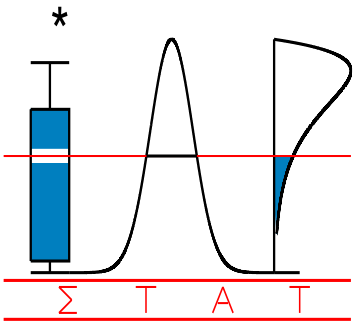
T E C H N I C A L

R E P O R T

0310

ESTIMATION AND INFERENCE IN TWO-STAGE, SEMI-PARAMETRIC MODELS OF PRODUCTION PROCESSES

L. SIMAR and P. W. WILSON



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes

Léopold Simar

and

Paul W. Wilson*

January 2003

Key Words: Data envelopment analysis, bootstrap, nonparametric efficiency estimation, two-stage estimation.

ABSTRACT

A plethora of papers have used multi-stage estimation procedures where nonparametric estimates of productive efficiency are obtained in the first stage and then regressed on environmental variables in a subsequent stage in attempts to account for exogenous factors that might affect firms' performance. None of these papers have described a coherent data-generating process (DGP). Moreover, conventional approaches to inference employed in these papers are invalid due to complicated, unknown serial correlation among the estimated efficiencies. We first describe a DGP wherein firms' efficiencies are influenced by environmental variables. We then propose a single and a double bootstrap procedure; both permit valid inference, and the double bootstrap procedure improves statistical efficiency in the second-stage regression. We examine the statistical performance of our estimators using Monte Carlo experiments.

*Simar: Institut de Statistique, Université Catholique de Louvain, Voie du Roman Pays 20, Louvain-la-Neuve, Belgium; email simar@stat.ucl.ac.be. Wilson: Department of Economics, University of Texas, Austin, Texas 78712 USA; email wilson@eco.utexas.edu. We are grateful to Irène Gijbels and participants at the $XXXIV$èmes Journées de Statistique, Société Française de Statistique, Bruxelles; the North American Productivity Workshop, Albany, New York; and the Workshop on Quantitative Methods for the Measurement of Organizational Efficiency, Institute for Fiscal Studies, University College, London, for comments on earlier versions. Research support from "Projet d'Actions de Recherche Concertées" (No. 98/03–217), from the "Interuniversity Attraction Pole", Phase V (No. P5/24) from the Belgian Government, and from the Texas Advanced Computation Center is gratefully acknowledged. Any remaining errors are are solely our responsibility.

1. INTRODUCTION

Linear-programming based measures of efficiency along the lines of Charnes *et al.* (1978, 1979) and Färe *et al.* (1985) are widely used in the analysis of efficiency of production. 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 data-generating process (DGP). 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. Many of these studies have used a two-stage approach, where efficiency is estimated in the first stage, and then the estimated efficiencies (or, in a few cases, ratios of estimated efficiencies, Malmquist indices, etc.) are regressed on covariates (typically different from those used in the first stage) that are viewed as representing environmental variables. Published examples of this approach include Byrnes *et al.* (1988), Ray (1988, 1991), McCarty and Yaisawarng (1993), Rhodes and Southwick (1993), Banker and Johnston (1994), Chirikos and Sear (1994), Kooreman (1994), Lovell *et al.* (1994), Sexton *et al.* (1994), Chilingerian (1995), Rosko *et al.* (1995), Arnold *et al.* (1996), De Borger and Kerstens (1996), González and Barber (1996), Luoma *et al.* (1996), Carrington *et al.* (1997), Burgess and Wilson (1998), Kirjavainen and Loikkanen (1998), McMillan and Datta (1998), Puig-Junoy (1998), Dietch and Weill (1999), Fried, Lovell, and Yaisawarng (1999), Garden and Ralston (1999), Resende (2000), Worthington and Dollery (2000), Chakraborty *et al.* (2001), Mukherjee *et al.* (2001), Ralston *et al.* (2001), Isik and Hassan (2002), and Stanton (2002). In a slight variation on this approach, Fried *et al.* (1993), Fried, Schmidt, and Yaisawarng (1999), and Fried *et al.* (2002) regressed radial and non-radial slacks on environmental variables. Since their dependent variables are functions of estimated efficiencies, the problems discussed

below apply here as well. In addition, Internet search engines such as google.com reveal dozens, perhaps hundreds of unpublished working papers that use the two-stage approach.

As far as we have been able to determine, none of the studies that employ two-stage approaches have described the underlying data-generating process. Since the DGP has not been described, there is some doubt about what is being estimated in the two-stage approaches. Among the studies that regress DEA estimates of efficiency on some covariates in the second stage, most have specified a censored (tobit) model for the second stage, but several (*e.g.,* Chirkos and Sears, 1994; Dietsch and Weill, 1999; Ray, 1991; Sexton *et al.,* 1994; and Stanton, 2002) have estimated a linear model by ordinary least squares (OLS). Authors have argued that DEA efficiency estimates are somehow censored since there are typically a mass of estimates at one, but no coherent account of how the censoring arises has been offered. Others have used OLS in the second-stage regression after transforming the DEA estimates of efficiency using log, logistic, or log-normal transformations and in some cases adding or subtracting an arbitrary constant to avoid division by zero or taking the log of zero (*e.g.,* Byrnes *et al.,* 1988; Ray, 1999; and Puig-Junoy, 1998). Lovell *et al.* (1994) and Burgess and Wilson (1998) avoided boundary problems in their second-stage regressions by using in their first-stage estimation a leave-one-out estimator of efficiency originally suggested by Andersen and Petersen (1993). Unfortunately, however, it is difficult to give a statistical interpretation to this estimator, even if the second-stage regressions are ignored. Still others have regressed ratios of efficiency estimates, Malmquist indices, or differences in efficiency estimates in the second stage; these have avoided boundary problems, but have still not provided a coherent description of a DGP that would make such regressions sensible.

A more serious problem in ***all*** of the the two-stage studies that we have found arises from the fact that DEA efficiency estimates are serially correlated; consequently, standard approaches to inference—used in all but two of the studies we have seen that employ the two-stage approach—are invalid. The two exceptions are Xue and Harker (1999) and

Hirschberg and Lloyd (2002); they recognize that DEA efficiency estimates are serially correlated, but both papers use a naive bootstrap method based on resampling from an empirical distribution in their attempts to correct the serial correlation problem. Unfortunately, the the naive bootstrap used by both Xue and Harker (1999) and Hirschberg and Lloyd (2002) is inconsistent in the context of nonparametric efficiency estimation, as demonstrated by Simar and Wilson (1999a, 1999b), and so the approaches by Xue and Harker (1999) and Hirschberg and Lloyd (2002) make little sense Moreover, neither of these studies describe a DGP for which their second-stage regressions would be appropriate.

This paper describes a DGP that is logically consistent with regression of nonparametric efficiency estimates on some covariates in a second stage. In addition, we demonstrate that while conventional inference methods are inconsistent in the second-stage regression, consistent inference is both possible and feasible.

2. A STATISTICAL MODEL

Let $\boldsymbol{x} \in \mathbb{R}^p_+$ denote a $(1 \times p)$ vector of inputs; $\boldsymbol{y} \in \mathbb{R}^q_+$ denote a $(1 \times q)$ vector of outputs; and $\boldsymbol{z} \in \mathbb{R}^r$ denote a $(1 \times r)$ vector of environmental variables. Elements of $\boldsymbol{z}$ may be either continuous or discrete; we will elaborate further on this later in the paper. How one might decide whether a particular variable is an environmental variable rather than an input or output of the production process is not entirely clear; our aim is not to provide answers to such questions, but rather to rationalize the two-stage analysis that is performed once this decision has been made.[1]

Studies that have used the two-stage approach have assumed, either explicitly or implicitly, that firms face certain environmental variables $\boldsymbol{z}$, and that these constrain their choices of inputs $\boldsymbol{x}$ and outputs $\boldsymbol{y}$. In the real world, the analyst is confronted with a set of observations $\mathcal{S}_n = \{(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{z}_i)\}_{i=1}^n$.

Assumption A1: *the sample observations* $(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{z}_i)$ *in* $\mathcal{S}_n$ *are realizations of identically,*

[1]Note that one can test whether a particular variable is an input, or an output, using the methods described in Simar and Wilson (2001a).

independently distributed random variables with probability density function $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ *which has support over* $\mathcal{P} \times \mathbb{R}^r$*, where* $\mathcal{P} \subset \mathbb{R}_+^{p+q}$ *is a production set defined by*

$$\mathcal{P} = \{(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x} \text{ can produce } \boldsymbol{y}\}. \tag{2.1}$$

In any interesting case, $\boldsymbol{z}$ is not independent with respect to $(\boldsymbol{x}, \boldsymbol{y})$, *i.e.,* $f(\boldsymbol{x}, \boldsymbol{y} \mid \boldsymbol{z}) \neq f(\boldsymbol{x}, \boldsymbol{y})$; independence between $\boldsymbol{z}$ and $(\boldsymbol{x}, \boldsymbol{y})$ can be tested using the methods surveyed by Wilson (2003). Otherwise, there would be no motivation for the second-stage regression. Assumption A1 means that the constraints on firms' choices of inputs $\boldsymbol{x}$ and outputs $\boldsymbol{y}$ due to the environmental variables that firms face operates through the dependence of $(\boldsymbol{x}, \boldsymbol{y})$ on $\boldsymbol{z}$ in $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$.

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. 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 the boundary of $\mathcal{P}$.

