$ are solutions to (6.6). Our Monte-Carlo experiments indicate that, in finite samples, the error $\Pr(D(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{I}_\alpha) - (1-\alpha)$ may sometimes be substantial. Of course, in practice, with real data, we cannot evaluate this error as in Monte-Carlo experiments.

We show in the present section that a method based on the iterated bootstrap may be used to estimate the true coverage,

$$\pi(\alpha) = \Pr(D(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{I}_\alpha), \tag{9.1}$$

and also to estimate the value of α for which the true coverage is equal to a predetermined nominal level $(1-\alpha_0)$, such as 0.95. If $\hat{\pi}(\alpha)$ is our estimator of $\pi(\alpha)$, then we can search for a solution $\hat{\alpha}$ in the equation

$$\hat{\pi}(\hat{\alpha}) = 1 - \alpha_0, \tag{9.2}$$

and then recalibrate our intitial confidence interval $\mathcal{I}_\alpha$, chosing instead $\mathcal{I}_{\hat{\alpha}}$ as our final, corrected confidence interval for $D(\boldsymbol{x}_0, \boldsymbol{y}_0)$.

The idea of iterating the bootstrap is discussed in details in Hall (1991), and has been used in parametric frontier models by Hall *et al.* (1995). Using the notation introduced above, the method may be defined as follows. After steps [5]-[6] of Algorithm #1, where $\mathcal{S}_n^*$ and $\widehat{D}^*(\boldsymbol{x}_0, \boldsymbol{y}_0)$ have been computed, we generate a second level bootstrap sample $\mathcal{S}_n^{**}$ from $\mathcal{S}_n^*$ along the same lines by which we generated $\mathcal{S}_n^*$ from $\mathcal{S}_n$.

In particular, the second level bootstrap estimator $\widehat{D}^{**}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ correspond to $\mathcal{S}_n^{**}$. So, from the bootstrap distribution of $\widehat{D}^{**}(\boldsymbol{x}_0, \boldsymbol{y}_0)$, we can find values $\widehat{a}^*_\alpha$ and $\widehat{b}^*_\alpha$ such that

$$\Pr\left(-\widehat{b}^*_\alpha \le \widehat{D}^{**}(\boldsymbol{x}_0, \boldsymbol{y}_0) - \widehat{D}^{*}(\boldsymbol{x}_0, \boldsymbol{y}_0) \le -\widehat{a}^*_\alpha \mid \widehat{\mathcal{F}}(\mathcal{S}^*_n)\right) = 1 - \alpha \tag{9.3}$$

where $\widehat{D}^{*}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ and $\widehat{\mathcal{F}}(\mathcal{S}^*_n)$ have replaced $\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ and $\widehat{\mathcal{F}}(\mathcal{S}_n)$ in (6.6).

This will provide a confidence interval for $\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ (in the second-level bootstrap) denoted $\mathcal{I}^*_\alpha$, just as $\mathcal{I}_\alpha$ was the confidence interval for $D(\boldsymbol{x}_0, \boldsymbol{y}_0)$ in the first-level bootstrap:

$$\mathcal{I}^*_\alpha = \left[\widehat{D}^{*}(\boldsymbol{x}_0, \boldsymbol{y}_0) + \widehat{a}^*_\alpha, \widehat{D}^{*}(\boldsymbol{x}_0, \boldsymbol{y}_0) + \widehat{b}^*_\alpha\right]. \tag{9.4}$$

However, here $\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ is known; so by repeating the first-level bootstrap many times, we can record $\hat{\pi}(\alpha)$, the proportion of times that $\mathcal{I}^*_\alpha$ covers $\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)$:

$$\hat{\pi}(\alpha) = \Pr\left(\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{I}^*_\alpha \mid \widehat{\mathcal{F}}(\mathcal{S}_n)\right). \tag{9.5}$$

This quantity is the estimator of the function $\pi(\alpha)$ defined in (9.1); by solving (9.2), we obtain the iterated bootstrap confidence interval, $\mathcal{I}_{\hat{\alpha}}$.

The algorithm can be implemented as follows:

****

Algorithm #2:

[1] For each $(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathcal{S}_n$, apply one of the distance function estimators in (3.5)–(3.7) to obtain estimates $\widehat{D}(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $i = 1, \ldots, n$. Here the reference set is $\mathcal{S}_n$.

[2] Repeat step [1] for $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ to obtain $\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)$.

