Testing Restrictions in Nonparametric Efficiency Models

Léopold Simar

Institut de Statistique
Université Catholique de Louvain
Voie du Roman Pays 20
Louvain-la-Neuve, Belgium

and

Paul W. Wilson*

Department of Economics
University of Texas
Austin, Texas 78712 USA

April 2000

ABSTRACT

This paper discusses statistical procedures for testing various restrictions in the context of nonparametric models of technical efficiency. In particular, tests for whether inputs or outputs are irrelevant, as well as tests of whether inputs or outputs may be aggregated are formulated. Bootstrap estimation procedures which yield appropriate critical values for the test statistics are also provided. Evidence on the true sizes and power of the proposed tests is obtained from Monte Carlo experiments.

*We are grateful for financial support from the contract "Projet d'Actions de Recherche Concertées" (PARC No. 98/03–217) of the Belgian Government and from the Management Science Group, US Department of Veterans Affairs. In addition, we are especially grateful to the Texas Advanced Computing Center (TACC) for a grant of computational time on their Cray T3E parallel machine, and to Robert Harkness and Kent Milfeld of the TACC staff for help in porting code to the T3E environment. Any remaining errors, of course, are solely our responsibility.

1. INTRODUCTION

Linear-programming based measures of efficiency along the lines of Charnes *et al.* (1978, 1979), Deprins *et al.* (1984), and Färe *et al.* (1985) are widely used in the analysis of efficiency of production.[1] These methods are based on definitions of technical and allocative efficiency in production provided by Debreu (1951) and Farrell (1957). Among this literature, those approaches which incorporate convexity assumptions are known as Data Envelopment Analysis (DEA).

DEA measures efficiency relative to a nonparametric, maximum likelihood *estimate* of an unobserved *true* frontier, conditional on observed data resulting from an underlying (and unobserved) data-generating process (DGP).[2] These methods have been widely applied to examine technical and allocative efficiency in a variety of industries; see Lovell (1993) and Seiford (1996, 1997) for comprehensive bibliographies of these applications. Aside from the production setting, the problem of estimating monotone concave boundaries also naturally occurs in portfolio management. In capital asset pricing models (CAPM), the objective is to analyze the performance of investment portfolios. Risk and average return on a portfolio are analogous to inputs and outputs in models of production; in CAPM, the attainable set of portfolios is naturally convex and the boundary of this set gives a benchmark relative to which the efficiency of a portfolio can be measured. These models were developed by Markovitz (1959) and others; Sengupta (1991) and Sengupta and Park (1993) provide links between CAPM and nonparametric estimation of frontiers as in DEA.

DEA and similar approaches to efficiency measurement are frequently referred to as *deterministic,* as if to suggest that DEA models have no statistical underpinnings. Yet, since efficiency is measured relative to an estimate of the frontier, estimates of efficiency from DEA models are subject to uncertainty due to sampling variation; bootstrap methods may be used to assess this uncertainty by estimating confidence intervals, *etc.* (see Simar and

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

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

Wilson, 1998a, 2000a, 2000b). In addition, there may be uncertainty about the structure of the underlying statistical model in terms of whether certain variables are relevant or whether subsets of variables may be aggregated. This paper addresses this second problem by providing tests of hypotheses about the model structure.

Banker (1993) proved the consistency of DEA *output*-oriented efficiency scores in the case of a *single* output, but gives no indication of the achieved rate of convergence. Korostelev *et al.* (1995) also analyzed the single-output problem and derived the speed of convergence for the estimated attainable production set (using the Lebesgue measure of symmetric differences between the true and the estimated production sets), but not for the estimated measures of efficiency. The theory of statistical consistency in DEA models has been extended to the general multi-input *and* multi-output case for both input- and output-oriented efficiency measures in Kneip *et al.* (1998), where the rates of convergence are also derived.

As with most nonparametric estimators, DEA efficiency estimators suffer from the well-known curse of dimensionality; *i.e.,* convergence rates become slower as the number of inputs and outputs is increased. In the context of nonparametric curve-fitting, various procedures for dimension-reduction have been proposed (see Scott, 1992, chapters 7–8, for discussion), but the radial nature of DEA estimators seems to preclude these approaches in DEA efficiency estimation. These facts make it imperative to avoid inclusion of irrelevant variables in the analysis, as well as to exploit any opportunities for aggregation that might exist, giving rise to the need for statistical tests of hypotheses about model structure.

Due to the complexity and multidimensional nature of DEA estimators, the sampling distributions of the estimators are not easily available. Consequently, distributions of test statistics constructed from these estimators remain unknown. In the very particular case of one-input and one-output, Gijbels *et al.* (1999) derived the asymptotic sampling distribution of the DEA estimator, with an expression for its asymptotic bias and variance. However, in the more useful multi-output and multi-input case, the bootstrap methodology seems, so far, to be the only way to investigate sampling properties of DEA estimators.

Simar and Wilson (1998a, 2000a) proposed bootstrap strategies for analyzing the sampling variation of efficiency measures.

Here, we extend the methods in Simar and Wilson (1998a, 2000a) to construct tests of various hypotheses and estimate sampling distributions of test statistics constructed from DEA estimators. Section 2 defines our notation and the statistical model. In section 3 we develop statistical tests for whether either subsets of inputs or of outputs may be irrelevant to the production process. In section 4 we propose tests of whether subsets of inputs or of outputs may be aggregated. Various test statistics are defined in section 5, and the bootstrap procedure for implementing the tests is discussed in section 6. Section 7 details results of Monte Carlo experiments to analyze the size and power of our tests, and conclusions are discussed in the final section.

2. NONPARAMETRIC PRODUCTION MODELS

2.1. The Economic Model:

Standard microeconomics texts develop the theory of the firm by positing a production set which describes how a set of inputs may somehow be converted into outputs. A brief outline of the standard story is necessary to define notation and quantities to be estimated. 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 closed and convex; $\mathcal{Y}(\boldsymbol{x})$ is closed, convex, and bounded for all $\boldsymbol{x} \in \mathbb{R}^p_+$; and $\mathcal{X}(\boldsymbol{y})$ is closed and convex for all $\boldsymbol{y} \in \mathbb{R}^q_+$.*

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

Assumption A2 merely says that there are no free lunches.[3]

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.*

The boundary of $\mathcal{P}$ is sometimes referred to as the ***technology*** or the ***production frontier,*** and is given by the intersection of $\mathcal{P}$ and the closure of its compliment. Assumption A3 is sometimes called free disposability and is equivalent to an assumption of monotonicity of the technology, which may now be defined as

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

Other assumptions are possible, but might require small modifications in the methods we propose later. 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.

Isoquants are defined by the intersection of $\mathcal{X}(\boldsymbol{y})$ and the closure of its complement, or

$$\mathcal{X}^{\partial}(\boldsymbol{y}) = \{\boldsymbol{x} \mid \boldsymbol{x} \in \mathcal{X}(\boldsymbol{y}),\ \theta\boldsymbol{x} \notin \mathcal{X}(\boldsymbol{y})\ \forall\ 0 < \theta < 1\}. \tag{2.5}$$

[3]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 $\boldsymbol{x}$.

Similarly, the intersection of $\mathcal{Y}(\boldsymbol{x})$ and the closure of its compliment give iso-output curves,

$$\mathcal{Y}^{\partial}(\boldsymbol{x}) = \{\boldsymbol{y} \mid \boldsymbol{y} \in \mathcal{Y}(\boldsymbol{x}),\ \theta^{-1}\boldsymbol{y} \notin \mathcal{Y}(\boldsymbol{x})\ \forall\ 0 < \theta < 1\}. \tag{2.6}$$

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 $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_{+}^{p+q}$ to $\mathcal{P}^{\partial}$ in a radial direction orthogonal to $\boldsymbol{x}$, and may be defined as

$$D^{out}(\boldsymbol{x}, \boldsymbol{y}) \equiv \inf\{\theta > 0 | (\boldsymbol{x}, \theta^{-1}\boldsymbol{y}) \in \mathcal{P}\}. \tag{2.7}$$

Clearly, $D^{out}(\boldsymbol{x}, \boldsymbol{y}) \leq 1$ for all $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$. If $D^{out}(\boldsymbol{x}, \boldsymbol{y}) = 1$ 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.

Alternatively, the Shephard (1970) input distance function provides a normalized measure of Euclidean distance from a point $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_{+}^{p+q}$ to $\mathcal{P}^{\partial}$ in a direction orthogonal to $\boldsymbol{y}$, and may be defined as

$$D^{in}(\boldsymbol{x}, \boldsymbol{y}) \equiv \sup\{\theta > 0 | (\theta^{-1}\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}\}, \tag{2.8}$$

with $D^{in}(\boldsymbol{x}, \boldsymbol{y}) \geq 1$ for all $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$. If $D^{in}(\boldsymbol{x}, \boldsymbol{y}) = 1$ then $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^{\partial}$; *i.e.,* if $D^{in}(\boldsymbol{x}, \boldsymbol{y}) = 1$ then the point $(\boldsymbol{x}, \boldsymbol{y})$ lies on the boundary of $\mathcal{P}$. Note that no behavioral assumptions are necessary for measuring technical efficiency. From a purely technical viewpoint, either the input or the output distance function can be used to measure technical efficiency—the only difference is in the direction in which distance to the technology is measured.[4]

Given the output distance function in (2.7), the point $(\boldsymbol{x}^{\partial}(\boldsymbol{y}), \boldsymbol{y})$, where

$$\boldsymbol{x}^{\partial}(\boldsymbol{y}) = \boldsymbol{x}/D^{in}(\boldsymbol{x}, \boldsymbol{y}) \tag{2.9}$$

represents the projection of $(\boldsymbol{x}, \boldsymbol{y})$ onto $\mathcal{P}^{\partial}$ along the ray $(\delta\boldsymbol{x}, \boldsymbol{y})$, $\delta \in [0, \infty)$. Similarly, the point $(\boldsymbol{x}, \boldsymbol{y}^{\partial}(\boldsymbol{x}))$, where

$$\boldsymbol{y}^{\partial}(\boldsymbol{x}) = \boldsymbol{y}/D^{out}(\boldsymbol{x}, \boldsymbol{y}) \tag{2.10}$$

[4] For purposes of assessing technical efficiency, one can also employ hyperbolic graph measures, where inputs are reduced and outputs are expanded (simultaneously, by the same proportions) to reach the technology $\mathcal{P}^{\partial}$. See Färe *et al.* (1985) for details.

represents the projection of $(\boldsymbol{x}, \boldsymbol{y})$ onto $\mathcal{P}^{\partial}$ along the ray $(\boldsymbol{x}, \delta\boldsymbol{y})$, $\delta \in [0, \infty)$. The points $(\boldsymbol{x}^{\partial}(\boldsymbol{y}), \boldsymbol{y})$ and $(\boldsymbol{x}, \boldsymbol{y}^{\partial}(\boldsymbol{x}))$ are technically efficient, and represent two possible targets for an inefficient firm producing at the point $(\boldsymbol{x}, \boldsymbol{y})$ in the interior of $\mathcal{P}$.