Various measures of technical efficiency are possible. Define a measure δ for some point $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathbb{R}_+^{p+q}$ such that

$$\delta_0 = \delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{P}) \equiv \sup\{\delta \mid (\boldsymbol{x}_0, \delta \boldsymbol{y}_0) \in \mathcal{P},\ \delta > 0\}. \tag{2.2}$$

This is simply the Farrell (1957) measure of output technical efficiency, which is the reciprocal of the Shephard (1970) output distance function. For $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{P}$, $\delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{P}) \geq 1$. Note that δ provides a measure of Euclidean distance from the point $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathbb{R}_+^{p+q}$ to the boundary of $\mathcal{P}$ in a direction parallel to the output axes and orthogonal to the input axes. In the the next assumption, and frequently in the discussion that follows, we will replace the subscripts "0" in (2.2) with subscript i to signify that the distance measure is evaluated for a particular, specific observation in $\mathcal{S}_n$.

For reasons that will become clear later, it is convenient to represent $\boldsymbol{y}$ in terms of its polar coordinates, while expressing the modulus in terms of distance from the boundary of

$\mathcal{P}$. Again for an arbitrary point $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathbb{R}_+^{p+q}$, and $\boldsymbol{y}_0 = [\, y_{01} \quad \ldots \quad y_{0q} \,]$, we can write the angles as

$$\eta_{0j} = \begin{cases} \arctan\left(y_{0,j+1}/y_{01}\right) & \text{for } y_{01} > 0 \\ \pi/2 & \text{if } y_{01} = 0 \end{cases} \tag{2.3}$$

for $j = 1, \ldots, q-1$. The corresponding modulus is given by $\omega(\boldsymbol{y}_0) = \boldsymbol{y}_0'\boldsymbol{y}_0$, which is related to the Farrell efficiency measure by

$$\delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{P}) = \frac{\omega(\delta\left(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{P}\right)\boldsymbol{y}_0)}{\omega(\boldsymbol{y}_0)}. \tag{2.4}$$

Thus, since $\mathcal{P}$ is fixed, we can characterize $\boldsymbol{y}_0$ by $(\boldsymbol{\eta}_0, \delta_0)$, where $\boldsymbol{\eta}_0 = [\, \eta_{01} \quad \ldots \quad \eta_{0,q-1} \,]$ and $\delta_0 = \delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{P})$.

The joint density $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ can now be described by a series of conditional densities and in terms of cylindrical coordinates:

$$f(\boldsymbol{x}_i, \boldsymbol{\eta}_i, \delta_i, \boldsymbol{z}_i) = f(\boldsymbol{x}_i, \boldsymbol{\eta}_i \mid \delta_i, \boldsymbol{z}_i) f(\delta_i \mid \boldsymbol{z}_i) f(\boldsymbol{z}_i). \tag{2.5}$$

The order of the conditioning on the right-hand side of (2.5) reflects the sequential nature of the DGP. Firm i is faced with environmental variables $\boldsymbol{z}_i$ drawn from $f(\boldsymbol{z})$. Given this $\boldsymbol{z}_i$, an efficiency level δ_i is drawn from $f(\delta_i \mid \boldsymbol{z}_i)$, and then $\boldsymbol{x}_i$ and $\boldsymbol{\eta}_i$ are drawn from $f(\boldsymbol{x}, \boldsymbol{\eta} \mid \delta, \boldsymbol{z})$, resulting in a realization $(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{z}_i)$ from the joint density $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ after transforming the polar coordinates $(\boldsymbol{\eta}_i, \delta_i)$ to Cartesian coordinates $\boldsymbol{y}_i$.

Assumption A2: *the conditioning in $f(\delta_i \mid \boldsymbol{z}_i)$ in (2.5) operates through the following mechanism:*

$$\delta_i = \psi(\boldsymbol{z}_i, \boldsymbol{\beta}) + \varepsilon_i \geq 1, \tag{2.6}$$

where ψ is a smooth, continuous, function, $\boldsymbol{\beta}$ is a vector of (possibly infinitely many) parameters, and ε_i is a continuous iid random variable, independent of $\boldsymbol{z}_i$.

Note that Assumptions A1–A2 amount to a separability condition; the production set $\mathcal{P}$ is assumed a subset of the entire sample space, while the effect of the covariates $\boldsymbol{z}$ operates through the dependence between $\boldsymbol{y}$ and $\boldsymbol{z}$ induced by (2.6).

Assumption A3: *ε_i in (2.6) is distributed $N(0, \sigma_\varepsilon^2)$ with left-truncation at $1 - \psi(z_i, \boldsymbol{\beta})$ for each i.*

Assumption A3 could be changed to impose some other distribution on the ε_i, but normality seems a natural choice. Alternatively, one could leave the distribution of ε unspecified and employ the semi-parametric method for truncated regression proposed by Honoré and Powell (1994); one could in addition leave the function $\psi(\cdot)$ unspecified and employ the nonparametric method for truncated regression proposed by Lewbel and Linton (2002). Here, however, we impose a distributional form for ε and later will assume a form for $\psi(\cdot)$, since that is what has been done in every two-stage applied study that we are aware of.

The production set $\mathcal{P}$ is sometimes described in terms of its sections

$$\mathcal{Y}(\boldsymbol{x}) \equiv \{\boldsymbol{y} | (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}\} \tag{2.7}$$

and

$$\mathcal{X}(\boldsymbol{y}) \equiv \{\boldsymbol{x} | (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}\}, \tag{2.8}$$

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 A4: *$\mathcal{P}$ 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 A5: *$(\boldsymbol{x}, \boldsymbol{y}) \notin \mathcal{P}$ if $\boldsymbol{x} = 0$, $\boldsymbol{y} \geq 0, \boldsymbol{y} \neq \mathbf{0}$, i.e., all production requires use of some inputs.*

Assumption A6: *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.*

Here and 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 number $\ell \in \{0,\ 1,\ \ldots,\ p\}$ of the corresponding elements of $\widetilde{\boldsymbol{x}}$ and $\boldsymbol{x}$ ere equal, while $(p-\ell)$ of the elements of $\widetilde{\boldsymbol{x}}$ are greater than the corresponding elements of $\boldsymbol{x}$. Assumption A5 merely says that there are no free lunches. Assumption A6 is sometimes called ***free disposability*** and is equivalent to an assumption of monotonicity of the technology.

In order for our estimators of $\mathcal{P}$ and $\delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{P})$ to be consistent, some additional assumptions are needed. In particular, the probability of observing firms in a neighborhood of the boundary of $\mathcal{P}$ must approach unity as the sample size increases:

Assumption A7: *for all* $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}$ *such that* $(\theta^{-1}\boldsymbol{x}, \boldsymbol{y} \notin \mathcal{P})$ *and* $(\boldsymbol{x}, \theta\boldsymbol{y}) \notin \mathcal{P})$ *for* $\theta > 1$, $f(\boldsymbol{x}, \boldsymbol{y} \mid \boldsymbol{z})$ *is strictly positive, and* $f(\boldsymbol{x}, \boldsymbol{y} \mid \boldsymbol{z})$ *is continuous in any direction toward the interior of* $\mathcal{P}$ *for all* $\boldsymbol{z}$.

Also, an assumption about the smoothness of the frontier is needed:

Assumption A8: *for all* $(\boldsymbol{x}, \boldsymbol{y})$ *in the interior of* $\mathcal{P}$, $\delta(\boldsymbol{x}, \boldsymbol{y} \mid \mathcal{P})$ *is differentiable in both its arguments.*

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.

Assumptions A4–A6 are standard in microeconomic theory of the firm; assumptions A7–A8 are based on those of Kneip *et al.* (1998), with some extensions to accommodate the environmental variables in $\boldsymbol{z}$. Assumptions A1–A3 have been introduced specifically to accommodate the environmental variables. Together, assumptions A1–A8 define the DGP $\mathcal{F}$ which yields the data in $\mathcal{S}_n$. The problem is to estimate $\{\delta_i\}_{i=1}^n$ and $\boldsymbol{\beta}$, and then to make inferences about these unknown quantities. For the case of the δ_i, we have provided two approaches to inference, in Simar and Wilson (1998, 2000a). Our focus here is on estimation and inference about $\boldsymbol{\beta}$, which describes the marginal effects of $\boldsymbol{z}$ on inefficiency.

3. Typical Two-Stage Approaches

The convex hull of the free disposal hull of the observed pairs $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ contained in $\mathcal{S}_n$ has frequently been used to estimate the production set $\mathcal{P}$. This estimator is described by

$$\widehat{\mathcal{P}} = \left\{(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{y} \leq \boldsymbol{Y}\boldsymbol{q},\ \boldsymbol{x} \geq \boldsymbol{X}\boldsymbol{q},\ \boldsymbol{i}'\boldsymbol{q} = 1,\ \boldsymbol{q} \in \mathbb{R}^n_+\right\}, \tag{3.1}$$

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. Korostelev *et al.* (1995) proved that $\widehat{\mathcal{P}}$ is a consistent estimator of $\mathcal{P}$ under conditions met by assumptions A1–A8 above. Estimators of the Farrell efficiency measure can be constructed by replacing $\mathcal{P}$ on the right-hand side of (2.4) with $\widehat{\mathcal{P}}$.

The estimator of $\delta_0 = \delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \mathcal{P})$ defined in (2.2) at a particular point $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathbb{R}^{p+q}_+$ can be written in terms of the linear program

$$\begin{aligned} \widehat{\delta_0} &= \delta(\boldsymbol{x}_0, \boldsymbol{y}_0 \mid \widehat{\mathcal{P}}) \\ &= \max\left\{\theta > 0 \mid \theta\boldsymbol{y}_0 \leq \boldsymbol{Y}\boldsymbol{q},\ \boldsymbol{x}_0 \geq \boldsymbol{X}\boldsymbol{q},\ \boldsymbol{i}'\boldsymbol{q} = 1,\ \boldsymbol{q} \in \mathbb{R}^n_+\right\}, \end{aligned} \tag{3.2}$$

where the maximization provides a solution for θ as well as $\boldsymbol{q}$. This is merely the empirical analog of the measure defined in (2.2).