[3] Reflect the n estimates $\widehat{D}(\boldsymbol{x}_i, \boldsymbol{y}_i)$ about unity, and determine the bandwidth parameter h via least-squares cross-validation.

[4] Use the computational shortcut in (7.2) to draw n bootstrap values D^*_i, $i = 1, \ldots, n$, from the kernel density estimate of $f(\widehat{D})$.

[5] Construct a pseudo dataset $\mathcal{S}^*_n$ with elements $(\boldsymbol{x}^*_i, \boldsymbol{y}^*_i)$ given by

$$\boldsymbol{y}^*_i = D^*_i \boldsymbol{y}_i / \widehat{D}(\boldsymbol{x}_i, \boldsymbol{y}_i)$$

and $\boldsymbol{x}^*_i = \boldsymbol{x}$.

[6] Use (6.3) (or an analog of (6.3) if the CCR or FDH estimators were used in step [1]) to compute the bootstrap estimate $\widehat{D}^*(\boldsymbol{x}_0, \boldsymbol{y}_0)$; here the reference set $\mathcal{S}_n^*$.

[6.1] For each $(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*) \in \mathcal{S}_n^*$, apply the same distance function estimators chosen in [1] to obtain estimates $\widehat{D}^*(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)$, $i = 1, \ldots, n$, where the reference set is $\mathcal{S}_n^*$.

[6.2] Reflect the n estimates $\widehat{D}^*(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)$ about unity.

[6.3] Use the computational shortcut in (7.2) to draw n bootstrap values D_i^{**}, $i = 1, \ldots, n$, from the kernel density estimate of $f(\widehat{D}^*)$.

[6.4] Construct a pseudo dataset $\mathcal{S}_n^{**}$ with elements $(\boldsymbol{x}_i^{**}, \boldsymbol{y}_i^{**})$ given by

$$\boldsymbol{y}_i^{**} = D_i^{**} \boldsymbol{y}_i^* / \widehat{D}^*(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)$$

and $\boldsymbol{x}_i^{**} = \boldsymbol{x}_i^*$.

[6.5] Use (6.3) (or an analog of (6.3) if the CCR or FDH estimators were used in step [1]) to compute the bootstrap estimate $\widehat{D}^{**}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ with the reference set $\mathcal{S}_n^{**}$.

[6.6] Repeat steps [6.3]–[6.5] B_2 times to obtain a set of B_2 bootstrap values

$$\left\{ \widehat{D}_{b_2}^{**}(\boldsymbol{x}_0, \boldsymbol{y}_0) \right\}_{b_2=1}^{B_2}.$$

[6.7] Use (9.3) to determine $\widehat{a}_\alpha^*$ and $\widehat{b}_\alpha^*$, and then use these in (9.4) together with the estimate $\widehat{D}^*(\boldsymbol{x}_0, \boldsymbol{y}_0)$ obtained in step [6] to obtain $\mathcal{I}_\alpha^*$, an estimated confidence interval for $D^*(\boldsymbol{x}_0, \boldsymbol{y}_0)$.

[7] Repeat steps [4]–[6.7] B_1 times to obtain a set of B_1 bootstrap estimates

$$\left\{ \widehat{D}_{b_1}^{*}(\boldsymbol{x}_0, \boldsymbol{y}_0) \right\}_{b_1=1}^{B_1}$$

and a set of confidence intervals $\left\{ \mathcal{I}_{\alpha,b_1}^* \right\}_{b_1=1}^{B_1}$.

[8] Compute the proportion of time $\mathcal{I}_{\alpha,b_1}^*$ covers $\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)$:

$$\hat{\pi}_{B_1 B_2}(\alpha) = \frac{1}{B_1} \sum_{b_1=1}^{B_1} I(\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{I}_{\alpha,b_1}^*)$$

where $I(\cdot)$ is the indicator function.

[9] Solve the equation $\widehat{\pi}_{B_1 B_2}(\alpha) = 1 - \alpha_0$ for α, where $1 - \alpha_0$ is the desired, nominal coverage. Denote the solution by $\widehat{\alpha}$.

[10] Use (6.6) to determine $\widehat{a}_{\widehat{\alpha}}$ and $\widehat{b}_{\widehat{\alpha}}$, and then use these in (6.8) together with the original estimate $\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ obtained in step [2] to obtain $\mathcal{I}_{\widehat{\alpha}}$, the estimated confidence interval for $D(\boldsymbol{x}_0, \boldsymbol{y}_0)$. In addition, the bootstrap estimates can be used with (6.10) to obtain an estimate of the bias of $\widehat{D}(\boldsymbol{x}_0, \boldsymbol{y}_0)$, and with (6.11) to obtain a bias-corrected estimator if the condition in (6.13) is satisfied.