Standard microeconomic theory suggests that with perfectly competitive input and output markets, firms which are either technically or allocatively inefficient will be driven from the market.[5] 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 (1995, 2000) provide support for this view through empirical evidence for banks operating in the US.

Equations (2.1)–(2.6), together with the assumptions listed above, constitute the ***true*** economic model of production. Unfortunately, these quantities, as well as the distance function values given by (2.7)–(2.8) and the quantities defined in (2.9)–(2.10), are unobservable and unknown, and consequently must be ***estimated.*** Hypotheses regarding $\mathcal{P}$ or $\mathcal{P}^{\partial}$ can then be tested using estimates of the distance functions defined in (2.7)–(2.8).

2.2. The Statistical Model:

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

$$\mathcal{S}_n = \{(\boldsymbol{x}_i, \boldsymbol{y}_i) \mid i = 1, \ldots, n\}. \tag{2.11}$$

Before anything can be estimated, however, a statistical model must be defined by augmenting the economic assumptions A1-A3 with some appropriate assumptions on the data-generating process (DGP). Our assumptions are based on those of Kneip *et al.* (1998).

[5]By *allocatively inefficient,* we mean suboptimal locations on $\mathcal{P}^{\partial}$ given vectors of input and output prices.

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 $\mathcal{P}$.*

Adopting an output orientation, 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 $\boldsymbol{y} \in \mathbb{R}_+^q$. The modulus is $\omega = \omega(\boldsymbol{y}) = \sqrt{\boldsymbol{y}'\boldsymbol{y}} \in \mathbb{R}_+^1$ 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

$$f(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x}, \omega, \boldsymbol{\eta}) = f(\omega \mid \boldsymbol{x}, \boldsymbol{\eta}) f(\boldsymbol{\eta} \mid \boldsymbol{x}) f(\boldsymbol{x}) \tag{2.12}$$

where 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

$$0 \leq D^{out}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\omega(\boldsymbol{y})}{\omega(\boldsymbol{y}^\partial(\boldsymbol{x}))} \leq 1. \tag{2.13}$$

Then the density $f(\omega \mid \boldsymbol{x}, \boldsymbol{\eta})$ on $[0, \omega(\boldsymbol{y}^\partial(\boldsymbol{x}))]$ implies a density $f(D^{out}(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x}, \boldsymbol{\eta})$ on the interval $[0, 1]$.

Switching to an input orientation for the moment, and applying similar reasoning, $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{p+q}$ can also be represented by a different set of cylindrical coordinates, namely $(\tau, \boldsymbol{\phi}, \boldsymbol{y})$ where $(\tau, \boldsymbol{\phi})$ gives the polar coordinates of $\boldsymbol{x}$ with modulus $\tau = \tau(\boldsymbol{x}) = \sqrt{\boldsymbol{x}'\boldsymbol{x}} \in \mathbb{R}_+^1$ and angles $\boldsymbol{\phi} = \boldsymbol{\phi}(\boldsymbol{x}) \in [0, \pi/2]^{p-1}$, where the jth element of $\boldsymbol{\phi}(\boldsymbol{x})$ is given by $\arctan(x^{j+1}/x^1)$ for $x^1 \neq 0$ (where x^j represents the jth element of $\boldsymbol{x}$) or 0 otherwise. Then (2.12) can be rewritten as

$$f(\boldsymbol{x}, \boldsymbol{y}) = f(\tau, \boldsymbol{\phi}, \boldsymbol{y}) = f(\tau \mid \boldsymbol{\phi}, \boldsymbol{y}) f(\boldsymbol{\phi} \mid \boldsymbol{y}) f(\boldsymbol{y}) \tag{2.14}$$

where again all the conditional densities exist; $f(\boldsymbol{y})$ is defined on $\mathbb{R}^q_+$, $f(\boldsymbol{\eta} \mid \boldsymbol{y})$ is defined on $[0, \pi/2]^{p-1}$, and $f(\tau \mid \boldsymbol{\phi}, \boldsymbol{y})$ is defined on $\mathbb{R}^1_+$. Similarly, (2.13) can be rewritten as

$$D^{in}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\tau(\boldsymbol{x})}{\tau(\boldsymbol{x}^{\partial}(\boldsymbol{y}))} \geq 1 \tag{2.15}$$

for $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$. Thus, similar to the output orientation, the density $f(\tau \mid \boldsymbol{\phi y})$ on the interval $\left[0,\ \tau(\boldsymbol{x}^{\partial}(\boldsymbol{y}))\right]$ implies a density $f\left(D^{in}(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{\phi}, \boldsymbol{y}\right)$ on the interval $[1, \infty)$.

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

Assumption A5: *for all $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}^{\partial}$, $f(\boldsymbol{x}, \boldsymbol{y})$ is strictly positive, and $f(\boldsymbol{x}, \boldsymbol{y})$ is continuous in any direction toward the interior of $\mathcal{P}$.*

In addition, an assumption about the smoothness of the frontier is needed:[6]

Assumption A6: *for all $(\boldsymbol{x}, \boldsymbol{y})$ in the interior of $\mathcal{P}$, $D^{out}(\boldsymbol{x}, \boldsymbol{y})$ and $D^{in}(\boldsymbol{x}, \boldsymbol{y})$ are differentiable in both their arguments.*

Assumptions A1–A6 define the DGP $\mathcal{F}$ which yields the data in $\mathcal{S}_n$.

2.3. Estimation:

The convex hull of the free disposal hull of the sample observations in $\mathcal{S}_n$, denoted $\widehat{\mathcal{P}}$, has frequently been used to estimate the production set $\mathcal{P}$. Korostelev *et al.* (1995) proved that $\widehat{\mathcal{P}}$ is a consistent estimator of $\mathcal{P}$ under conditions met by assumptions A1–A5 above. Estimators of the distance functions $D^{out}(\boldsymbol{x}, \boldsymbol{y})$ and $D^{in}(\boldsymbol{x}, \boldsymbol{y})$ can be constructed by replacing $\mathcal{P}$ on the right-hand sides of (2.7)–(2.8) with $\widehat{\mathcal{P}}$. For practical purposes, the estimators can be written in terms of linear programs

$$\left[\widehat{D}^{out}(\boldsymbol{x}, \boldsymbol{y})\right]^{-1} = \max\left\{\theta > 0 \mid \theta \boldsymbol{y} \leq \boldsymbol{Y}\boldsymbol{q},\ \boldsymbol{x} \geq \boldsymbol{X}\boldsymbol{q},\ \boldsymbol{i}'\boldsymbol{q} = 1,\ \boldsymbol{q} \in \mathbb{R}^n_+\right\} \tag{2.16}$$

[6] Our characterization of the smoothness condition here is stronger than required; Kneip *et al.* (1998) require only Lipschitz continuity for the distance functions, which is implied by the simpler, but stronger requirement presented here.

and

$$\left[\widehat{D}^{in}(\boldsymbol{x},\boldsymbol{y})\right]^{-1} = \min\left\{\theta > 0 \mid \boldsymbol{Y}\boldsymbol{q} \geq \boldsymbol{y},\ \boldsymbol{X}\boldsymbol{q} \leq \theta\boldsymbol{x},\ \boldsymbol{i}'\boldsymbol{q} = 1,\ \boldsymbol{q} \in \mathbb{R}^n_+\right\}, \tag{2.17}$$

where $\boldsymbol{Y} = [\,\boldsymbol{y}_1 \quad \ldots \quad \boldsymbol{y}_n\,]$, $\boldsymbol{X} = [\,\boldsymbol{x}_1 \quad \ldots \quad \boldsymbol{x}_n\,]$, $\boldsymbol{i}$ denotes an $(n \times 1)$ vector of ones, and $\boldsymbol{q}$ is an $(n \times 1)$ vector of intensity variables whose values are determined by solution of the linear programs in each case.