It is straightforward to prove that $\widehat{\delta_0}$ is a consistent estimator of δ_0 under assumptions A1–A8 by altering the notation in Kneip *et al.* (1998). In particular,

$$\widehat{\delta_0} = \delta_0 + \mathcal{O}_p(n^{-\frac{2}{p+q+1}}). \tag{3.3}$$

The rate of convergence is low, as is typical in nonparametric estimation, and the rate slows as $p + q$ is increased—this is the well-known curse of dimensionality. Moreover, by construction, $\widehat{\delta_0}$ is biased downward.

Few results exist on the sampling distributions of nonparametric efficiency estimators such as the one in (3.2). Gijbels *et al.* (1999) derived the asymptotic distribution of the Shephard (1970) output distance function 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 second-stage regression problem considered here, the bootstrap also appears to be useful.

In principal, one could assume $\boldsymbol{\beta}$ in (2.2) has finite dimensions, fully specify the density $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$, and then estimate $\boldsymbol{\beta}$ by maximum likelihood. More often, however, researchers have employed some variant of the approach outlined below. The two-stage studies that have appeared in the literature typically specify $\psi(\boldsymbol{z}_i, \boldsymbol{\beta}) = \boldsymbol{z}_i\boldsymbol{\beta}$ so that (2.6) can be written as

$$\delta_i = \boldsymbol{z}_i\boldsymbol{\beta} + \varepsilon_i \geq 1, \tag{3.4}$$

where $\delta_i = \delta(\boldsymbol{x}_i, \boldsymbol{y}_i \mid \widehat{\mathcal{P}})$ in the notation of (3.2). These studies then (i) use the observed pairs $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ in $\mathcal{S}_n$ to estimate δ_i for all $i = 1, \ldots, n$, yielding a set of estimates $\{\widehat{\delta}_i\}_{i=1}^n$; (ii) replace the unobserved δ_i on the left-hand side of (3.4) with estimates $\widehat{\delta}_i$ obtained from step (i); and then (iii) estimate

$$\widehat{\delta}_i = \boldsymbol{z}_i\boldsymbol{\beta} + \xi_i \geq 1 \tag{3.5}$$

using censored (tobit) regression, or in a few cases, ordinary least squares.[2]

Regardless of the form chosen for ψ, the two-stage approach outlined above presents problems for inference. First, the dependent variable in (2.6) and (3.4) is unobserved, and must be replaced by an estimate in the actual regression that is estimated. The

[2]Some (*e.g.,* Puig-Junoy *et al.,* 1998) have transformed the dependent variable while introducing an arbitrary constant to avoid taking the log of zero. Those who have used a tobit specification in the second stage have justified their approach by the fact that typically several, perhaps many, efficiency estimates equal unity in a given application. As far as we are aware, all who have specified a censored regression model in the second stage have constrained the processes that determine the probability of censoring and that govern the uncensored observations to be the same. See the Appendix for a discussion of censored and truncated regression models, their differences, and a discussion of the last point.

$\widehat{\delta}_i$ that are used in the second-step estimation of (3.5) are serially correlated, and in a complicated, unknown way. To understand this, note that $\delta_i = \widehat{\delta}(\boldsymbol{x}_i, \boldsymbol{y}_i \mid \widehat{\mathcal{P}})$ depends on all the observations $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ in $\mathcal{S}_n$ through $\widehat{\mathcal{P}}$, and consequently so must the error term ξ_i in (3.5). Moreover, while the observations in $\mathcal{S}_n$ are assumed independently drawn in assumption A1, $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ are correlated with $\boldsymbol{z}_i$ due to assumption A2—otherwise, there would be no motivation for the second-stage regression. This in turn means the the error term ξ_i in (3.5) is correlated with $\boldsymbol{z}_i$.

Both the correlation among the ξ_i as well as the correlation between ξ_i and $\boldsymbol{z}_i$ disappear asymptotically, but only at the same slow rate given in (3.3) with which $\widehat{\delta}_i$ converges. This means that maximum likelihood estimates of $\boldsymbol{\beta}$ in the second-stage regression will be consistent, but will not have the usual, parametric convergence rate of $n^{-1/2}$. More troubling, however, for $p + q > 3$, the correlation among the ξ_i does not disappear quickly enough for standard approaches to inference (based on the inverse of the negative Hessian of the log-likelihood) to be valid.

Further consideration reveals an additional problem. Note that we can always write

$$\widehat{\delta}_i = E(\widehat{\delta}_i) + u_i, \tag{3.6}$$

where $E(u_i) = 0$. In addition, the bias of the estimator $\widehat{\delta}_i$ is defined by

$$\text{BIAS}(\widehat{\delta}_i) \equiv E(\widehat{\delta}_i) - \delta_i. \tag{3.7}$$

Substituting for $E(\widehat{\delta}_i)$ from (3.6) in (3.7) and re-arranging terms yields

$$\delta_i = \widehat{\delta}_i - \text{BIAS}(\widehat{\delta}_i) - u_i. \tag{3.8}$$

Substituting for δ_i in (3.4) gives

$$\widehat{\delta}_i - \text{BIAS}(\widehat{\delta}_i) - u_i = \boldsymbol{z}_i\boldsymbol{\beta} + \varepsilon_i \geq 1. \tag{3.9}$$

Since $\widehat{\delta}_i$ is a consistent estimator, the u_i become negligible asymptotically, as does $\text{BIAS}(\widehat{\delta}_i)$. These facts provide justification for writing (3.5), the equation that is typically estimated in two-stage applications.

Although the u_i in (3.6) and (3.8)–(3.9) have zero mean, the term $\text{BIAS}(\widehat{\delta}_i)$ does not. Rather, the bias of $\widehat{\delta}_i$ is always strictly negative in finite samples. The u_i are unknown and cannot be estimated, but the bias term can be estimated by bootstrap methods; see Efron and Tibshirani (1993) for discussion, and Simar and Wilson (2000a) for an example similar to the present context. The bootstrap bias estimate equals the true bias plus a residual:

$$\widehat{\text{BIAS}}(\widehat{\delta}_i) = \text{BIAS}(\widehat{\delta}_i) + v_i. \tag{3.10}$$

The variance of the residual v_i diminishes as $n \to \infty$, and hence v_i is typically of smaller magnitude than $\text{BIAS}(\widehat{\delta}_i)$ for reasonable sample sizes n. The bootstrap estimator of bias can in turn be used to construct a bias-corrected estimator of δ:

$$\widehat{\widehat{\delta}}_i = \widehat{\delta}_i - \widehat{\text{BIAS}}(\widehat{\delta}_i). \tag{3.11}$$

Substituting for $\widehat{\text{BIAS}}(\widehat{\delta}_i)$ in (3.11) from (3.10), re-arranging terms, and then substituting for $\text{BIAS}(\widehat{\delta}_i)$ in (3.9) yields

$$\widehat{\widehat{\delta}}_i + v_i - u_i = \boldsymbol{z}_i\boldsymbol{\beta} + \varepsilon_i \geq 1. \tag{3.12}$$

As noted, both the terms v_i and u_i become negligible asymptotically; hence maximum likelihood estimation on

$$\widehat{\widehat{\delta}}_i \approx \boldsymbol{z}_i\boldsymbol{\beta} + \varepsilon_i \geq 1 \tag{3.13}$$

will yield consistent estimates.

Comparing (3.5) and (3.9), it is clear that estimation of (3.5) ignores both $\text{BIAS}(\widehat{\delta}_i)$ and u_i. Estimation of (3.13) ignores v_i and u_i. This aspect of the estimation problem resembles the story of measurement error in the dependent variable that is told in every undergraduate econometrics textbook. If v_i is indeed of smaller magnitude than $\text{BIAS}(\widehat{\delta}_i)$, we would expect estimates of $\boldsymbol{\beta}$ from (3.13) to be more statistically efficient than those from (3.5). We examine this question in our Monte Carlo experiments that follow.

In every two-stage application that we have found, the bias term in (3.9) has been ignored. In the special case where $p = q = 1$, Gijbels *et al.* (1999) demonstrate that $\text{BIAS}(\widehat{\delta}_i)$ is affected by the curvature of the boundary of $\mathcal{P}$. Although results do not exist for more general cases, one can speculate that a similar phenomenon exists in higher-dimensional spaces. In addition, it seems clear that $\text{BIAS}(\widehat{\delta}_i)$ will be larger in regions of $\mathcal{P}$ where data are sparse relative to other regions of $\mathcal{P}$ where the data are more dense. In regions where the data are sparse, there is less information about where the boundary of $\mathcal{P}$ lies, and hence the bias in estimated efficiency is likely to be larger.[3] All of this suggests that $\text{BIAS}(\widehat{\delta}_i)$, which is incorporated in the error term of (3.5) when it is ignored, is correlated with $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$, and hence with $\boldsymbol{z}_i$. While this correlation, as well as the bias itself, disappears asymptotically, it is reasonable to suppose that including an estimate of bias in the second-stage regression might improve efficiency of estimation of $\boldsymbol{\beta}$ in finite samples, and perhaps also improve the coverage of estimated confidence intervals for $\boldsymbol{\beta}$.

As a final remark, we note that almost all researchers have estimated (3.5) by assuming a censored normal (tobit) specification for ξ_i. The tobit specification is sometimes motivated by the observation that several values in $\{\widehat{\delta}_i\}_{i=1}^n$ are equal to unity, suggesting a probability mass at 1. However, it is important to recall that the underlying ***true*** model in (3.4) (or (2.6)) does not have this property. The process that determines whether $\widehat{\delta}_i = 1$ is primarily an artifact of finite samples, and has nothing to do with the process described by (2.6) in Assumption A2.