10. CONCLUSIONS

As noted in the introduction, DEA methods are still quite young compared to many existing parametric statistical techniques. Becauses of their nonparametric nature and the resulting lack of structure, obtaining asymptotic results is difficult, and even when they can be obtained, their practical use remains to be demonstrated. The bootstrap is a natural tool to apply in such cases. Our Monte Carlo evidence confirms, however, that the structure of the underlying, true model plays a crucial role in determining how well the bootstrap will perform in a given applied setting. In those cases, where the researcher does not have access to the true model as we did in our Monte Carlo experiments, the iterated bootstrap offers a convenient, analogous approach for evaluating the performance of the bootstrap and providing corrections if needed.

REFERENCES

Aigner, D., C. A. K. Lovell, and P. Schmidt (1977), Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* 6, 21–37.

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

Banker, R.D., A. Charnes, and W.W. Cooper (1984), Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science* 30, 1078–1092.

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

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

Burley, H. T. (1980), Productive efficiency in U. S. manufacturing: a linear programming approach, *Review of Economics and Statistics* 62, 619–622.

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.

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

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.

Färe, R. (1988), *Fundamentals of Production Theory,* Berlin: Springer-Verlag.

Färe, R., and C. A. K. Lovell (1978), Measuring the technical efficiency of production, *Journal of Economic Theory* 19, 150–162.

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

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

Gijbels, I., E. Mammen, B.U. Park, and L. Simar (1999), On estimation of monotone and concave frontier functions, *Journal of the American Statistical Association*, 94, 220–228.

Greene, W. H. (1980a), Maximum likelihood estimation of econometric frontier functions, *Journal of Econometrics* 13, 27–56.

Greene, W. H. (1980b), On the estimation of a felxible frontier production model, *Journal of Econometrics* 13, 101–115.

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

Hall, P., W. Härdle, and L. Simar (1995), Iterated bootstrap with application to frontier models, *The Journal of Productivity Analysis* 6, 63-76.

Jondrow, J., C. A. K. Lovell, I. S. Materov, and P. Schmidt (1982), On the estimation of technical inefficiency in the stochastic frontier production function model, *Journal of Econometrics* 19, 233–238.

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.

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.*

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

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

Meeusen, W. and J. van den Broeck (1977), Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review* 18, 435–444.

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, forthcoming in *Econometric Theory.*

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

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

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

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

Silverman, B. W. (1978), Choosing the window width when estimating a density, *Biometrika* 65, 1–11.

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

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 (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, forthcoming in *Journal of Applied Statistics.*

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.

Simar, L., and P.W. Wilson (2000), "Statistical Inference in Nonparametric Frontier Models: The State of the Art,“ *Journal of Productivity Analysis*,13, 49-78.

Stevenson, R. (1980), Likelihood functions ofr generalized stochastic frontier estimation, *Journal of Econometrics* 13, 58–66.

Swanepoel, J.W.H. (1986), A note on proving that the (modified) bootstrap works, *Communications in Statistics: Theory and Methods* 15, 3193–3203.

Wheelock, D.C., and P.W. Wilson (1996), Explaining bank failures: deposit insurance, regulation, and efficiency, *Review of Economics and Statistics* 77 (1995), 689–700.

Wheelock, D.C., and P.W. Wilson (1999), Why do banks disappear? The determinants of US bank failures and acquisitions, *Review of Economics and Statistics,* forthcoming.

Winsten, C.B. (1957), Discussion on Mr. Farrell's paper, *Journal of the Royal Statistical Society* Series A, 120, 282–284.

Zieschang, K. O. (1984), An extended Farrell technical efficiency measure, *Journal of Economic Theory* 33, 387–396.