The estimators $\widehat{D}^{out}(\boldsymbol{x},\boldsymbol{y})$ and $\widehat{D}^{in}(\boldsymbol{x},\boldsymbol{y})$ measure distance in particular directions from a point $(\boldsymbol{x},\boldsymbol{y})$ to the boundary of $\widehat{\mathcal{P}}$, and hence give estimates of the distances in these directions to $\mathcal{P}^{\partial}$. Kneip *et al.* (1998) proved that for a fixed point $(\boldsymbol{x},\boldsymbol{y})$, given assumptions A1–A6, $\widehat{D}^{in}(\boldsymbol{x},\boldsymbol{y})$ consistently estimates $D^{in}(\boldsymbol{x},\boldsymbol{y})$ with

$$\widehat{D}^{in}(\boldsymbol{x},\boldsymbol{y}) - D^{in}(\boldsymbol{x},\boldsymbol{y}) = O_p(n^{-\frac{2}{p+q+1}}). \tag{2.18}$$

The rate of convergence is low, as is typical in nonparametric estimation, and the rate slows as $p+q$ is increased—this is the curse of dimensionality mentioned in the introduction. Moreover, by construction, $\widehat{D}^{in}(\boldsymbol{x},\boldsymbol{y})$ is biased downward. It is straightforward to prove that $\widehat{D}^{out}(\boldsymbol{x},\boldsymbol{y})$ is a consistent estimator of $D^{out}(\boldsymbol{x},\boldsymbol{y})$ with the same rate of convergence by altering the notation in Kneip *et al.* (1998). Few results exist on the sampling distribution of the distance function estimators in (2.16)–(2.17). Gijbels *et al.* (1999) derived the asymptotic distribution of $\widehat{D}^{out}(\boldsymbol{x},\boldsymbol{y})$ in the special case of one input and one output ($p = q = 1$), along with an analytic expression for its large sample bias and variance, and it is similarly straightforward to extend these results to the input-oriented case by appropriate changes in notation. Unfortunately, in the more general multivariate setting where $p+q > 2$, the radial nature of the distance functions and the complexity of the estimated frontier complicates the derivations. So far, the bootstrap appears to offer the only way to approximate asymptotic distribution of the distance function estimators in multivariate settings. For the case of test statistics constructed from the distance function estimators, the resulting complexity begs for bootstrap methods to obtain critical values even in the special case where $p = q = 1$.

3. TESTS FOR IRRELEVANT INPUTS OR OUTPUTS

3.1. Irrelevant Inputs:

Both the economic and the statistical models defined in the previous section assume that p and q are known. However, in applications, there is often uncertainty about one or perhaps both of these. Given the curse of dimensionality in DEA estimation, it is very important to eliminate any inputs or outputs that are not truly part of the production process. We first consider the case of possibly irrelevant inputs. Suppose the researcher is willing to accept that $p - r > 0$ inputs are truly part of the model, but that some uncertainty exists about whether a subset of size r, $0 < r < p$, of the p inputs under consideration are truly part of the production process.

Let $\boldsymbol{x} = [\,\boldsymbol{x}^{1\prime} \quad \boldsymbol{x}^{2\prime}\,]' \in \mathbb{R}_+^p$, with $\boldsymbol{x}^1 \in \mathbb{R}_+^{p-r}$ representing the $(p-r)$-vector of "known" inputs and $\boldsymbol{x}^2 \in \mathbb{R}_+^r$ representing the r-vector of possibly irrelevant inputs.[7] We wish to test the null hypothesis that $\boldsymbol{x}^2$ does not contribute to the production of $\boldsymbol{y}$. If this null is true, then $\boldsymbol{y}$ is only influenced by $\boldsymbol{x}^1$ and not by $\boldsymbol{x}^2$; but if the null is false, then the feasible quantities of $\boldsymbol{y}$ depend not only on $\boldsymbol{x}^1$ but also on $\boldsymbol{x}^2$. Under the null, the production set $\mathcal{P}$ is still defined by (2.1), but has a particular shape. The null and alternative hypotheses may be stated formally in terms of the distance functions introduced in section 2.1:

$$\text{Test \#1} \quad \begin{cases} H_0: & D^{out}(\boldsymbol{x}^1, \boldsymbol{x}^2, \boldsymbol{y}) = D^{out}(\boldsymbol{x}^1, \widetilde{\boldsymbol{x}}^2, \boldsymbol{y}) \leq 1 \text{ for all} \\ & \qquad\qquad (\boldsymbol{x}^1, \boldsymbol{x}^2, \boldsymbol{y}),\ (\boldsymbol{x}^1, \widetilde{\boldsymbol{x}}^2, \boldsymbol{y}) \in \mathcal{P}; \\ H_1: & D^{out}(\boldsymbol{x}^1, \boldsymbol{x}^2 \boldsymbol{y}) < D^{out}(\boldsymbol{x}^1, \widetilde{\boldsymbol{x}}^2, \boldsymbol{y}) \leq 1 \text{ for some} \\ & \qquad\qquad (\boldsymbol{x}^1, \boldsymbol{x}^2, \boldsymbol{y}),\ (\boldsymbol{x}^1, \widetilde{\boldsymbol{x}}^2, \boldsymbol{y}) \in \mathcal{P},\ \widetilde{\boldsymbol{x}}^2 \underset{\neq}{\geq} \boldsymbol{x}^2. \end{cases}$$

If the null hypothesis is true, then computing

$$\left[\widehat{D}_0^{out}(\boldsymbol{x}^1, \boldsymbol{y})\right]^{-1} = \max\left\{\theta > 0 \mid \theta \boldsymbol{y} \leq \boldsymbol{Y}\boldsymbol{q},\ \boldsymbol{x}^1 \geq \boldsymbol{X}^1 \boldsymbol{q},\ \boldsymbol{i}'\boldsymbol{q} = 1,\ \boldsymbol{q} \in \mathbb{R}_+^n\right\}, \tag{3.1}$$

where $\boldsymbol{X}^1 = [\,\boldsymbol{x}_1^1 \quad \ldots \quad \boldsymbol{x}_n^1\,]$, should give values $\widehat{D}_0^{out}(\boldsymbol{x}^1, \boldsymbol{y})$ that are "close" to those obtained by computing the distance function estimator defined in (2.16) for the same point $(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}_1, \boldsymbol{x}^2, \boldsymbol{y})$ (noting that $\boldsymbol{x}^2$ is not used in (3.1)). In (3.1), any possible

[7] Having already used subscripts in the previous section to index observations in $\mathcal{S}_n$, here we use superscripts to denote elements of the partition of $\boldsymbol{x}$.

role of $\boldsymbol{x}^2$ is ignored, and the space containing the production set has been collapsed to a subset of $\mathbb{R}_+^{p-r+q}$ from a subset of $\mathbb{R}_+^{p+q}$. Of course, in finite samples, sampling variation and the geometry of the estimator $\widehat{\mathcal{P}}$ will lead to $1 \geq \widehat{D}_0^{out}(\boldsymbol{x}^1, \boldsymbol{y}) \geq \widehat{D}^{out}(\boldsymbol{x}, \boldsymbol{y})$, and thus the question becomes one of whether these differences are large enough to cast doubt on the null hypothesis H_0 in Test #1.

To answer this question, we can define a test statistic and then use bootstrap methods to estimate its distribution, which will yield critical values and thus allow us to implement Test #1. Unfortunately, however, existing evidence provides almost no guidance to tell us the proper scaling needed to define pivotal, or even asymptotically pivotal statistics. Consequently, we define a variety of statistics in section 5 and investigate the sizes of the resulting tests through Monte Carlo experiments as discussed in section 7.

3.2 Irrelevant outputs:

Rather than testing for irrelevant inputs, the researcher may wish to test for irrelevant outputs. Similar to the preceding story, suppose the researcher is willing to accept that $q - r > 0$ outputs are truly part of the model, but that some uncertainty exists about whether a subset of size r, $0 < r < q$, of the q outputs under consideration are truly part of the model.

Let $\boldsymbol{y} = [\,\boldsymbol{y}^{1\prime} \quad \boldsymbol{y}^{2\prime}\,]' \in \mathbb{R}_+^q$, with $\boldsymbol{y}^1 \in \mathbb{R}_+^{q-r}$ representing the vector of "known" outputs and $\boldsymbol{y}^2 \in \mathbb{R}_+^r$ representing the vector of possibly irrelevant outputs. The problem here is analogous to the previous case; with $\boldsymbol{y}^2$ possibly irrelevant, the null hypothesis is that $\boldsymbol{x}$ influences the level of $\boldsymbol{y}^1$, but not of $\boldsymbol{y}^2$. Also similar to the previous case, the null and alternative hypotheses may be stated in terms of input distance functions:

$$\text{Test \#2} \quad \begin{cases} H_0: & D^{in}(\boldsymbol{x}, \boldsymbol{y}^1, \boldsymbol{y}^2) = D^{in}(\boldsymbol{x}, \boldsymbol{y}^1, \widetilde{\boldsymbol{y}}^2) \geq 1 \text{ for all} \\ & \qquad\qquad (\boldsymbol{x}, \boldsymbol{y}^1, \boldsymbol{y}^2),\ (\boldsymbol{x}, \boldsymbol{y}^1, \widetilde{\boldsymbol{y}}^2) \in \mathcal{P}; \\ H_1: & D^{in}(\boldsymbol{x}, \boldsymbol{y}^1, \boldsymbol{y}^2) > D^{in}(\boldsymbol{x}, \boldsymbol{y}^1, \widetilde{\boldsymbol{y}}^2) \geq 1 \text{ for some} \\ & \qquad\qquad (\boldsymbol{x}, \boldsymbol{y}^1, \boldsymbol{y}^2),\ (\boldsymbol{x}, \boldsymbol{y}^1, \widetilde{\boldsymbol{y}}^2) \in \mathcal{P};\ \widetilde{\boldsymbol{y}}^2 \underset{\neq}{\geq} \boldsymbol{y}^2. \end{cases}$$

Analogous to (3.1), but in the input orientation, we can compute

$$\left[\widehat{D}_0^{in}(\boldsymbol{x}, \boldsymbol{y}^1)\right]^{-1} = \min\left\{\theta > 0 \mid \boldsymbol{y}^1 \leq \boldsymbol{Y}^1\boldsymbol{q},\ \boldsymbol{X}\boldsymbol{q} \leq \theta\boldsymbol{x},\ \boldsymbol{i}\boldsymbol{\tau} = 1,\ \boldsymbol{q} \in \mathbb{R}_+^n\right\}, \tag{3.2}$$

where $\boldsymbol{Y}^1 = [\, \boldsymbol{y}_1^1 \quad \ldots \quad \boldsymbol{y}_n^1 \,]$. If the null hypothesis in Test #2 is true, then (3.2) should yield values $\widehat{D}_0^{in}(\boldsymbol{x}, \boldsymbol{y}^1)$ that are "close" to those obtained by computing the distance function estimator defined in (2.17) for the same point, namely $(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{y}^1, \boldsymbol{y}^2)$. Following the reasoning in section 3.1, sampling variation and the geometry of $\widehat{\mathcal{P}}$ will yield values such that $\widehat{D}_0^{in}(\boldsymbol{x}, \boldsymbol{y}^1) \geq \widehat{D}^{in}(\boldsymbol{x}, \boldsymbol{y}^1, \boldsymbol{y}^2) \geq 1$, and so the question is whether $\widehat{D}_0^{in}(\boldsymbol{x}, \boldsymbol{y}^1)$ and $\widehat{D}^{in}(\boldsymbol{x}, \boldsymbol{y}^1, \boldsymbol{y}^2)$ differ enough to cast doubt on the null hypothesis in Test #2. Before answering this question, however, we examine a related set of tests which will involve similar questions.