4. TOWARD BETTER ESTIMATION AND INFERENCE

The problems associated with estimating (3.5) and making inference about $\boldsymbol{\beta}$ arise from the serial correlation and bias of $\widehat{\delta}_i$, and the correlation between ξ_i and $\boldsymbol{z}_i$. The structure of these phenomena are unknown, and difficult to guess. Therefore, we propose a bootstrap procedure to overcome these difficulties; we describe the procedure in this section, and

[3] It is common for data to be rather unevenly distributed over $\mathcal{P}$ in applications; examples include data for banks (Wheelock and Wilson, 2000) and hospitals (Wilson and Carey, 2001).

present Monte Carlo evidence in the next section.

In order to implement a bootstrap procedure, we must draw iid bootstrap samples (*i.e.,* pseudo-data) $(\boldsymbol{x}_i^*, \boldsymbol{y}_i^*, \boldsymbol{z}_i^*)$ from a density $\widehat{f}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$. By now it is well-known that the ordinary, naive bootstrap based on resampling from the empirical distribution of the data is inconsistent in the present context due to the bounded nature of the DGP; see Simar and Wilson (1999a, 1999b, 2000b) for discussion. Simar and Wilson (1998, 2000a) described smoothed bootstrap procedures that overcomes these problems, but for the case with no environmental variables.

Note that Assumption A2 provides an extra piece of information that was not available in the model considered by Simar and Wilson (2000a). The bootstrap procedure proposed by Simar and Wilson (2000a) allows for heterogeneity in the distribution of δ, but incorporates no assumptions on the form of the heterogeneity; here, however, the form is made explicit by Assumption A2. Fortunately, the information provided by Assumption A2 allows considerable simplification in the bootstrap procedures we propose in this section.

We propose two bootstrap procedures for the two-stage efficiency estimation problem. The first procedure is designed to improve on inference, but without taking account of the bias term in (3.9):

Algorithm #1:

[1] Using the original data in $\mathcal{S}_n$, compute $\widehat{\delta}_i = \widehat{\delta}(x_i, y_i \mid \widehat{\mathcal{P}}) \; \forall \; i = 1, \ldots, n$ using (3.2).

[2] Use the method of maximum likelihood to obtain an estimate $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ as well as an estimate $\widehat{\sigma}_\varepsilon$ of σ_ε in the truncated regression of $\widehat{\delta}_i$ on $\boldsymbol{z}_i$ in (3.5).

[3] Loop over the next three steps ([3.1]–[3.3]) L times to obtain a set of bootstrap estimates $\mathcal{A} = \left\{ (\widehat{\boldsymbol{\beta}}^*, \widehat{\sigma}_\varepsilon^*)_b \right\}_{b=1}^{L}$:

[3.1] For each $i = 1, \ldots, n$, draw ε_i from the $N(0, \widehat{\sigma}_\varepsilon)$ distribution with left-truncation at $(1 - \boldsymbol{z}_i\widehat{\boldsymbol{\beta}})$.[4]

[4]See the Appendix for details on how to draw from a left-truncated normal distribution.

[3.2] Again for each $i = 1, \ldots, n$, compute $\delta_i^* = \boldsymbol{z}_i \widehat{\boldsymbol{\beta}} + \varepsilon_i$.

[3.3] Use the maximum likelihood method to estimate the truncated regression of δ_i^* on $\boldsymbol{z}_i$, yielding estimates $(\widehat{\boldsymbol{\beta}}^*, \widehat{\sigma}_\varepsilon^*)$.

[4] Use the bootstrap values in $\mathcal{A}$ and the original estimates $\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_\varepsilon$ to construct estimated confidence intervals for each element of $\boldsymbol{\beta}$ and for σ_ε as described below.

Alternatively, $\widehat{\widehat{\delta}}_i$ can be regressed on $\boldsymbol{z}_i$ in (3.13), and the following bootstrap procedure can be used to provide inference about $\boldsymbol{\beta}$.

Algorithm #2:

[1] Using the original data in $\mathcal{S}_n$, compute $\widehat{\delta}_i = \widehat{\delta}(x_i, y_i \mid \widehat{\mathcal{P}}) \; \forall \; i = 1, \ldots, n$ using (3.2).

[2] Use the method of maximum likelihood to obtain an estimate $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ as well as an estimate $\widehat{\sigma}_\varepsilon$ of σ_ε in the truncated regression of $\widehat{\delta}_i$ on $\boldsymbol{z}_i$ in (3.5).

[3] Loop over the next four steps ([3.1]–[3.4]) L_1 times to obtain n sets of bootstrap estimates $\mathcal{B}_i = \left\{ \widehat{\delta}_{ib}^* \right\}_{b=1}^{L_1}$:

[3.1] For each $i = 1, \ldots, n$, draw ε_i from the $N(0, \widehat{\sigma}_\varepsilon)$ distribution with left-truncation at $(1 - \boldsymbol{z}_i \widehat{\boldsymbol{\beta}})$.

[3.2] Again for each $i = 1, \ldots, n$, compute $\delta_i^* = \boldsymbol{z}_i \widehat{\boldsymbol{\beta}} + \varepsilon_i$.

[3.3] Set $\boldsymbol{x}_i^* = \boldsymbol{x}_i, \; \boldsymbol{y}_i^* = \boldsymbol{y}_i \widehat{\delta}_i / \delta_i^*$ for all $i = 1, \ldots, n$.

[3.4] Compute $\widehat{\delta}_i^* = \delta(\boldsymbol{x}_i, \boldsymbol{y}_i \mid \widehat{\mathcal{P}}^*) \; \forall \; i = 1, \ldots, n$, where $\widehat{\mathcal{P}}^*$ is obtained by replacing $\boldsymbol{Y}, \; \boldsymbol{X}$ in (3.1) with $\boldsymbol{Y}^* = [\, \boldsymbol{y}_1^* \quad \ldots \quad \boldsymbol{y}_n^* \,]$, $\boldsymbol{X}^* = [\, \boldsymbol{x}_1^* \quad \ldots \quad \boldsymbol{x}_n^* \,]$.

[4] For each $i = 1, \ldots, n$, compute the bias-corrected estimator $\widehat{\widehat{\delta}}_i$ defined by (3.11) using the bootstrap estimates in $\mathcal{B}_i$ obtained in step [3.4] and the original estimate $\widehat{\delta}_i$.

[5] Use the method of maximum likelihood to estimate the truncated regression of $\widehat{\widehat{\delta}}_i$ on $\boldsymbol{z}_i$, yielding estimates $(\widehat{\widehat{\boldsymbol{\beta}}}, \widehat{\widehat{\sigma}})$.

[6] Loop over the next three steps ([6.1]–[6.3]) L_2 times to obtain a set of bootstrap estimates $\mathcal{C} = \left\{(\widehat{\boldsymbol{\beta}}^*, \widehat{\sigma}^*_\varepsilon)_b\right\}_{b=1}^{L_2}$:

[6.1] For each $i = 1, \ldots, n$, draw ε_i from the $N(0, \widehat{\widehat{\sigma}})$ distribution with left-truncation at $(1 - \boldsymbol{z}_i\widehat{\widehat{\boldsymbol{\beta}}})$.

[6.2] Again for each $i = 1, \ldots, n$, compute $\delta_i^{**} = \boldsymbol{z}_i\widehat{\widehat{\boldsymbol{\beta}}} + \varepsilon_i$.

[6.3] Use the maximum likelihood method to estimate the truncated regression of δ_i^{**} on $\boldsymbol{z}_i$, yielding estimates $(\widehat{\widehat{\boldsymbol{\beta}}}^*, \widehat{\widehat{\sigma}}^*)$.

[7] Use the bootstrap values in $\mathcal{C}$ and the original estimates $\widehat{\widehat{\boldsymbol{\beta}}}, \widehat{\widehat{\sigma}}$ to construct estimated confidence intervals for each element of $\boldsymbol{\beta}$ and for σ_ε as described below.

Algorithm #2 employs a bootstrap procedure to compute $\widehat{\widehat{\delta}}_i$ in step [4]. The bootstrap in step [6] is similar to the one in Algorithm #1, except that the one in Algorithm #2 is based on the bias-corrected distance function estimates.

In either case, once bootstrap the set of bootstrap values $\mathcal{A}$ or $\mathcal{C}$ has been obtained either in step 3 of Algorithm #1 or step 6 of Algorithm #2, percentile bootstrap confidence intervals can be constructed. To illustrate, suppose interest lies in β_j, the jth element of $\boldsymbol{\beta}$, which has been estimated by $\widehat{\widehat{\beta}}_j$, the jth element of $\widehat{\widehat{\boldsymbol{\beta}}}$. If the distribution of $(\widehat{\widehat{\beta}}_j - \beta_j)$ were known, it would be trivial to find values $a_\alpha,\ b_\alpha$ such that

$$\Pr\left[-b_\alpha \leq (\widehat{\widehat{\beta}}_j - \beta_j) \leq -a_\alpha\right] = 1 - \alpha \tag{4.1}$$

for some small value of α, $0 < \alpha < 1$, say $\alpha = .05$. Since the distribution of $(\widehat{\widehat{\beta}}_j - \beta_j)$ is unknown, we can use the jth element of each bootstrap value $\widehat{\widehat{\boldsymbol{\beta}}}^*$ to find values $a^*_\alpha,\ b^*_\alpha$ such that

$$\Pr\left[-b_\alpha \leq (\widehat{\widehat{\beta}}^*_j - \widehat{\widehat{\beta}}_j) \leq -a^*_\alpha\right] \approx 1 - \alpha, \tag{4.2}$$

with improving approximation as $L_2 \to \infty$. Substituting $a^*_\alpha,\ b^*_\alpha$ for $a_\alpha,\ b_\alpha$ in (4.1) leads to an estimated confidence interval $\left[\widehat{\widehat{\beta}}_j + a^*_\alpha,\ \widehat{\widehat{\beta}}_j + b^*_\alpha\right]$.