TABLE 1

Monte Carlo Estimates of Confidence Interval Coverages
($p = q = 1$; *i.e.*, one input, one output)

n	Significance level	(1) Smooth	(2) x, y	(3) $\widehat{D}$	(4) Smooth	(5) x, y	(6) $\widehat{D}$
10	.80	0.745	0.757	0.757	0.626	0.757	0.450
	.90	0.858	0.789	0.789	0.754	0.830	0.609
	.95	0.916	0.899	0.899	0.852	0.917	0.740
	.975	0.948	0.936	0.936	0.895	0.938	0.826
	.99	0.976	0.957	0.957	0.941	0.969	0.894
25	.80	0.750	0.770	0.771	0.646	0.683	0.393
	.90	0.861	0.776	0.775	0.777	0.808	0.562
	.95	0.932	0.894	0.890	0.843	0.889	0.689
	.975	0.965	0.928	0.922	0.905	0.926	0.807
	.99	0.980	0.967	0.950	0.944	0.963	0.885
50	.80	0.752	0.769	0.765	0.645	0.683	0.369
	.90	0.863	0.783	0.778	0.770	0.808	0.515
	.95	0.920	0.896	0.891	0.842	0.889	0.640
	.975	0.950	0.934	0.941	0.901	0.926	0.751
	.99	0.972	0.962	0.960	0.947	0.963	0.850
100	.80	0.766	0.767	0.749	0.620	0.580	0.285
	.90	0.864	0.797	0.800	0.749	0.724	0.438
	.95	0.921	0.889	0.891	0.830	0.810	0.565
	.975	0.956	0.929	0.947	0.882	0.869	0.676
	.99	0.980	0.955	0.970	0.925	0.927	0.786
200	.80	0.780	0.757	0.736	0.600	0.531	0.280
	.90	0.877	0.795	0.803	0.728	0.674	0.406
	.95	0.937	0.879	0.888	0.811	0.774	0.514
	.975	0.963	0.939	0.939	0.867	0.825	0.607
	.99	0.984	0.961	0.963	0.902	0.886	0.726
400	.80	0.785	0.759	0.772	0.632	0.540	0.306
	.90	0.878	0.809	0.811	0.758	0.665	0.430
	.95	0.936	0.883	0.889	0.845	0.754	0.529
	.975	0.970	0.937	0.936	0.891	0.828	0.626
	.99	0.990	0.963	0.961	0.926	0.896	0.720

TABLE 1 (continued)

n	Significance level	(1) Smooth	(2) x, y	(3) $\widehat{D}$	(4) Smooth	(5) x, y	(6) $\widehat{D}$
800	.80	0.767	0.769	0.733	0.614	0.487	0.281
	.90	0.884	0.811	0.791	0.730	0.616	0.380
	.95	0.950	0.886	0.871	0.810	0.708	0.478
	.975	0.971	0.938	0.930	0.858	0.790	0.571
	.99	0.984	0.966	0.955	0.907	0.853	0.667
1600	.80	0.786	0.752	0.735	0.663	0.522	0.302
	.90	0.897	0.799	0.793	0.777	0.657	0.423
	.95	0.957	0.876	0.868	0.846	0.739	0.517
	.975	0.981	0.945	0.929	0.896	0.815	0.629
	.99	0.990	0.970	0.960	0.944	0.887	0.730
3200	.80	0.801	0.758	0.765	0.653	0.500	0.303
	.90	0.904	0.822	0.813	0.775	0.639	0.430
	.95	0.951	0.897	0.864	0.860	0.717	0.527
	.975	0.976	0.951	0.924	0.899	0.785	0.611
	.99	0.992	0.973	0.951	0.938	0.865	0.715
6400	.80	0.839	0.765	0.743	0.666	0.537	0.337
	.90	0.917	0.816	0.811	0.789	0.656	0.462
	.95	0.960	0.878	0.868	0.868	0.737	0.588
	.975	0.977	0.943	0.937	0.909	0.806	0.671
	.99	0.987	0.964	0.962	0.942	0.873	0.744

TABLE 2

Monte Carlo Estimates of Confidence Interval Coverages
Showing the Influence of Range of the Input for
the Variable Returns to Scale Case
($p = q = 1$, $n = 200$)

Significance	$x \sim U(6.5, 8.5)$	$x \sim U(5.5, 9.5)$	$x \sim U(4.5, 10.5)$	$x \sim U(3.5, 11.5)$
		Smooth Bootstrap		
.80	0.734	0.727	0.659	0.633
.90	0.846	0.829	0.790	0.748
.95	0.906	0.888	0.858	0.830
.975	0.946	0.930	0.902	0.886
.99	0.967	0.963	0.946	0.920
		Resample (x, y)		
.80	0.686	0.620	0.565	0.540
.90	0.813	0.763	0.719	0.689
.95	0.897	0.849	0.808	0.769
.975	0.941	0.913	0.872	0.839
.99	0.969	0.955	0.932	0.902
		Resample $\widehat{D}$		
.80	0.326	0.276	0.274	0.279
.90	0.489	0.442	0.406	0.398
.95	0.640	0.568	0.526	0.516
.975	0.744	0.686	0.635	0.615
.99	0.858	0.801	0.753	0.725