4. TESTS FOR ADDITIVITY OF INPUTS OR OUTPUTS

4.1. Additive Inputs:

Even if the researcher is willing to accept a set of p inputs and q outputs as belonging to the true model, he may wish to test whether certain outputs or inputs may be aggregated. Again, given the curse of dimensionality in nonparametric estimation reflected by the convergence rate shown in (2.18), it is clearly desirable to reduce the dimensions of the input/output space by aggregation if it is appropriate.[8]

To illustrate our ideas, let $1 < r \leq p$ and consider the question of whether the technology is such that r inputs $(p - r + 1), \ldots, p$ may be aggregated. Partition the input vector as before, so that $\boldsymbol{x} = [\, \boldsymbol{x}^{1\prime} \quad \boldsymbol{x}^{2\prime} \,]' \in \mathbb{R}_+^p$ with $\boldsymbol{x}^1 \in \mathbb{R}_+^{p-r}$ and $\boldsymbol{x}^2 \in \mathbb{R}_+^r$ (note that when $r = p$, $\boldsymbol{x}^1$ has zero dimension; *i.e.,* it does not exist). Let $\boldsymbol{i}_r$ denote an $r \times 1$ vector of ones, and let $\boldsymbol{0}_{r-1}$ represent an $(r-1) \times 1$ vector of zeros. Define $\boldsymbol{x}^a = [\, \boldsymbol{x}^{1\prime} \quad \boldsymbol{i}_r' \boldsymbol{x}^2 \quad \boldsymbol{0}_{r-1}' \,]'$. Then the null and alternative hypotheses may be stated as

$$\text{Test \#3} \quad \begin{cases} H_0: & 1 \geq D^{out}(\boldsymbol{x}^a, \boldsymbol{y}) = D^{out}(\boldsymbol{x}, \boldsymbol{y}) \;\forall\; (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}; \\ H_1: & 1 \geq D^{out}(\boldsymbol{x}^a, \boldsymbol{y}) > D^{out}(\boldsymbol{x}, \boldsymbol{y}) \text{ for some } (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}. \end{cases}$$

The direction of the inequality $D^{out}(\boldsymbol{x}^a, \boldsymbol{y}) > D^{out}(\boldsymbol{x}, \boldsymbol{y})$ under H_1 is due to convexity of the input requirement sets defined in (2.3).

[8]The discussion in this section assumes that outputs or inputs which are to be aggregated are measured in the same units. Otherwise, it may be necessary to introduce a conversion or scale factor in the aggregation.

Now define $\boldsymbol{x}^+ = [\,\boldsymbol{x}^{1'} \quad \boldsymbol{i}_r'\boldsymbol{x}^2\,]'$. Then we can compute

$$\left[\widehat{D}_0^{out}(\boldsymbol{x}^+, \boldsymbol{y})\right]^{-1} = \max\left\{\theta > 0 \mid \theta\boldsymbol{y} \le \boldsymbol{Y}\boldsymbol{q},\ \boldsymbol{x}^+ \ge \boldsymbol{X}^+\boldsymbol{q},\ \boldsymbol{i}\boldsymbol{\tau} = 1,\ \boldsymbol{q} \in \mathbb{R}_+^n\right\}, \qquad (4.1)$$

where $\boldsymbol{X}^+ = [\,\boldsymbol{x}_1^+ \quad \ldots \quad \boldsymbol{x}_n^+\,]$. For a given point $(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}^1, \boldsymbol{x}^2, \boldsymbol{y})$, $\widehat{D}_0^{out}(\boldsymbol{x}^+, \boldsymbol{y})$ should be "close" to $\widehat{D}^{out}(\boldsymbol{x}, \boldsymbol{y})$ when the null hypothesis in Test #3 is true. Similar to our earlier remarks, however, in finite samples, sampling variation and the geometry of $\widehat{\mathcal{P}}$ will cause $\widehat{D}_0^{out}(\boldsymbol{x}^+, \boldsymbol{y}) \ge \widehat{D}^{out}(\boldsymbol{x}, \boldsymbol{y}) \ge 1$, and so the question is whether the differences between $\widehat{D}_0^{out}(\boldsymbol{x}^+, \boldsymbol{y})$ and $\widehat{D}^{out}(\boldsymbol{x}, \boldsymbol{y})$ are large enough to cast doubt on the null.

4.2. Additive Outputs:

Alternatively, for $1 < r \le q$, the researcher may wish to test whether the outputs $(q - r + 1), \ldots, q$ may be aggregated. Partition the output vector so that $\boldsymbol{y} = [\,\boldsymbol{y}^{1'} \quad \boldsymbol{y}^{2'}\,]' \in \mathbb{R}_+^q$ with $\boldsymbol{y}^1 \in \mathbb{R}_+^{q-r}$ and $\boldsymbol{y}^2 \in \mathbb{R}_+^r$. With $\boldsymbol{i}_r$ and $\boldsymbol{0}_{r-1}$ defined as above, let $\boldsymbol{y}^a = [\,\boldsymbol{y}^{1'} \quad \boldsymbol{i}_r\boldsymbol{y}^2 \quad \boldsymbol{0}_{r-1}'\,]'$. Then the null and alternative hypotheses may be stated as

$$\text{Test \#4} \quad \begin{cases} H_0: & 1 \le D^{in}(\boldsymbol{x}^a, \boldsymbol{y}) = D^{in}(\boldsymbol{x}, \boldsymbol{y})\ \forall\ (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}; \\ H_1: & 1 \le D^{in}(\boldsymbol{x}^a, \boldsymbol{y}) < D^{in}(\boldsymbol{x}, \boldsymbol{y}) \text{ for some } (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}. \end{cases}$$

Analogous to (4.1), we can compute

$$\left[\widehat{D}_0^{in}(\boldsymbol{x}, \boldsymbol{y}^+)\right]^{-1} = \max\left\{\theta > 0 \mid \theta\boldsymbol{y}^+ \le \boldsymbol{Y}^+\boldsymbol{q},\ \boldsymbol{x} \ge \boldsymbol{X}\boldsymbol{q},\ \boldsymbol{i}\boldsymbol{\tau} = 1,\ \boldsymbol{q} \in \mathbb{R}_+^n\right\}, \qquad (4.2)$$

where $\boldsymbol{Y}^+ = [\,\boldsymbol{y}_1^+ \quad \ldots \quad \boldsymbol{y}_n^+\,]$ and $\boldsymbol{y}^+ = [\,\boldsymbol{y}^{1'} \quad \boldsymbol{i}_r'\boldsymbol{y}^2\,]'$. Under the null in Test #4, $\widehat{D}_0^{in}(\boldsymbol{x}, \boldsymbol{y}^+)$ should be "close" to $\widehat{D}_0^{in}(\boldsymbol{x}, \boldsymbol{y})$, but in finite samples, $\widehat{D}_0^{in}(\boldsymbol{x}, \boldsymbol{y}^+) \le \widehat{D}_0^{in}(\boldsymbol{x}, \boldsymbol{y})$ due to sampling variation. Again, the question is whether the differences between $\widehat{D}_0^{in}(\boldsymbol{x}, \boldsymbol{y}^+)$ and $\widehat{D}_0^{in}(\boldsymbol{x}, \boldsymbol{y})$ are large enough to cast doubt on the null.

5. SOME TEST STATISTICS

As noted earlier, we defined a variety of test statistics for Tests #1–4 in sections 3 and 4. Among those that we considered, our Monte Carlo experiments found six that gave reasonable performance in terms of size. We discuss only those six here.