In Simar and Wilson (2001b), the bootstrap principle was iterated to assess the accuracy of bootstrap confidence interval estimates, along the lines described by Hall (1992). Although a similar procedure could be employed to assess the accuracy of the confidence intervals estimated with Algorithms #1 and #2, the double bootstrap in Algorithm #2 uses sequential bootstraps, rather than iterated bootstraps. Moreover, the first loop in step [3] of Algorithm #2 is used only to construct bias-corrected distance function estimates, and so it is reasonable to use fewer replications than when estimating confidence intervals in the second loop. Consequently, the computational burden incurred by Algorithm #2 is far less than what would typically be incurred with an iterated bootstrap as in Simar and Wilson (2001b).

5. MONTE CARLO EXPERIMENTS

To examine the performance of the various approaches to inference in the second-stage regression, we conducted several Monte-Carlo experiments. In each case, we generated data from a known process and applied our bootstrap algorithms on each of M Monte Carlo trials.

Data for the ith observation in each Monte Carlo trial were generated using the following algorithm.

Algorithm #3:

[1] Set $z_{i1} = 1$; draw $z_{ij} \sim N(\mu_z, \sigma_z^2)$ for $j = 2, \ldots, r$.

[2] Draw $v \sim$ uniform on the interval $(\Phi(1 - \boldsymbol{z}_i\boldsymbol{\beta}),\ 1)$, and set $\varepsilon_i = \Phi^{-1}(v)$, where $\Phi(\cdot)$ denotes the standard Gaussian distribution function and $\Phi^{-1}(\cdot)$ denotes the standard Gaussian quantile function. Thus ε_i is a $N(0, \sigma_\varepsilon^2)$ deviate left-truncated $1 - \boldsymbol{z}_i\boldsymbol{\beta}$.

[3] Set $\delta_i = \boldsymbol{z}_i\boldsymbol{\beta} + \varepsilon_i$.

[4] For each $j = 1, \ldots, p$, draw $x_{ij} \sim$ uniform$(10, 20)$.

[5] If $q = 1$, then set $y_i = \delta_i^{-1} \sum_{j=1}^{p} (x_{ij} - 4)^{3/4}$. If $q > 1$, then perform steps [5.1]–[5.4]:

[5.1] Set $\zeta = \delta_i^{-1} \sum_{j=1}^{p} (x_{ij} - 4)^{3/4}$.

[5.2] Draw $\alpha_1 \sim \text{uniform}(0, 1)$.

[5.3] If $q > 2$, then for each $\ell = 2, \ldots, q-1$ draw $\alpha_\ell \sim \text{uniform}\left(0, 1 - \sum_{\ell=1}^{q-2} \alpha_\ell\right)$.

[5.4] For $j = 1, \ldots, q-1$, set $y_{ij} = \delta_i^{-1} \alpha_j \zeta$; set $y_{iq} = \delta_i^{-1} \left(1 - \sum_{\ell=1}^{q-1} \alpha_\ell\right) \zeta$.

In steps [5.1]–[5.4] for the case $q > 1$, aggregate output ζ is computed in step [5.1] and then disaggregated among the q individual outputs in steps [5.2]–[5.4]. Although not required by Assumptions A1–A8, steps [5.2]–[5.4] result in independence between the mix of outputs, characterized by the angles $\boldsymbol{\eta}$, and $(\boldsymbol{x}, \delta, \boldsymbol{z})$.

In each of our experiments, we set $r = 2$, $\mu_z = 2$, $\sigma_z = 2$, $\sigma_\varepsilon = 1$. In addition, values of each element of $\boldsymbol{\beta}$ must be set. In order to simplify scaling in estimation of the second-stage truncated regression, we set $\beta_1 = \beta_2 = 0.5$.

In each experiment, we ran 1000 Monte Carlo trials. For Algorithm #1, we use $L = 2000$ bootstrap replications. For Algorithm #2, we use $L_1 = 100$ replications for the first loop used to compute the bias-corrected efficiency estimates, and $L_2 = 2000$ replications for the second loop where the truncated regression model is bootstrapped.

Our first set of experiments involve generating data using Algorithm #3, applying the single bootstrap in Algorithm #1, and repeating these steps over 1000 Monte Carlo trials. Table 1 gives the proportions of the 1000 Monte Carlo trials in each experiment where the estimated confidence interval at a particular nominal significance level (.80, .90, .95, and .99) includes the true value for a given parameter. For $p = q = 1$ and $n = 100$, the coverages are too small, but not by a large amount. Coverages improve when the sample size is increased to 400, as expected. As p and q are increased, the coverages become slightly worse, but not too much so; coverages with $p = q = 3$ and $n = 400$ are roughly as accurate as the case where $p = q = 1$ and $n = 100$.

In each of the experiments used to produce the results in Table 1, we also applied conventional inference methods on each Monte Carlo trial to obtain the coverages shown in Table 2. Although coverages by the estimated confidence intervals for the slope and

variance terms are roughly similar to those in Table 1, the coverages for the intercept term are poor, and equal 1.0000 even at a nominal significance of .80 in several cases. Tobit regression with conventional inference was also applied to the data generated on each Monte Carlo Trial, producing the coverage results shown in Table 3. This approach involves a specification error in the sense that the model estimated is mis-specified. Not surprisingly, the results in Table 3 indicate that this approach results in catastrophe.

Additional insight is gained by examining the distributions of the estimators over the 1000 Monte Carlo trials in our experiments. Figure 1 shows kernel density estimates for these distributions for the case where $p = q = 2$ and $n = 400$.[5] Figure 1 contains six plots of density estimates; the rows (from top to bottom) correspond to $\widehat{\beta}_1$, $\widehat{\beta}_2$, and $\widehat{\sigma}_\varepsilon$; the first column corresponds to the Tobit estimates described in Table 3, while the second column corresponds to the truncated regression estimates described in Tables 1–2. True values of the parameters are indicated by the dashed lines. Figure 1 makes clear why the coverages of estimated confidence intervals from the Tobit model are so poor—none of the estimated values lie near the true values. The situation is much better for the truncated regression; although the distributions for $\widehat{\beta}_1$ and $\widehat{\beta}_2$ show small amounts of skewness, they are rather well-centered over the true values.

A similar set of Monte Carlo experiments were performed to obtain the results in Tables 4–5. In the case of Table 4, we employed the double bootstrap in Algorithm #2. Comparison of these results with those in Table 1 reveal little advantage in the double bootstrap in terms of coverage for the intercept, but a significant improvement in coverages of the slope parameter and the variance term, at each value of p, q, and n that we examined (coverage of the intercept term was already quite good with the single bootstrap). Given that Algorithm #2 involves only a small increase in computational burden over Algorithm #1, the improved performance of Algorithm #2 seems to justify its use.

[5]The kernel density estimates were obtained using a Gaussian kernel and with bandwidths chosen by least-squares cross-validation.

In this second set of Monte Carlo experiments we also applied conventional inference after the regression of $\widehat{\widehat{\delta}}_i$ on $\boldsymbol{z}_i$ in step [5] of Algorithm #2. Coverage results from this approach are shown in Table 5. The problems with this approach are similar to those found in Table 2, except that here in addition to coverage of the intercept being corrupted, coverages of the slope and variance parameters are also implausibly large in a number of instances. For example, with $p = q = 3$ $n = 100$, and .90 nominal significance, this approach yields coverage of 0.937; when the sample size is increased to $n = 400$, coverage also increases to 0.982 instead of moving close to the nominal level. This pattern is found in a number of instances in Table 5.

Kernel estimates of the densities of the estimators from the second set of experiments are shown in Figure 2, again for the case $p = q = 2$, $n = 400$, arranged as in Figure 1, with estimates from the regression of $\widehat{\delta}_i$ on $\boldsymbol{z}_i$ in the first column, and estimates from the regression of $\widehat{\widehat{\delta}}_i$ on $\boldsymbol{z}_i$ in the second column (true parameter values are indicated by dashed lines). The densities in the first column exhibit slightly greater skewness than those in the second column; on the whole, the densities for the regression of $\widehat{\widehat{\delta}}_i$ are better-centered on the true values than the densities for the regression of $\widehat{\delta}_i$.

The second set of experiments conducted to produce the results in Table 4–5 also allow comparison of root-mean-square-error (RMSE) of each estimator in regressions of $\widehat{\delta}_i$ on $\boldsymbol{z}_i$ versus regressions of $\widehat{\widehat{\delta}}_i$ on $\boldsymbol{z}_i$. Table 6 shows RMSE for each estimator (computed over 1000 Monte Carlo Trials) with varying model dimensions and sample sizes. For $p = q = 1$, 2, or 3 and a sample size of 100, RMSE for each estimator is smaller when $\widehat{\delta}_i$ is regressed on $\boldsymbol{z}_i$ than when $\widehat{\widehat{\delta}}_i$ is regressed on $\boldsymbol{z}_i$. However, when the sample size is increased to 400, with $p = q = 1$ or 2 use of $\widehat{\widehat{\delta}}_i$ yields lower RMSE for the intercept and slope estimators than $\widehat{\delta}_i$. For $p = q = 3$, use of $\widehat{\widehat{\delta}}_i$ yields lower RMSE for the intercept and slope estimators than $\widehat{\delta}_i$ when the sample size is increased to 800. At small sample sizes, the curse of dimensionality is more critical for models with larger dimensions than for simpler models. In each case shown in Table 6, RMSE for the variance estimator is larger when $\widehat{\widehat{\delta}}_i$ is used

as the dependent variable than when $\widehat{\delta}_i$ is used, perhaps reflecting the fact that additional noise has been introduced by adding the estimate of bias to $\widehat{\delta}_i$.

6. SUMMARY AND CONCLUSIONS

The Monte Carlo results presented in the previous assumption illustrate a number of problems with existing two-stage studies that employ nonparametric distance function estimators similar to (3.2) in the first stage, and then employ tobit regression in the second stage with conventional inference based on the inverse negative Hessian of the log-likelihood function. As noted in the introduction, none of the published studies that we are aware of have defined a DGP that might be estimated, and it is difficult to imagine one where tobit regression would be sensible in this context. In terms of coverage of estimated confidence intervals, tobit regression is catastrophic in our Monte Carlo experiments.