TABLE 3

Widths of Estimated Confidence Intervals for Constant Returns to Scale Case

n	Mean	Std. Dev.	Range
		Smooth Bootstrap	
10	0.1384	0.0335	0.0485—0.2838
25	0.0551	0.0099	0.0271—0.0935
50	0.0283	0.0044	0.0183—0.0477
100	0.0146	0.0019	0.0094—0.0255
200	0.0076	0.0008	0.0055—0.0110
400	0.0039	0.0004	0.0029—0.0053
800	0.0019	0.0002	0.0016—0.0026
1600	0.0010	0.0001	0.0008—0.0012
3200	0.0005	0.0001	0.0004—0.0007
6400	0.0003	0.0001	0.0002—0.0003
		Resample (x, y)	
10	0.2018	0.1404	0.0079—0.9945
25	0.0664	0.0384	0.0048—0.2556
50	0.0320	0.0186	0.0018—0.1391
100	0.0154	0.0087	0.0011—0.0527
200	0.0078	0.0046	0.0001—0.0306
400	0.0037	0.0021	0.0001—0.0141
800	0.0019	0.0011	0.0002—0.0066
1600	0.0009	0.0005	0.0 —0.0040
3200	0.0005	0.0003	0.0 —0.0018
6400	0.0002	0.0001	0.0 —0.0009
		Resample $\widehat{D}$	
10	0.2018	0.1404	0.0079—0.9945
25	0.0693	0.0431	0.0038—0.3975
50	0.0315	0.0175	0.0018—0.1218
100	0.0157	0.0089	0.0011—0.0580
200	0.0074	0.0040	0.0003—0.0295
400	0.0038	0.0022	0.0002—0.0141
800	0.0019	0.0010	0.0001—0.0072
1600	0.0009	0.0006	0.0001—0.0034
3200	0.0005	0.0003	0.0 —0.0017
6400	0.0002	0.0001	0.0 —0.0009

TABLE 4

Widths of Estimated Confidence Intervals for
Variable Returns to Scale Case

n	Mean	Std. Dev.	Range
		Smooth Bootstrap	
10	0.3020	0.1520	0.0499—2.3832
25	0.1331	0.0407	0.0426—0.3941
50	0.0741	0.0195	0.0208—0.1619
100	0.0426	0.0103	0.0181—0.0819
200	0.0254	0.0053	0.0133—0.0460
400	0.0161	0.0030	0.0081—0.0271
8000	0.0102	0.0018	0.0057—0.0167
1600	0.0066	0.0011	0.0036—0.0105
3200	0.0042	0.0007	0.0022—0.0062
6400	0.0028	0.0004	0.0016—0.0043
		Resample (x, y)	
10	0.4911	0.2834	0.0603—2.9616
25	0.1667	0.0741	0.0237—0.4875
50	0.1667	0.0741	0.0237—0.4875
100	0.0447	0.0185	0.0101—0.1252
200	0.0245	0.0098	0.0046—0.0723
400	0.0147	0.0060	0.0035—0.0417
800	0.0089	0.0034	0.0021—0.0262
1600	0.0057	0.0022	0.0012—0.0153
3200	0.0035	0.0013	0.0007—0.0088
6400	0.0022	0.0009	0.0002—0.0059
		Resample $\widehat{D}$	
10	0.2701	0.1746	0.0264—1.6783
25	0.1067	0.0529	0.0169—0.4355
50	0.0529	0.0241	0.0039—0.1742
100	0.0270	0.0125	0.0057—0.0800
200	0.0147	0.0064	0.0014—0.0392
400	0.0090	0.0037	0.0015—0.0229
800	0.0054	0.0021	0.0011—0.0123
1600	0.0036	0.0013	0.0003—0.0092
3200	0.0023	0.0008	0.0005—0.0053
6400	0.0015	0.0005	0.0003—0.0031