To simplify notation, let

$$\widehat{\rho}_{1i} = \widehat{D}^{out}(\boldsymbol{x}_i, \boldsymbol{y}_i)/\widehat{D}_0^{out}(\boldsymbol{x}_i^1, \boldsymbol{y}_i) \tag{5.1}$$

and

$$\widehat{\delta}_{1i} = \widehat{D}^{out}(\boldsymbol{x}_i, \boldsymbol{y}_i) - \widehat{D}_0^{out}(\boldsymbol{x}_i^1, \boldsymbol{y}_i). \tag{5.2}$$

Also, let $\widehat{\rho}_{1(j)}$ denotes the jth smallest value among the $\widehat{\rho}_{1i}$, $i = 1, \ldots, n$, so that $\widehat{\rho}_{1(j)}$ defines an order statistic; $\widehat{\delta}_{1(j)}$ may be defined similarly. Then for purposes of Test #1, where we test for irrelevant inputs, the statistics which performed best in our Monte Carlo experiments may be written as:

$$\widehat{\gamma}_{11}(\mathcal{S}_n) = \sum_{i=1}^{n} (\widehat{\rho}_{1i} - 1) \geq 0, \tag{5.3}$$

$$\widehat{\gamma}_{12}(\mathcal{S}_n) = \sum_{i=1}^{n} \left(\frac{\widehat{\delta}_{1i}}{\widehat{D}_0^{out}(\boldsymbol{x}_i, \boldsymbol{y}_i)} \right)^2 \geq 0, \tag{5.4}$$

$$\widehat{\gamma}_{13}(\mathcal{S}_n) = \sum_{i=1}^{n} \left(\frac{\widehat{\delta}_{1i}}{\widehat{D}_0^{out}(\boldsymbol{x}_i, \boldsymbol{y}_i)} \right) \geq 0, \tag{5.5}$$

$$\widehat{\gamma}_{14}(\mathcal{S}_n) = \text{median of } \{\widehat{\rho}_{1i}\}_{i=1}^{n} \geq 0, \tag{5.6}$$

$$\widehat{\gamma}_{15}(\mathcal{S}_n) = \text{median of } \left\{\widehat{\delta}_{1i}\right\}_{i=1}^{n} - 1 \geq 0, \tag{5.7}$$

and

$$\widehat{\gamma}_{16}(\mathcal{S}_n) = \left(\widehat{\rho}_{(1)} - 1\right) + \sum_{j=2}^{n} \left[\left(\widehat{\rho}_{(j)} - \widehat{\rho}_{(j-1)}\right) \frac{(n-j+1)}{n} \right] \geq 0. \tag{5.8}$$

Recall that for Test #1, $\widehat{\rho}_{1i} \geq 1$ and $\widehat{\delta}_{1i} \geq 0$; if the null hypothesis is true, then in the limit, $\widehat{\rho}_{1i} = 1$ and $\widehat{\delta}_{1i} = 0$. The first five statistics in (5.3)–(5.7) seem rather obvious at first glance. The last statistic in (5.8) is based on the idea of Kolmogorov-Smirnov statistics; it is merely the integrated difference between (i) the limiting distribution function of the $\widehat{\rho}_{1i}$ under the null and (ii) the empirical distribution function of the $\widehat{\rho}_{1i}$ for $i = 1, \ldots, n$. In the limit, if the null hypothesis is true, the $\widehat{\rho}_{1i}$ must all equal unity, and the difference between (i) and (ii) must equal 0. Implementations of Test #1 may be based on any one of these statistics, and the null hypothesis is to be rejected if the computed value of the chosen statistic (a function of $\mathcal{S}_n$) exceeds an appropriate critical value.

Before discussing how critical values may be obtained, we define similar statistics for the remaining tests. For Test #2, let

$$\widehat{\rho}_{2i} = \widehat{D}_0^{in}(\boldsymbol{x}_i, \boldsymbol{y}_i^1)/\widehat{D}^{in}(\boldsymbol{x}_i, \boldsymbol{y}_i^1, \boldsymbol{y}_i^2) \geq 1 \tag{5.9}$$

and

$$\widehat{\delta}_{2i} = \widehat{D}_0^{in}(\boldsymbol{x}_i, \boldsymbol{y}_i^1) - \widehat{D}^{in}(\boldsymbol{x}_i, \boldsymbol{y}_i^1, \boldsymbol{y}_i^2) \geq 0. \tag{5.10}$$

Then statistics $\widehat{\gamma}_{21}(\mathcal{S}_n), \ \ldots, \widehat{\gamma}_{26}(\mathcal{S}_n)$ may be defined by replacing $\widehat{\rho}_{1i}$ with $\widehat{\rho}_{2i}$ and $\widehat{\delta}_{1i}$ with $\widehat{\delta}_{2i}$ in (5.3)–(5.7), respectively. For Test #3, define

$$\widehat{\rho}_{3i} = \widehat{D}_0^{out}(\boldsymbol{x}_i^+, \boldsymbol{y}_i)/\widehat{D}^{out}(\boldsymbol{x}, \boldsymbol{y}_i) \tag{5.11}$$

and

$$\widehat{\delta}_{3i} = \widehat{D}_0^{out}(\boldsymbol{x}_i^+, \boldsymbol{y}_i) - \widehat{D}^{out}(\boldsymbol{x}, \boldsymbol{y}_i). \tag{5.12}$$

Then statistics $\widehat{\gamma}_{31}(\mathcal{S}_n), \ \ldots, \widehat{\gamma}_{36}(\mathcal{S}_n)$ may be defined by replacing $\widehat{\rho}_{1i}$ with $\widehat{\rho}_{3i}$ and $\widehat{\delta}_{1i}$ with $\widehat{\delta}_{3i}$ in (5.3)–(5.7), respectively. Finally, for Test #4, define

$$\widehat{\rho}_{4i} = \widehat{D}^{in}(\boldsymbol{x}_i, \boldsymbol{y}_i)/\widehat{D}_0^{in}(\boldsymbol{x}, \boldsymbol{y}_i^+) \tag{5.13}$$

and

$$\widehat{\delta}_{4i} = \widehat{D}^{in}(\boldsymbol{x}_i, \boldsymbol{y}_i) - \widehat{D}_0^{in}(\boldsymbol{x}, \boldsymbol{y}_i^+), \tag{5.14}$$

and then define statistics $\widehat{\gamma}_{41}(\mathcal{S}_n), \ \ldots, \widehat{\gamma}_{46}(\mathcal{S}_n)$ by replacing $\widehat{\rho}_{1i}$ with $\widehat{\rho}_{4i}$ and $\widehat{\delta}_{1i}$ with $\widehat{\delta}_{4i}$ in (5.3)–(5.7), respectively.

In each case we have defined test statistics such that their support is bounded on $[0, \infty)$.

6. IMPLEMENTING THE TESTS

Since analytic results establishing distributions for our various test statistics are not available, the bootstrap (Efron, 1979, 1982) seems to offer the only route to obtaining critical values for our tests. Suppose we are using a particular statistic $\widehat{\gamma}(\mathcal{S}_n)$ for one of Tests #1–4 (here, we suppress subscripts on $\widehat{\gamma}(\mathcal{S}_n)$, with the understanding that the story we tell may apply to any of the statistics defined in the previous section), and let $\widehat{\gamma}(\mathcal{S}_n)$ denote a realized value of the random variable $\widehat{\gamma}$. We would like to know the distribution of $(\widehat{\gamma}-\gamma)$ under the null hypothesis H_0, where γ is the value estimated by $\widehat{\gamma}(\mathcal{S}_n)$. If this distribution were known, it would be trivial to find a value c_α such that

$$\text{Prob}\left(\widehat{\gamma}-\gamma \leq c_\alpha \mid H_0, \mathcal{F}\right) = 1-\alpha \tag{6.1}$$

for some small value of α, *e.g.,* 0.05. Unfortunately, however, this distribution is not known, but it (fortunately) can be estimated by the bootstrap.

The bootstrap idea is based on drawing an iid sample $\mathcal{S}_n^* = \{(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*)\}_{i=1}^n$ from an estimate $\widehat{\mathcal{F}}$ of the DGP $\mathcal{F}$ defined by assumptions A1–A6 in sections 2.1–2.2. Given $\mathcal{S}_n^*$, it is straightforward to apply the original estimation methods to the data in $\mathcal{S}_n^*$ to obtain a bootstrap version $\widehat{\gamma}(\mathcal{S}_n^*)$ of our statistic. For example, if we are using $\widehat{\gamma}_{11}(\mathcal{S}_n)$ to test the null hypothesis in Test #1, we can compute, for each $i = 1, \ldots, n$,

$$\left[\widehat{D}^{out*}(\boldsymbol{x}_i, \boldsymbol{y}_i)\right]^{-1} = \max\left\{\theta > 0 \mid \theta \boldsymbol{y}_i \leq \boldsymbol{Y}^*\boldsymbol{q},\ \boldsymbol{x}_i \geq \boldsymbol{X}^*\boldsymbol{q},\ \boldsymbol{i}'\boldsymbol{q} = 1,\ \boldsymbol{q} \in \mathbb{R}_+^n\right\} \tag{6.2}$$

and

$$\left[\widehat{D}_0^{out*}(\boldsymbol{x}_i^1, \boldsymbol{y}_i)\right]^{-1} = \max\left\{\theta > 0 \mid \theta \boldsymbol{y}_i \leq \boldsymbol{Y}^*\boldsymbol{q},\ \boldsymbol{x}_i^1 \geq \boldsymbol{X}^{1^*}\boldsymbol{q},\ \boldsymbol{i}'\boldsymbol{q} = 1,\ \boldsymbol{q} \in \mathbb{R}_+^n\right\}, \tag{6.3}$$

which are analogous to (2.16) and (3.1), and where $\boldsymbol{Y}^* = [\,\boldsymbol{y}_1^* \;\; \ldots \;\; \boldsymbol{y}_n^*\,]$ and $\boldsymbol{X}^* = [\,\boldsymbol{x}_1^* \;\; \ldots \;\; \boldsymbol{x}_n^*\,]$. In (6.2), the distance function estimate measures distance from the observed sample point $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ to the boundary of the convex hull of the free-disposal hull of the points in the pseudo-sample $\mathcal{S}_n^*$. The distance function estimate in (6.3) is similar, but in the space $\mathbb{R}_+^{p-r+q}$ instead of $\mathbb{R}_+^{p+q}$ as in (6.2). Next, analogous to (5.1), we can compute

$$\widehat{\rho}_{1i}^* = \widehat{D}^{out*}(\boldsymbol{x}_i, \boldsymbol{y}_i)/\widehat{D}_0^{out*}(\boldsymbol{x}_i^1, \boldsymbol{y}_i), \tag{6.4}$$

and then we can replace $\widehat{\rho}_{1i}$ in (5.3) with $\widehat{\rho}^*_{1i}$ to obtain $\widehat{\gamma}_{11}(\mathcal{S}^*_n)$. For any of our other test statistics, the procedure would be similar to what we have described here.