Truncated regression estimates the correct model in our experiments, but conventional inference—either when the uncorrected or bias-corrected distance function estimates are used—does not perform well. In terms of coverage of estimated confidence intervals, our single bootstrap is shown to perform well, but our double bootstrap performs even better, with little increase in computational burden over the single bootstrap. The double bootstrap has the additional advantage that RMSE of the intercept and slope estimators in the second-stage regression declines more rapidly than when the single bootstrap is used, resulting in lower RMSE at moderate sample sizes depending on model dimensionality in the first stage. The double bootstrap is thus our preferred choice.

APPENDIX: TECHNICAL DETAILS

Censored versus Truncated Regression:

Consider the regression model

$$\vartheta_i = \boldsymbol{z}_i\boldsymbol{\beta} + \varepsilon_i, \tag{A.1}$$

where $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ is identically, independently distributed for all $i = 1, \ldots, n$. The left-hand side variable ϑ is said to be ***censored*** when, instead of observing ϑ_i for all

observations, we observe

$$\theta_i = \begin{cases} \boldsymbol{z}_i + \varepsilon_i & \text{if } \boldsymbol{z}_i + \varepsilon_i > c_i \\ c_i & \text{otherwise.} \end{cases} \tag{A.2}$$

In this case, ϑ is left-censored at the constant c_i, which may vary over observations. Alternatively, ϑ_i is said to be ***truncated*** if we observe $\theta_i = \vartheta_i$ for all $\vartheta_i \geq c_i$, but observe nothing otherwise.

Relative to the classical linear regression model with unbounded error terms, both censoring and truncation involve a loss of information about the dependent variable. In the case of censoring, some information is lost about the dependent variable for censored observations, but the right-hand side variables are observed for such observations. In the case of truncation, neither the left-hand nor the right-hand side variables are observed for some observations. The information loss is thus more severe in the case of truncation than in the case of censoring.

In the case of truncation, if the ϑ_i are assumed normal with left-truncation at c_i, $\boldsymbol{\beta}$ in (A.1) can be estimated by maximizing the likelihood function

$$\mathcal{L}_1 = \prod_{i=1}^{n} \frac{1}{\sigma_\varepsilon} \phi\left(\frac{\theta_i - \boldsymbol{z}_i\boldsymbol{\beta}}{\sigma_\varepsilon}\right) \left[1 - \Phi\left(\frac{c_i - \boldsymbol{z}_i\boldsymbol{\beta}}{\sigma_\varepsilon}\right)\right]^{-1}, \tag{A.3}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ represent the standard normal density and distribution functions, respectively. This resembles the likelihood for regression models with normal errors that are neither censored nor truncated, except for the term in square brackets. Division by this term is necessary to re-scale the normal density $\phi(\cdot)$ so that it integrates to unity after truncation.

In the case of left-censoring at c_i where one assumes that the un-censored observations are normally distributed, $\boldsymbol{\beta}$ in (A.1) can be estimated by specifying a functional form for $\text{Prob}(\vartheta_i > c_i \mid \boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is a vector of parameters, and writing the likelihood function

as

$$\mathcal{L}_2 = \prod_{i|\theta_i > c_i} \frac{1}{\sigma_\varepsilon}\phi\left(\frac{\theta_i - \boldsymbol{z}_i\boldsymbol{\beta}}{\sigma_\varepsilon}\right)\left[1 - \Phi\left(\frac{c_i - \boldsymbol{z}_i\boldsymbol{\beta}}{\sigma_\varepsilon}\right)\right]^{-1} \times \text{Prob}(\vartheta_i > c_i \mid \boldsymbol{\alpha}) \prod_{i|\theta_i = c_i} \left[1 - \text{Prob}(\vartheta_i > c_i \mid \boldsymbol{\alpha})\right]. \tag{A.4}$$

Here, the probability that $\vartheta_i > c_i$ is allowed to result from a different process than the one that determines ϑ when $\vartheta_i > c_i$. The probability $\text{Prob}(\vartheta_i > c_i \mid \boldsymbol{\alpha})$ would typically be specified as probit or logit, but many other specifications are possible.

The tobit model is a special case of (A.4), where the probability that $\vartheta_i > c_i$ is assumed to be controlled by the same process that that determines ϑ when $\vartheta_i > c_i$. In other words, the tobit model incorporates the additional assumption that

$$\text{Prob}(\vartheta_i > c_i) = \left[1 - \Phi\left(\frac{c_i - \boldsymbol{z}_i\boldsymbol{\beta}}{\sigma_\varepsilon}\right)\right] \tag{A.5}$$

Substitution of the expression in (A.5) for $\text{Prob}(\vartheta_i > c_i \mid \boldsymbol{\alpha})$ in (A.4) yields the tobit likelihood function,

$$\mathcal{L}_3 = \prod_{i|\theta_i > c_i} \frac{1}{\sigma_\varepsilon}\phi\left(\frac{\theta_i - \boldsymbol{z}_i\boldsymbol{\beta}}{\sigma_\varepsilon}\right) \prod_{i|\theta_i = c_i} \Phi\left(\frac{c_i - \boldsymbol{z}_i\boldsymbol{\beta}}{\sigma_\varepsilon}\right). \tag{A.6}$$

The terms in the second product expression reflect that fact that all that is known about the censored observations is that ϑ_i lies below c_i.

Even if the censored model was the correct specification, use of the Tobit version of this model would be curious in the context of the two-stage models discussed in this paper. Note that the likelihood $\mathcal{L}_2$ in (A.4) is multiplicatively separable in $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. Both the censored as well as the un-censored observations identify $\boldsymbol{\alpha}$, while only the un-censored observations identify $\boldsymbol{\beta}$. In fact, the separability means that $\mathcal{L}_2$ can be maximized with respect to each parameter vector independently; maximization with respect to $\boldsymbol{\beta}$ is equivalent to maximization of the likelihood $\mathcal{L}_1$ in (A.3) for the truncated model. By contrast, when the tobit specification is used, both the censored as well as the un-censored

observations determine $\widehat{\boldsymbol{\beta}}$ when the likelihood $\mathcal{L}_3$ in (A.6) is maximized. A better approach would be to specify $\text{Prob}(\vartheta_i > c_i \mid \boldsymbol{\alpha})$ as a probit probability, maximize both $\mathcal{L}_2$ and $\mathcal{L}_3$, and perform a likelihood-ratio test of the null hypothesis $\boldsymbol{\beta} = \boldsymbol{\alpha}$. None of the two-stage studies that have used the tobit specification in the second stage that we are aware of have done this. Of course, the censored model is not the correct specification, so perhaps the point is moot.

Both the truncated and the tobit regression models can easily be estimated by maximum likelihood. In either case, the task is made easier when Olsen's (1978) re-parameterization is used. With this re-parameterization, maximization by Newton's method typically converges very quickly, requiring only a few iterations. For the practitioner, commands that perform truncated regression are available in a number of popular software packages.

Obtaining Draws from a Left-Truncated Normal Distribution:

Both bootstrap algorithms (Algorithms #1 and #2), as well as the simulation algorithm in section 5 (Algorithm #3), require iid draws from a $N(0, \sigma^2)$ distribution with left-truncation at a constant c. This can be accomplished quickly and easily using a modified transformation method. Let $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ denote the standard normal distribution function and the standard normal quantile function, respectively, so that $u = \Phi^{-1}(\Phi(u))$. Generate v uniform on $(0, 1)$, let $c' = c/\sigma$, and set $v' = \Phi(c') + [1 - \Phi(c')]v$. Then compute $u = \sigma\Phi^{-1}(v')$ to obtain the desired left-truncated normal deviate.

REFERENCES

Andersen, P. and N.C. Petersen (1993), A procedure for ranking efficient units in Data Envelopment Analysis, *Management Science* 39, 1261–1264.

Arnold, V.L., I.R. Bardhan, W.W. Cooper, and S.C. Kumbhakar (1996), New uses of DEA and statistical regressions for efficiency and estimation: Texas schools, *Annals of Operations Research* 66, 255–277.

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

Banker, R.D. and H.H. Johnston (1994), Evaluating the impacts of operating strategies on efficiency in the US airline industry, in *Data Envelopment Analysis: Theory, Methodology, and Application,* ed. by A. Charnes, W.W. Cooper, A.Y. Levin, and L.M. Seiford, Boston: Kluwer Academic Publishers, Inc., pp. 97–128.

Burgess, J.F. and P.W. Wilson (1998), Variation in inefficiency among US hospitals, *Canadian Journal of Operational Research and Information Processing (INFOR)* 36, 84–102.

Byrnes, P., R. Färe, S. Grosskopf, and C.A.K. Lovell (1988), The effect of unions on productivity: U.S. surface mining of coal, *Management Science* 34, 1037–1053.

Carrington, R., N. Puthucheary, D. Rose, and S. Yaisawarng (1997), Performance measurement in government service provision: The case of police services in New South Wales, *Journal of Productivity Analysis* 8, 415–430.

Chakraborty, K., B. Biswas, and W.C. Lewis (2001), Measurement of technical efficiency in public education: A stochastic and nonstochastic production approach, *Southern Economic Journal* 67, 889–905.

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

Charnes, A., Cooper, W.W. and Rhodes E. (1979), Measuring the efficiency of decision making units, *European J. Oper. Res.* 3, 339.

Chilingerian, J.A. (1995), Evaluating physician efficiency in hospitals: A multivariate analysis of best practices, *European Journal of Operational Research* 80, 548–574.

Chirkos, T.N. and A.M. Sears (1994), Technical efficiency and the competitive behavior of hospitals, *Socio-Economic Planning Science* 28, 219–227.