TABLE 5

Bootstrap Bias Estimates
Constant Returns to Scale Case

n	"True"	Mean	Std. Dev.	Range
		Smooth Bootstrap		
10	0.0517	0.0362	0.0085	0.0133—0.0756
25	0.0203	0.0147	0.0025	0.0077—0.0241
50	0.0101	0.0076	0.0011	0.0050—0.0123
100	0.0048	0.0039	0.0005	0.0025—0.0067
200	0.0024	0.0020	0.0002	0.0015—0.0029
400	0.0012	0.0010	0.0001	0.0008—0.0014
800	0.0006	0.0005	0.0001	0.0004—0.0007
1600	0.0003	0.0003	0.0000	0.0002—0.0004
3200	0.0002	0.0001	0.0000	0.0001—0.0002
6400	0.0001	0.0001	0.0	0.0001—0.0001
		Resample (x, y)		
10	0.0517	0.0324	0.0250	0.0025—0.2503
25	0.0203	0.0117	0.0082	0.0008—0.0585
50	0.0101	0.0058	0.0039	0.0003—0.0278
100	0.0048	0.0028	0.0019	0.0003—0.0152
200	0.0024	0.0014	0.0010	0.0001—0.0066
400	0.0012	0.0007	0.0005	0.0 —0.0034
800	0.0006	0.0004	0.0003	0.0 —0.0017
1600	0.0003	0.0002	0.0001	0.0 —0.0009
3200	0.0002	0.0001	0.0001	0.0 —0.0005
6400	0.0001	0.0000	0.0001	0.0 —0.0003
		Resample $\widehat{D}$		
10	0.0517	0.0324	0.0250	0.0025—0.2503
25	0.0205	0.0121	0.0085	0.0008—0.0650
50	0.0102	0.0057	0.0037	0.0003—0.0302
100	0.0049	0.0030	0.0020	0.0003—0.0135
200	0.0025	0.0014	0.0009	0.0001—0.0071
400	0.0012	0.0007	0.0005	0.0001—0.0036
800	0.0006	0.0004	0.0002	0.0 —0.0016
1600	0.0003	0.0002	0.0001	0.0 —0.0009
3200	0.0002	0.0001	0.0001	0.0 —0.0006
6400	0.0001	0.0000	0.0000	0.0 —0.0002

TABLE 6

Bootstrap Bias Estimates
Variable Returns to Scale Case

n	"True"	Mean	Std. Dev.	Range
		Smooth Bootstrap		
10	0.1615	0.1017	0.0487	0.0149—0.7556
25	0.0735	0.0511	0.0171	0.0140—0.1452
50	0.0403	0.0298	0.0090	0.0074—0.0745
100	0.0247	0.0178	0.0050	0.0066—0.0403
200	0.0150	0.0110	0.0029	0.0049—0.0216
400	0.0088	0.0071	0.0017	0.0031—0.0132
800	0.0060	0.0045	0.0010	0.0022—0.0090
1600	0.0035	0.0029	0.0007	0.0013—0.0054
3200	0.0023	0.0019	0.0004	0.0009—0.0032
6400	0.0014	0.0013	0.0003	0.0007—0.0021
		Resample (x, y)		
10	0.1615	0.2263	0.5805	−3.3019—0.3282
25	0.0735	0.0337	0.0320	−0.5655—0.1450
50	0.0735	0.0337	0.0320	−0.5655—0.1450
100	0.0247	0.0100	0.0048	0.0015—0.0346
200	0.0150	0.0057	0.0028	0.0013—0.0183
400	0.0088	0.0034	0.0016	0.0006—0.0123
800	0.0060	0.0020	0.0010	0.0004—0.0064
1600	0.0035	0.0013	0.0006	0.0002—0.0037
3200	0.0023	0.0008	0.0004	0.0001—0.0021
6400	0.0013	0.0005	0.0002	0.0001—0.0017
		Resample $\widehat{D}$		
10	0.1615	0.0506	0.0412	0.0030—0.6932
25	0.0735	0.0207	0.0142	0.0017—0.1226
50	0.0403	0.0101	0.0069	0.0005—0.0642
100	0.0247	0.0052	0.0035	0.0004—0.0224
200	0.0150	0.0029	0.0020	0.0002—0.0126
400	0.0088	0.0018	0.0012	0.0001—0.0076
800	0.0060	0.0011	0.0007	0.0001—0.0044
1600	0.0035	0.0008	0.0005	0.0 —0.0030
3200	0.0023	0.0005	0.0003	0.0001—0.0016
6400	0.0013	0.0003	0.0002	0.0 —0.0010