The process described above can be repeated B times to yield B different pseudo samples $\mathcal{S}^*_{n,b}$, $b = 1, \ldots, B$, and hence a set of B bootstrap values $\left\{\widehat{\gamma}(\mathcal{S}^*_{n,b})\right\}_{b=1}^{B}$. These bootstrap values can then be used to determine the significance of the computed test statistic, or its p-value, as in standard bootstrap methods.[9]

The sole remaining question concerns how one might generate the pseudo datasets $\mathcal{S}^*_{n,b}$, which is a question of how the DGP $\mathcal{F}$ is to be estimated. These questions have been discussed extensively for the present setting in Simar and Wilson (1998a, 2000a, 2000b). In the general case, where there are no restrictions on $f(\boldsymbol{x}, \boldsymbol{y})$, one can use the method described in Simar and Wilson (2000a). However, in Simar and Wilson (1998a), we imposed a homogeneity restriction equivalent to specifying

$$f(\omega \mid \boldsymbol{x}, \boldsymbol{\eta}) = f(\omega) \tag{6.5}$$

in (2.12) or

$$f(\tau \mid \boldsymbol{\phi}, \boldsymbol{y}) = f(\tau) \tag{6.6}$$

in (2.14). In either case, smoothing is required due to the boundary condition at $\mathcal{P}^{\partial}$. The naive bootstrap based on resampling either from (i) the empirical distribution of the data in $\mathcal{S}_n$ or (ii) the empirical distribution of the distance function estimates (and then using (2.13) or (2.15) to construct pseudo observations for either outputs or inputs) has been shown to be inconsistent due to the boundary problem (Simar and Wilson, 1998a, 1999a, 1999b, 2000a, 2000b). The problem with the naive bootstrap in this setting is very similar to the problem resulting from distributions with bounded support in the univariate setting; *e.g.,* see Bickel and Freedman (1981), Beran and Ducharme (1991), and Efron and Tibshirani (1993).

Since we wish to conduct Monte Carlo experiments to evaluate the performance of

[9] See Efron and Tibshirani (1993, chapter 16) for details; there, the p-value is referred to as the "achieved significance level."

various test statistics, we adopt the homogeneity restrictions in (6.5)–(6.6) for both our simulated DGP and our estimation methods to reduce the computational burden.

In bootstrapping critical values for test statistics, the key is to generate the bootstrap pseudo-data $\mathbb{S}^*_{n,b}$ under the conditions of the null hypothesis for the particular test being considered. For the case of Test #1 with the homogeneity restriction in (6.5), this means maintaining the restriction that $\boldsymbol{x}^2$ is irrelevant in the production process, which is accomplished by drawing bootstrap values D^* from an estimate of the distribution of the n distance function estimates $\widehat{D}_0^{out}(\boldsymbol{x}_i^1, \boldsymbol{y}_i)$.

In Simar and Wilson (1998a), we show how how to draw from a kernel estimate of the density of distance function estimates for cases where the homogeneity restriction in (6.5) is imposed; one must use the reflection method as in Simar and Wilson (1998a) or some other device to avoid bias problems near the boundary of the distance function estimates when using kernel density estimators. For cases where the homogeneity restriction in (6.5) is relaxed, one can use the multivariate framework described in Simar and Wilson (2000a).

It is straightforward to adapt the methods described here to test the null hypotheses in Tests #2–3. For Test #3, one must merely change notation in (6.3)–(6.4). For Tests #2 and #4, where statistics are defined in terms of input distance function estimators, a few additional, small changes in the above procedure must be made. For example, in Test #2, (6.2)–(6.3) must be replaced by

$$\left[\widehat{D}^{in*}(\boldsymbol{x}_i, \boldsymbol{y}_i)\right]^{-1} = \min\left\{\theta > 0 \mid \boldsymbol{Y}^*\boldsymbol{q} \geq \boldsymbol{y}_i,\ \boldsymbol{X}^*\boldsymbol{q} \leq \theta\boldsymbol{x}_i,\ \boldsymbol{i}'\boldsymbol{q} = 1,\ \boldsymbol{q} \in \mathbb{R}^n_+\right\} \tag{6.7}$$

and

$$\left[\widehat{D}_0^{in*}(\boldsymbol{x}_i, \boldsymbol{y}_i^1)\right]^{-1} = \min\left\{\theta > 0 \mid \boldsymbol{y}_i^1 \leq \boldsymbol{Y}^{1*}\boldsymbol{q},\ \boldsymbol{X}^*\boldsymbol{q} \leq \theta\boldsymbol{x}_i,\ \boldsymbol{i}\boldsymbol{\tau} = 1,\ \boldsymbol{q} \in \mathbb{R}^n_+\right\}, \tag{6.8}$$

respectively.

7. MONTE CARLO EXPERIMENTS

As noted earlier, the test statistics defined in section 5 are a subset of a larger number of statistics we investigated via Monte Carlo experiments. Here, we describe those experiments, and discuss the performance of the test statistics listed in section 5 for each of the Tests #1–4 in terms of actual versus nominal sizes of the tests.

In each of our experiments, we simulated a DGP consistent with the null hypothesis in the particular test being considered. Let x_{ji}, y_{ji} denote the jth elements of $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$. Let u_{1i}, u_{2i}, and v_i be independent random variables, with u_{1i} and u_{2i} each distributed uniformly on (0,1), and $v_i \sim N(0,1)$.[10] Then for Test #1, we set $p=2$, $q=1$, and our simulated DGP amounts to setting

$$x_{1i} = 1 + 8u_{1i}, \tag{7.1}$$

$$x_{2i} = 1 + 8u_{2i}, \tag{7.2}$$

and

$$y_{1i} = x_{1i}^{2/3} e^{|-v_i|} \tag{7.3}$$

for each $i = 1, \ldots, n$. Clearly, x_{2i} does not play a role in determining y_i, as required by the null hypothesis in Test #1. Our DGP for examining Test #2 has $p=1$, $q=2$ and is described by (7.1), (7.3), and

$$y_{2i} = 0.5 + 3.827u_{2i}. \tag{7.4}$$

The range of irrelevant output, y_{2i}, was chosen to roughly match the range of y_{1i}, the relevant output.

In our experiments with Test #3, we set $p=2$, $q=1$ and simulated the inputs as in the case of Test #1, *i.e.*, according to (7.1)–(7.2). The output variable was then determined by

$$y_{1i} = (x_{1i} + x_{2i})^{2/3} e^{-|v_i|} \tag{7.5}$$

so that the inputs are additive, as required by the null hypothesis in Test #3. Similarly, in our experiments with Test #4, we have $p=1$, $q=2$, with the input determined as in (7.1). We then computed the outputs as

$$y_{1i} = (0.1 + 0.8u_{2i}) x_{1i}^{2/3} e^{-|v_i|} \tag{7.6}$$

[10]Uniform pseudo-random deviates were generated using the multiplicative congruential generator with modulus $(2^{31}-1)$ and multiplier 16807 (see Lewis *et al.,* 1969). Pseudo random normal deviates were generated via the Box-Muller method (see Press *et al.,* 1986, pp. 202–203).

and

$$y_{2i} = x_{1i}^{2/3} e^{-|v_i|} - y_{1i} \tag{7.7}$$

so that the outputs are additive as required by the null hypothesis in Test #4.

We considered sample sizes of n = 25, 50, 100, 200, and 400. We used 1000 trials in each of our Monte Carlo experiments, with $B = 2000$ bootstrap replications on each trial. On each trial, we first simulated the data according to one of the protocols described above, then estimated two distance functions for each observation in order to compute the various test statistics. Least-squares cross validation was then used to determine an optimal bandwidth for the kernel density estimation, and then 2000 pseudo datasets were drawn, with $2n$ distance function estimates computed for each pseudo dataset. The computational burden of these experiments was therefore severe, limiting the number of cases we were able to consider.[11]

Our simulation results are reported in Tables 1–4, corresponding to Tests #1–4. For each sample size, and for various nominal sizes, we report the estimated sizes of the tests using each statistic. The estimated sizes were determined by counting how many times over 1000 Monte Carlo trials the null hypothesis for a particular test, with a particular statistic, was rejected, and then dividing this count by 1000.

In general, the estimated sizes of the tests, for each statistic, improve as the sample size is increased. The estimated sizes are frequently larger than the nominal sizes, although not in every case. In particular, for small sizes (α = .05 or .01), the estimated sizes are more often too small than is the case for larger sizes.

Nonetheless, the results are encouraging. With $n = 400$, the estimated sizes are in many cases quite close to the corresponding nominal sizes. Performances appear slightly worse with Tests #2 and #4, but this perhaps merely a consequence of our choices simulated

[11] For example, in each experiment with n = 400, 4,002,000 linear programs (with 401 variables in each one) were solved. On a Cray T3E massively parallel "supercomputer," using 32 Digital Equipment Corporation Alpha EV4 RISC microprocessors, each experiment with $n = 400$ required roughly 36 hours CPU time *on each processor.* The computational burden is largely a consequence of the fact that we are in a Monte Carlo environment. Necessarily, for an applied researcher with the same sample size and the same number of inputs and outputs, the computational burden would be reduced by three orders of magnitude, and so should not be prohibitive.

DGPs for those experiments.

Of course, in applied settings, one would likely face a situation where $p + q > 3$, and increased dimensionality should worsen the performance of the tests. This is only a consequence of the well-known curse of dimensionality which plagues most, if not all, nonparametric estimators. The degree to which distance function estimators suffer from the curse of dimensionality is reflected by the convergence rate in (2.18). The solution, of course, is to avoid small samples when estimating in high dimensions. The fact that our tests might not work well with small samples in high dimensions is hardly surprising or troubling, since the estimates of the distance functions themselves are likely to be mostly meaningless in such a situation.

It is also possible to make some corrections to bootstrapped test statistics in applied problems using the iterated bootstrap discussed by Beran and Ducharme (1991), Hall (1992), and Simar and Wilson (1999c). This procedure involves a second bootstrap loop inside the first bootstrap loop represented by step [7] in Algorithm #1 above, and extending the analogy principle underlying all bootstrap methods. We have not examined this technique in a Monte Carlo setting because it is computationally prohibitive to do so. In an applied setting, however, computational costs should not be too severe.