De Borger, B. and K. Kerstens (1996), Cost efficiency of Belgian local governments: A comparative analysis of FDH, DEA, and econometric approaches, *Regional Science and Urban Economics* 26, 145–170.

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

Dietsch, M. and L. Weill (1999), Les performances des banques de dépôts françaises: une évaluation par la méthod DEA, dans *La Méthod DEA,* ed. par P.Y. Badillo et J.C. Paradi, Paris: Hermes Science Publications.

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

Fan, J. and I. Gijbels (1996), *Local Polynomial Modelling and Its Applications,* London: Chapman & Hall.1l

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

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

Farrell, M.J. (1957), The measurement of productive efficiency, *J. Roy. Statist. Soc. Ser. A* 120, 253-281.

Fried, H.O., C.A.K. Lovell, S.S. Schmidt, and S. Yaisawarng (2002), Accounting for environmental effects and statistical noise in data envelopment analysis, *Journal of Productivity Analysis* 17, 157–174.

Fried, H.O., C.A.K. Lovell, and P. Vanden Eeckaut (1993), Evaluating the performance of U.S. Credit Unions, *Journal of Banking and Finance* 17, 251–265.

Fried, H.O., C.A.K. Lovell, and S. Yaisawarng (1999), The impact of mergers on credit union service provision, *Journal of Banking and Finance* 23, 367–386.

Fried, H.O., S.S. Schmidt, and S. Yaisawarng (1999), Incorporating the operating environment into a nonparametric measure of technical efficiency, *Journal of Productivity Analysis* 12, 249–267.

Garden, K.A. and D.E. Ralston (1999), The x-efficiency and allocative efficiency effects of credit union mergers, *Journal of International Financial Markets* 9, 285–301.

Gijbels, I., Mammen, E., Park, B.U. and Simar, L. (1999), On estimation of monotone and concave frontier functions, *J. Amer. Statist. Assoc.* 94, 220–228.

González, B. and P. Barber (1996), Changes in the efficiency of Spanish public hospitals after the introduction of program-contracts, *Investigaciones Económicas* 20, 377–402.

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

Hirschberg, J.G. and P.J. Lloyd (2002), "Does the Technology of Foreign-Invested Enterprises Spill Over to Other Enterprises in China? An Application of Post-DEA Bootstrap Regression Analysis," in *Modelling the Chinese Economy,* ed. by P.J. Lloyd and X.G. Zang, London: Edward Elgar Press.

Honoré, B.E. and J.L. Powell (1994), Pairwise difference estimators of censored and truncated regression models, *Journal of Econometrics* 64, 241–278.

Isik, I., and M.K. Hassan (2002), Technical, scale, and allocative efficiencies of Turkish banking industry, *Journal of Banking and Finance* 26, 719–766.

Kirjavainen, T. and H.A. Loikkanen (1998), Efficiency differences of Finnish senior secondary schools: An application of DEA and tobit analysis, *Economics of Education Review* 17, 377–394.

Kneip, A., Park, B.U. and Simar, L. (1998), A note on the convergence of nonparametric DEA estimators for production efficiency scores, *Econometric Theory*, 14, 783–793.

Kooreman, P. (1994), Nursing home care in The Netherlands: A nonparametric efficiency analysis, *Journal of Health Economics* 13, 301–316.

Korostelev, A., Simar, L. and Tsybakov, A.B. (1995), On estimation of monotone and convex boundaries, *Publ. Statist. Univ. Paris XXXIX* 1, 3–18.

Lewbel, A. and O. Linton (2002), Nonparametric censored and truncated regression, *Econometrica* 70, 765–779.

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.

Lovell, C.A.K., L.C. Walters, and L.L. Wood (1994), "Stratified Models of Education Production Using Modified DEA and Regression Analysis," in *Data Envelopment Analysis: Theory, Methodology, and Applications,* ed. by A. Charnes, W.W. Cooper, A.Y. Lewin, and L.M. Seiford, Boston: Kluwer Academic Publishers.

Luoma, K., M-L. Järviö, I. Suoniemi, R.T. Hjerppe (1996), Financial incentives and productive efficiency in Finnish health services, *Health Economics* 5, 435–445.

McCarty, T.A. and S. Yaisawarng (1993), Technical efficiency in New Jersey school districts, in *The Measurement of Productive Efficiency,* ed. by H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, New York: Oxford University Press.

McMillan, M.L. and D. Datta (1998), The relative efficiencies of Canadian universitites: A DEA perspective, *Canadian Public Policy—Analyse de Politiques* 24, 485–511.

Mukherjee, K., S.C. Ray, and S.M. Miller (2001), Productivity growth in large US commercial banks: The initial post-deregulation experience, *Journal of Banking and Finance* 25, 913–939.

Olsen, R. (1978), A note on the uniqueness of the maximum likelihood estimator in the tobit model, *Econometrica* 46, 1211–1215.

Puig-Junoy, J. (1998), Technical efficiency in the clinical management of critically ill patients, *Health Economics* 7, 263–277.

Ralston, D., A. Wright, and K. Garden (2001), Can mergers ensure the survival of credit unions in the third millennium? *Journal of Banking and Finance* 25, 2277–2304.

Ray, S.C. (1988), Data envelopment analysis, nondiscretionary inputs and efficiency: An alternative interpretation, *Socio-Economic Planning Science* 22, 167–176.

Ray, S.C. (1991), Resource-use efficiency in public schools: A study of Connecticut data, *Management Science* 37, 1620–1628.

Resende, M. (2000), Regulatory regimes and efficiency in US local telephony, *Oxford Economic Papers* 52, 447–470.

Rhodes, E.L. and L. Southwick, Jr. (1993), "Variations in Public and Private University Efficiency," in *Applications of Management Science,* volume 7, ed. by E.L. Rhodes, Greenwich, Connecticut: JAI Press, Inc.

Rosko, M.D., J.A. Chilingerian, J.S. Zinn, and W.E. Aaronson (1995), The effects of ownership, operating environment, and strategic choices on nursing home efficiency, *Medical Care* 33, 1001–xxxx.

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

Seiford, L. M. (1997), A bibliography for data envelopment analysis (1978–1996), *Ann. Oper. Res.* 73, 393–438.

Sexton, T.R., S. Sleeper, and R.E. Taggart, Jr. (1994), Improving pupil transportation in North Carolina, *Interfaces* 24, 87–103.

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

Simar, L. and Wilson, P.W. (1998), Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models, *Management Sci.* 44(11), 49–61.

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

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

Simar, L. and Wilson, P.W. (2000a), A general methodology for bootstrapping in nonparametric frontier models, *J. Appl. Statist.* 27, 779–802.

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

Simar, L. and Wilson, P.W. (2001a), Testing restrictions in nonparametric efficiency models, *Communications in Statistics* 30, 159–184.

Simar, L. and Wilson, P.W. (2001b), "Aplicación del 'Bootstrap' para Estimadores D.E.A.," in *La Medición de la Eficiencia y la Productividad,* edited by A. Álvarez, Madrid: Pirámide (2001). Translation of "Performance of the Bootstrap for DEA Estimators and Iterating the Principle," Discussion Paper No. 0002, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Simar, L. and Wilson, P.W. (2002), Nonparametric tests of returns to scale, *European Journal of Operational Research* 139, 115–132.

Stanton, K.R. (2002), Trends in relationship lending and factors affecting relationship lending efficiency, *Journal of Banking and Finance* 26, 127–152.

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

Wilson, P.W. (2003), Testing independence in models of productive efficiency, *Journal of Productivity Analysis,* forthcoming.

Wilson, P.W., and K. Carey (2002), "Nonparametric Analysis of Returns to Scale and Product Mix among US Hospitals," unpublished working paper, Department of Economics, University of Texas, Austin, Texas 78712 USA.

Worthington, A.C. and B.E. Dollery (2000), Productive efficiency and the Australian local government grants process, *Australian Journal of Regional Studies* 6, 95–121.

Xue, M. and P.T. Harker (2002), "Overcoming the Inherent Dependency of DEA Efficiency Scores: A Bootstrap Approach," unpublished working paper, Wharton Financial Institutions Center, University of Pennsylvania.

TABLE 1

Coverage of Estimated Confidence Intervals:
Truncated Regression of $\widehat{\delta}_i$ on $\boldsymbol{z}_i$,
Inference by Single Bootstrap (Algorithm #1)

			nominal significance			
$p = q$	n	parameter	.80	.90	.95	.99
1	100	β_1	0.814	0.871	0.893	0.925
		β_2	0.743	0.820	0.859	0.916
		σ_ε	0.747	0.811	0.851	0.905
1	400	β_1	0.807	0.910	0.947	0.974
		β_2	0.780	0.874	0.922	0.965
		σ_ε	0.771	0.871	0.907	0.953
2	100	β_1	0.886	0.929	0.948	0.971
		β_2	0.657	0.727	0.781	0.847
		σ_ε	0.663	0.743	0.790	0.852
2	400	β_1	0.770	0.929	0.979	0.991
		β_2	0.729	0.821	0.883	0.938
		σ_ε	0.735	0.812	0.870	0.923
3	100	β_1	0.895	0.942	0.958	0.970
		β_2	0.621	0.695	0.749	0.806
		σ_ε	0.656	0.717	0.761	0.817
3	400	β_1	0.624	0.890	0.990	0.999
		β_2	0.649	0.760	0.820	0.900
		σ_ε	0.717	0.791	0.849	0.921