We also examined the power of our tests for the case of Test #1. In these experiments, we modified the simulated null DGP slightly by setting the first input as in (7.1), but setting the second input as

$$x_{2i} = \omega(1 + 8u_{2i}) + (1 - \omega)x_{1i} \tag{7.8}$$

with $\omega \in \{0.2,\ 0.6,\ 1.0\}$ to induce correlation between x_{1i} and x_{2i}; for $\omega = 1$, (7.8) reduces to (7.2). The output variable was then simulated by setting

$$y_i = x_{1i}^{2/3} x_{2i}^{\kappa} e^{|v_i|}, \tag{7.9}$$

with $\kappa \in \{0,\ .01,\ .02,\ .04,\ .08,\ .12,\ .20\}$. In each experiment, we performed 1000 Monte Carlo trials as before, with $B = 2000$ bootstrap replications. For each experiment, we recorded the number of trials where the null hypothesis was rejected, and then divided this total by 1000 to obtain the results in Tables 5–7. In the experiments where $\kappa = 0$, the null hypothesis is maintained in the simulated DGP; these experiments thus allow examination of the size obtained with each test statistic for each value of ω. As κ increases from zero, we have increasing departure from the null hypothesis in Test #1, and therefore should see more rejections of the null.

Table 5 confirms this observation for the case of $\omega = 1,\ n = 200$.[12] Tests using the statistic $\widehat{\gamma}_{12}(\mathcal{S}_n)$ have low power, even for substantial departure from the null ($\kappa = .20$), but the other statistics perform well at nominal sizes of .1 and .05. The number of rejections of the null increases most quickly when $\widehat{\gamma}_{14}(\mathcal{S}_n)$ (median of ratios) and $\widehat{\gamma}_{15}(\mathcal{S}_n)$ (median of differences) are used. For $\kappa = .20$, these statistics yield power of 98.8 percent at a nominal size of .01, but the other statistics give somewhat less power (only 26.6 percent) for $\kappa = .20$ and a nominal size of .01.

Smaller values of ω result in relatively more correlation between the relevant input (x_{1i}) and the irrelevant input (x_{2i}), which should be expected to reduce the power of the tests to discriminate between the null and alternative hypotheses. Table 6 gives results for $\omega = 0.6$, and Table 7 give results for $\omega = 0.2$. Relative to the results for $\omega = 1$ in Table 5, power for corresponding values of κ and corresponding nominal sizes when $\omega = 0.6$ in Table 6 is only slightly lower. Lowering ω to 0.2 as in Table 7 further reduces the power provided by each statistic, as expected. Nonetheless, $\widehat{\gamma}_{14}(\mathcal{S}_n)$ and $\widehat{\gamma}_{15}(\mathcal{S}_n)$ still dominate the other statistics in terms of power performance.

8. Conclusions

We have provided two sets of tests for use with nonparametric frontier or efficiency estimators. Our tests are akin to t-tests and F-tests that are used routinely with the classical linear regression model in the sense that our tests allow testing of the same types of hypotheses as those tests in the linear model. The types of hypotheses we can test

[12] All of our experiments to examine power were conducted with $n = 200$ simulated observations to limit the computational burden.

are frequently of intrinsic interest to economists or management specialists, and are also interesting from an estimation viewpoint. If one is willing to accept the null hypothesis in any of the tests we have proposed after failure to reject, then it will be possible to mitigate, at least to some extent, the curse of dimensionality for a given sample size either by aggregating inputs or outputs, or by deleting irrelevant inputs or outputs from the estimation procedure.

REFERENCES

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

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

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

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

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

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

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

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

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

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

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 (1995), 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.

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.

Markovitz, H. M. (1959), *Portfolio Selection: Efficient Diversification of Investments.* John Wiley and Sons, Inc., New York.

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

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

Sengupta, J. K. (1991), Maximum probability dominance and portfolio theory, *Journal of Optimization Theory and Applications* 71, 341–357.

Sengupta, J. K., and H. S. Park (1993), Portfolio efficiency tests based on stochastic dominance and cointegration, *International Journal of Systems Science* 24, 2135–2158.

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

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

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, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Simar, L., and P.W. Wilson (1999a), Some problems with the Ferrier/Hirschberg bootstrap idea, *Journal of Productivity Analysis* 11, 67–80.

Simar, L., and P.W. Wilson (1999b), Of course we can bootstrap DEA scores! But does it mean anything? Logic trumps wishful thinking, *Journal of Productivity Analysis* 11, 93–97.

Simar, L., and P.W. Wilson (1999c), "Performance of the Bootstrap for DEA Estimators and Iterating the Principle," Discussion paper #0002, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Simar, L., and P.W. Wilson (2000a), A general methodology for bootstrapping in non-parametric frontier models, *Journal of Applied Statistics,* forthcoming.

Simar, L., and P.W. Wilson (2000b), Statistical inference in nonparametric frontier models: The state of the art, *Journal of Productivity Analysis* 13, 49–78.

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

Wheelock, D. C. and P. W. Wilson (2000), Why do banks disappear? The determinants of US bank failures and acquisitions, *Review of Economics and Statistics* 82, 127–138.

TABLE 1

Monte Carlo Estimates of Size for Test #1

n	Nominal Size	$\widehat{\gamma}_{11}(\mathcal{S}_n)$	$\widehat{\gamma}_{12}(\mathcal{S}_n)$	$\widehat{\gamma}_{13}(\mathcal{S}_n)$	$\widehat{\gamma}_{14}(\mathcal{S}_n)$	$\widehat{\gamma}_{15}(\mathcal{S}_n)$	$\widehat{\gamma}_{16}(\mathcal{S}_n)$
25	0.20	0.333	0.302	0.333	0.265	0.259	0.333
	0.15	0.249	0.217	0.249	0.214	0.206	0.249
	0.10	0.151	0.123	0.151	0.163	0.145	0.151
	0.05	0.052	0.044	0.052	0.088	0.077	0.052
	0.01	0.002	0.002	0.002	0.023	0.019	0.002
50	0.20	0.325	0.273	0.325	0.267	0.259	0.325
	0.15	0.224	0.169	0.224	0.217	0.209	0.224
	0.10	0.127	0.099	0.127	0.164	0.154	0.127
	0.05	0.048	0.040	0.048	0.099	0.092	0.048
	0.01	0.003	0.002	0.003	0.028	0.028	0.003
100	0.20	0.260	0.212	0.260	0.257	0.252	0.260
	0.15	0.188	0.146	0.188	0.216	0.205	0.188
	0.10	0.112	0.089	0.112	0.158	0.156	0.112
	0.05	0.045	0.021	0.045	0.109	0.101	0.045
	0.01	0.002	0.000	0.002	0.036	0.031	0.002
200	0.20	0.262	0.219	0.262	0.235	0.231	0.262
	0.15	0.202	0.170	0.202	0.196	0.193	0.202
	0.10	0.129	0.105	0.129	0.145	0.141	0.129
	0.05	0.060	0.040	0.060	0.081	0.083	0.060
	0.01	0.004	0.003	0.004	0.029	0.029	0.004
400	0.20	0.243	0.205	0.243	0.261	0.261	0.243
	0.15	0.182	0.157	0.182	0.216	0.211	0.182
	0.10	0.109	0.091	0.109	0.155	0.154	0.109
	0.05	0.048	0.025	0.048	0.083	0.081	0.048
	0.01	0.002	0.000	0.002	0.031	0.029	0.002

TABLE 2

Monte Carlo Estimates of Size for Test #2

n	Nominal Size	$\widehat{\gamma}_{21}(\mathcal{S}_n)$	$\widehat{\gamma}_{22}(\mathcal{S}_n)$	$\widehat{\gamma}_{23}(\mathcal{S}_n)$	$\widehat{\gamma}_{24}(\mathcal{S}_n)$	$\widehat{\gamma}_{25}(\mathcal{S}_n)$	$\widehat{\gamma}_{26}(\mathcal{S}_n)$
25	0.20	0.366	0.365	0.378	0.255	0.257	0.366
	0.15	0.281	0.302	0.304	0.210	0.214	0.281
	0.10	0.211	0.217	0.221	0.167	0.176	0.211
	0.05	0.120	0.133	0.152	0.119	0.134	0.120
	0.01	0.050	0.050	0.055	0.038	0.054	0.050
50	0.20	0.295	0.302	0.336	0.249	0.261	0.295
	0.15	0.247	0.251	0.273	0.203	0.208	0.247
	0.10	0.173	0.195	0.204	0.150	0.154	0.173
	0.05	0.109	0.115	0.127	0.091	0.093	0.109
	0.01	0.032	0.033	0.041	0.028	0.032	0.032
100	0.20	0.282	0.297	0.349	0.267	0.271	0.282
	0.15	0.232	0.243	0.292	0.213	0.226	0.232
	0.10	0.161	0.176	0.223	0.156	0.162	0.161
	0.05	0.101	0.116	0.145	0.094	0.097	0.101
	0.01	0.038	0.043	0.044	0.023	0.032	0.038
200	0.20	0.251	0.265	0.349	0.270	0.262	0.251
	0.15	0.211	0.223	0.288	0.227	0.224	0.211
	0.10	0.146	0.170	0.227	0.170	0.171	0.146
	0.05	0.091	0.117	0.152	0.093	0.094	0.091
	0.01	0.025	0.041	0.065	0.034	0.033	0.025
400	0.20	0.258	0.255	0.370	0.267	0.261	0.258
	0.15	0.207	0.210	0.307	0.211	0.204	0.207
	0.10	0.152	0.160	0.247	0.149	0.144	0.152
	0.05	0.104	0.102	0.174	0.090	0.087	0.104
	0.01	0.030	0.040	0.072	0.033	0.033	0.030

TABLE 3

Monte Carlo Estimates of Size for Test #3

n	Nominal Size	$\widehat{\gamma}_{31}(\mathcal{S}_n)$	$\widehat{\gamma}_{32}(\mathcal{S}_n)$	$\widehat{\gamma}_{33}(\mathcal{S}_n)$	$\widehat{\gamma}_{34}(\mathcal{S}_n)$	$\widehat{\gamma}_{35}(\mathcal{S}_n)$	$\widehat{\gamma}_{36}(\mathcal{S}_n)$
25	0.20	0.370	0.301	0.370	0.495	0.419	0.370
	0.15	0.269	0.208	0.269	0.404	0.348	0.269
	0.10	0.154	0.110	0.154	0.306	0.253	0.154
	0.05	0.055	0.033	0.055	0.187	0.143	0.055
	0.01	0.004	0.002	0.004	0.050	0.022	0.004
50	0.20	0.316	0.245	0.316	0.444	0.397	0.316
	0.15	0.238	0.172	0.238	0.367	0.312	0.238
	0.10	0.140	0.099	0.140	0.280	0.222	0.140
	0.05	0.052	0.034	0.052	0.164	0.129	0.052
	0.01	0.000	0.000	0.000	0.038	0.028	0.000
100	0.20	0.253	0.199	0.253	0.395	0.363	0.253
	0.15	0.190	0.143	0.190	0.319	0.287	0.190
	0.10	0.120	0.078	0.120	0.235	0.212	0.120
	0.05	0.047	0.028	0.047	0.136	0.110	0.047
	0.01	0.002	0.000	0.002	0.036	0.022	0.002
200	0.20	0.277	0.214	0.277	0.383	0.364	0.277
	0.15	0.212	0.155	0.212	0.322	0.294	0.212
	0.10	0.126	0.098	0.126	0.239	0.223	0.126
	0.05	0.056	0.043	0.056	0.144	0.130	0.056
	0.01	0.004	0.002	0.004	0.038	0.028	0.004
400	0.20	0.254	0.207	0.254	0.355	0.346	0.254
	0.15	0.184	0.153	0.184	0.291	0.263	0.184
	0.10	0.118	0.087	0.118	0.187	0.180	0.118
	0.05	0.049	0.026	0.049	0.101	0.091	0.049
	0.01	0.002	0.001	0.002	0.028	0.026	0.002

TABLE 4

Monte Carlo Estimates of Size for Test #4

n	Nominal Size	$\widehat{\gamma}_{41}(\mathcal{S}_n)$	$\widehat{\gamma}_{42}(\mathcal{S}_n)$	$\widehat{\gamma}_{43}(\mathcal{S}_n)$	$\widehat{\gamma}_{44}(\mathcal{S}_n)$	$\widehat{\gamma}_{45}(\mathcal{S}_n)$	$\widehat{\gamma}_{46}(\mathcal{S}_n)$
25	0.20	0.159	0.154	0.295	0.428	0.522	0.159
	0.15	0.094	0.099	0.227	0.338	0.454	0.094
	0.10	0.059	0.065	0.144	0.256	0.369	0.059
	0.05	0.027	0.032	0.071	0.140	0.257	0.027
	0.01	0.004	0.005	0.018	0.038	0.082	0.004
50	0.20	0.109	0.108	0.268	0.413	0.496	0.109
	0.15	0.064	0.071	0.200	0.347	0.435	0.064
	0.10	0.034	0.043	0.129	0.258	0.343	0.034
	0.05	0.014	0.017	0.060	0.131	0.202	0.014
	0.01	0.002	0.004	0.013	0.041	0.071	0.002
100	0.20	0.117	0.102	0.266	0.435	0.487	0.117
	0.15	0.078	0.067	0.203	0.370	0.421	0.078
	0.10	0.041	0.038	0.160	0.289	0.328	0.041
	0.05	0.015	0.012	0.085	0.174	0.206	0.015
	0.01	0.000	0.003	0.014	0.055	0.064	0.000
200	0.20	0.135	0.092	0.317	0.472	0.480	0.135
	0.15	0.093	0.069	0.270	0.399	0.403	0.093
	0.10	0.064	0.051	0.196	0.303	0.320	0.064
	0.05	0.028	0.025	0.101	0.193	0.209	0.028
	0.01	0.007	0.009	0.033	0.065	0.070	0.007
400	0.20	0.243	0.205	0.243	0.261	0.261	0.243
	0.15	0.182	0.157	0.182	0.216	0.211	0.182
	0.10	0.109	0.091	0.109	0.155	0.154	0.109
	0.05	0.048	0.025	0.048	0.083	0.081	0.048
	0.01	0.002	0.000	0.002	0.031	0.029	0.002

TABLE 5

Monte Carlo Estimates of Power for Test #1: $\omega = 1.0$

Statistic	α	$\kappa = .00$	$\kappa = .01$	$\kappa = .02$	$\kappa = .04$	$\kappa = .08$	$\kappa = .12$	$\kappa = .20$
$\widehat{\gamma}_{11}(\mathcal{S}_n)$	.1	.129	.153	.188	.277	.558	.847	.995
	.05	.060	.072	.085	.130	.295	.563	.945
	.01	.004	.004	.004	.005	.022	.064	.266
$\widehat{\gamma}_{12}(\mathcal{S}_n)$	.1	.105	.108	.111	.120	.144	.174	.235
	.05	.040	.042	.045	.045	.052	.063	.092
	.01	.003	.003	.001	.001	.002	.002	.003
$\widehat{\gamma}_{13}(\mathcal{S}_n)$	.1	.129	.153	.188	.277	.858	.847	.995
	.05	.060	.072	.085	.130	.295	.563	.945
	.01	.004	.004	.004	.005	.022	.064	.266
$\widehat{\gamma}_{14}(\mathcal{S}_n)$	.1	.145	.246	.400	.662	.912	.973	.997
	.05	.081	.166	.279	.545	.853	.962	.997
	.01	.029	.045	.102	.300	.705	.895	.988
$\widehat{\gamma}_{15}(\mathcal{S}_n)$	.1	.141	.238	.394	.653	.911	.973	.999
	.05	.083	.163	.275	.530	.844	.960	.997
	.01	.029	.045	.085	.285	.687	.883	.988
$\widehat{\gamma}_{16}(\mathcal{S}_n)$	.1	.129	.153	.188	.277	.558	.847	.995
	.05	.060	.072	.085	.130	.295	.563	.945
	.01	.004	.004	.004	.005	.022	.064	.266

TABLE 6

Monte Carlo Estimates of Power for Test #1: $\omega = 0.6$

Statistic	α	$\kappa = .00$	$\kappa = .01$	$\kappa = .02$	$\kappa = .04$	$\kappa = .08$	$\kappa = .12$	$\kappa = .20$
$\widehat{\gamma}_{11}(\mathcal{S}_n)$	.1	.128	.141	.168	.226	.400	.639	.929
	.05	.062	.071	.077	.113	.219	.372	.781
	.01	.004	.005	.007	.011	.031	.057	.197
$\widehat{\gamma}_{12}(\mathcal{S}_n)$	.1	.093	.094	.095	.099	.113	.130	.198
	.05	.043	.044	.047	.053	.057	.067	.093
	.01	.004	.004	.005	.004	.006	.007	.016
$\widehat{\gamma}_{13}(\mathcal{S}_n)$	.1	.128	.141	.168	.326	.400	.639	.929
	.05	.062	.071	.077	.113	.219	.372	.781
	.01	.004	.005	.007	.011	.031	.057	.197
$\widehat{\gamma}_{14}(\mathcal{S}_n)$	.1	.148	.202	.267	.438	.734	.874	.977
	.05	.081	.128	.175	.324	.610	.808	.964
	.01	.030	.040	.052	.122	.364	.620	.895
$\widehat{\gamma}_{15}(\mathcal{S}_n)$	.1	.143	.194	.262	.434	.718	.866	.977
	.05	.084	.123	.169	.315	.617	.798	.965
	.01	.031	.039	.048	.114	.349	.613	.888
$\widehat{\gamma}_{16}(\mathcal{S}_n)$	.1	.128	.141	.168	.226	.400	.639	.929
	.05	.062	.071	.077	.113	.219	.372	.781
	.01	.004	.005	.007	.011	.031	.057	.197

TABLE 7

Monte Carlo Estimates of Power for Test #1: $\omega = 0.2$

Statistic	α	$\kappa = .00$	$\kappa = .01$	$\kappa = .02$	$\kappa = .04$	$\kappa = .08$	$\kappa = .12$	$\kappa = .20$
$\widehat{\gamma}_{11}(\mathcal{S}_n)$	.1	.118	.128	.137	.150	.202	.275	.434
	.05	.056	.058	.064	.072	.094	.136	.256
	.01	.008	.009	.011	.012	.015	.019	.037
$\widehat{\gamma}_{12}(\mathcal{S}_n)$	.1	.090	.091	.093	.095	.101	.106	.127
	.05	.043	.043	.044	.046	.049	.052	.064
	.01	.006	.007	.007	.007	.008	.008	.008
$\widehat{\gamma}_{13}(\mathcal{S}_n)$	.1	.118	.128	.137	.150	.202	.275	.434
	.05	.056	.058	.064	.072	.094	.136	.256
	.01	.008	.009	.011	.012	.015	.019	.037
$\widehat{\gamma}_{14}(\mathcal{S}_n)$	.1	.151	.172	.185	.222	.343	.459	.692
	.05	.086	.099	.115	.145	.236	.344	.576
	.01	.033	.034	.036	.047	.077	.141	.330
$\widehat{\gamma}_{15}(\mathcal{S}_n)$	.1	.150	.170	.180	.224	.332	.453	.684
	.05	.083	.094	.111	.114	.233	.330	.564
	.01	.030	.031	.036	.048	.072	.139	.312
$\widehat{\gamma}_{16}(\mathcal{S}_n)$	.1	.118	.128	.137	.150	.202	.275	.434
	.05	.056	.058	.064	.072	.094	.136	.256
	.01	.008	.009	.011	.012	.015	.019	.037