TABLE 2

Coverage of Estimated Confidence Intervals:
Truncated Regression of $\widehat{\delta}_i$ on $\boldsymbol{z}_i$,
Conventional Inference

			nominal significance			
$p = q$	n	parameter	.80	.90	.95	.99
1	100	β_1	1.000	1.000	1.000	1.000
		β_2	0.783	0.847	0.885	0.940
		σ_ε	0.754	0.816	0.869	0.935
1	400	β_1	0.938	0.995	1.000	1.000
		β_2	0.830	0.901	0.936	0.975
		σ_ε	0.810	0.882	0.914	0.966
2	100	β_1	1.000	1.000	1.000	1.000
		β_2	0.684	0.765	0.817	0.896
		σ_ε	0.666	0.755	0.813	0.895
2	400	β_1	0.853	0.988	1.000	1.000
		β_2	0.750	0.840	0.900	0.959
		σ_ε	0.741	0.814	0.881	0.935
3	100	β_1	1.000	1.000	1.000	1.000
		β_2	0.666	0.745	0.794	0.867
		σ_ε	0.667	0.742	0.797	0.864
3	400	β_1	0.841	0.996	1.000	1.000
		β_2	0.662	0.777	0.841	0.936
		σ_ε	0.720	0.797	0.860	0.937

TABLE 3

Coverage of Estimated Confidence Intervals:
Tobit Regression of $\widehat{\delta}_i$ on $\boldsymbol{z}_i$,
Conventional Inference

$p = q$	n	parameter	nominal significance .80	.90	.95	.99
1	100	β_1	1.000	1.000	1.000	1.000
		β_2	0.000	0.000	0.000	0.000
		σ_ε	0.000	0.000	0.000	0.000
1	400	β_1	1.000	1.000	1.000	1.000
		β_2	0.031	0.044	0.059	0.096
		σ_ε	0.000	0.000	0.000	0.000
2	100	β_1	1.000	1.000	1.000	1.000
		β_2	0.042	0.059	0.086	0.145
		σ_ε	0.000	0.000	0.000	0.000
2	400	β_1	1.000	1.000	1.000	1.000
		β_2	0.001	0.001	0.001	0.001
		σ_ε	0.000	0.000	0.000	0.000
3	100	β_1	1.000	1.000	1.000	1.000
		β_2	0.079	0.112	0.147	0.217
		σ_ε	0.000	0.000	0.000	0.000
3	400	β_1	1.000	1.000	1.000	1.000
		β_2	0.003	0.006	0.006	0.009
		σ_ε	0.000	0.000	0.000	0.000

TABLE 4

Coverage of Estimated Confidence Intervals:
Truncated Regression of $\widehat{\widehat{\delta}}_i$ on $\boldsymbol{z}_i$,
Inference by Double Bootstrap (Algorithm #2)

			nominal significance			
$p = q$	n	parameter	.80	.90	.95	.99
1	100	β_1	0.801	0.867	0.893	0.928
		β_2	0.782	0.875	0.912	0.947
		σ_ε	0.806	0.886	0.907	0.956
1	400	β_1	0.798	0.895	0.936	0.973
		β_2	0.799	0.897	0.943	0.980
		σ_ε	0.814	0.895	0.945	0.979
2	100	β_1	0.850	0.967	0.978	0.985
		β_2	0.805	0.902	0.933	0.962
		σ_ε	0.813	0.925	0.950	0.972
2	400	β_1	0.753	0.910	0.975	0.990
		β_2	0.794	0.906	0.960	0.987
		σ_ε	0.777	0.910	0.964	0.992
3	100	β_1	0.726	0.970	0.993	1.000
		β_2	0.793	0.908	0.935	0.966
		σ_ε	0.820	0.922	0.951	0.973
3	400	β_1	0.380	0.641	0.869	0.999
		β_2	0.734	0.873	0.961	0.994
		σ_ε	0.702	0.878	0.961	0.998

TABLE 5

Coverage of Estimated Confidence Intervals:
Truncated Regression of $\widehat{\widehat{\delta}}_i$ on $\boldsymbol{z}_i$,
Conventional Inference

			nominal significance			
$p = q$	n	parameter	.80	.90	.95	.99
1	100	β_1	1.000	1.000	1.000	1.000
		β_2	0.853	0.897	0.925	0.966
		σ_ε	0.854	0.888	0.924	0.966
1	400	β_1	0.943	0.998	1.000	1.000
		β_2	0.873	0.929	0.963	0.990
		σ_ε	0.879	0.919	0.952	0.986
2	100	β_1	1.000	1.000	1.000	1.000
		β_2	0.885	0.926	0.949	0.977
		σ_ε	0.895	0.934	0.956	0.986
2	400	β_1	0.854	0.988	1.000	1.000
		β_2	0.919	0.955	0.977	0.993
		σ_ε	0.933	0.967	0.979	0.995
3	100	β_1	1.000	1.000	1.000	1.000
		β_2	0.906	0.937	0.962	0.980
		σ_ε	0.912	0.948	0.965	0.985
3	400	β_1	0.731	0.9790	1.000	1.00
		β_2	0.966	0.982	0.991	0.997
		σ_ε	0.974	0.989	0.996	0.999

TABLE 6

Root-Mean Square Error of Parameter Estimators
in Truncated Regression of $\widehat{\delta}_i$, $\widehat{\widehat{\delta}}_i$ on $\boldsymbol{z}_i$

$p = q$	n	parameter	RMSE $\widehat{\delta}_i$	RMSE $\widehat{\widehat{\delta}}_i$
1	100	β_1	0.4985	0.5261
		β_2	0.1107	0.1118
		σ_ε	0.1481	0.1446
1	400	β_1	0.2259	0.2255
		β_2	0.0504	0.0497
		σ_ε	0.0687	0.0656
2	100	β_1	0.7812	0.9515
		β_2	0.1418	0.1498
		σ_ε	0.1925	0.1886
2	400	β_1	0.3121	0.3002
		β_2	0.0585	0.0563
		σ_ε	0.0811	0.0775
3	100	β_1	1.1762	1.8447
		β_2	0.1773	0.2288
		σ_ε	0.2292	0.2536
3	400	β_1	0.5071	0.6203
		β_2	0.0743	0.0795
		σ_ε	0.0960	0.1061
3	800	β_1	0.3558	0.3542
		β_2	0.0527	0.0498
		σ_ε	0.0646	0.0797

Figure 1
Estimates of Sampling Distributions, Algorithm #1
($p = q = 2$, $n = 400$)

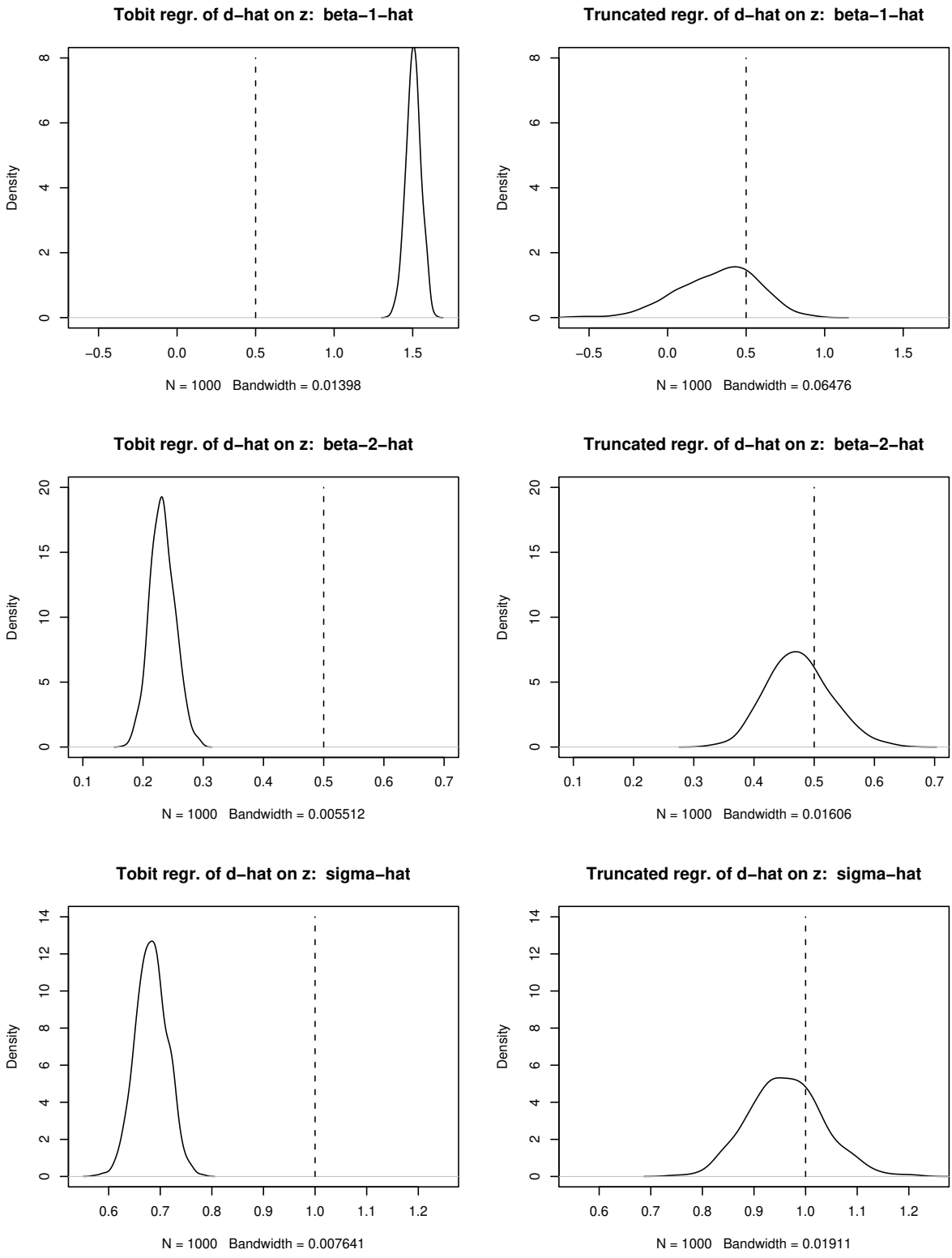
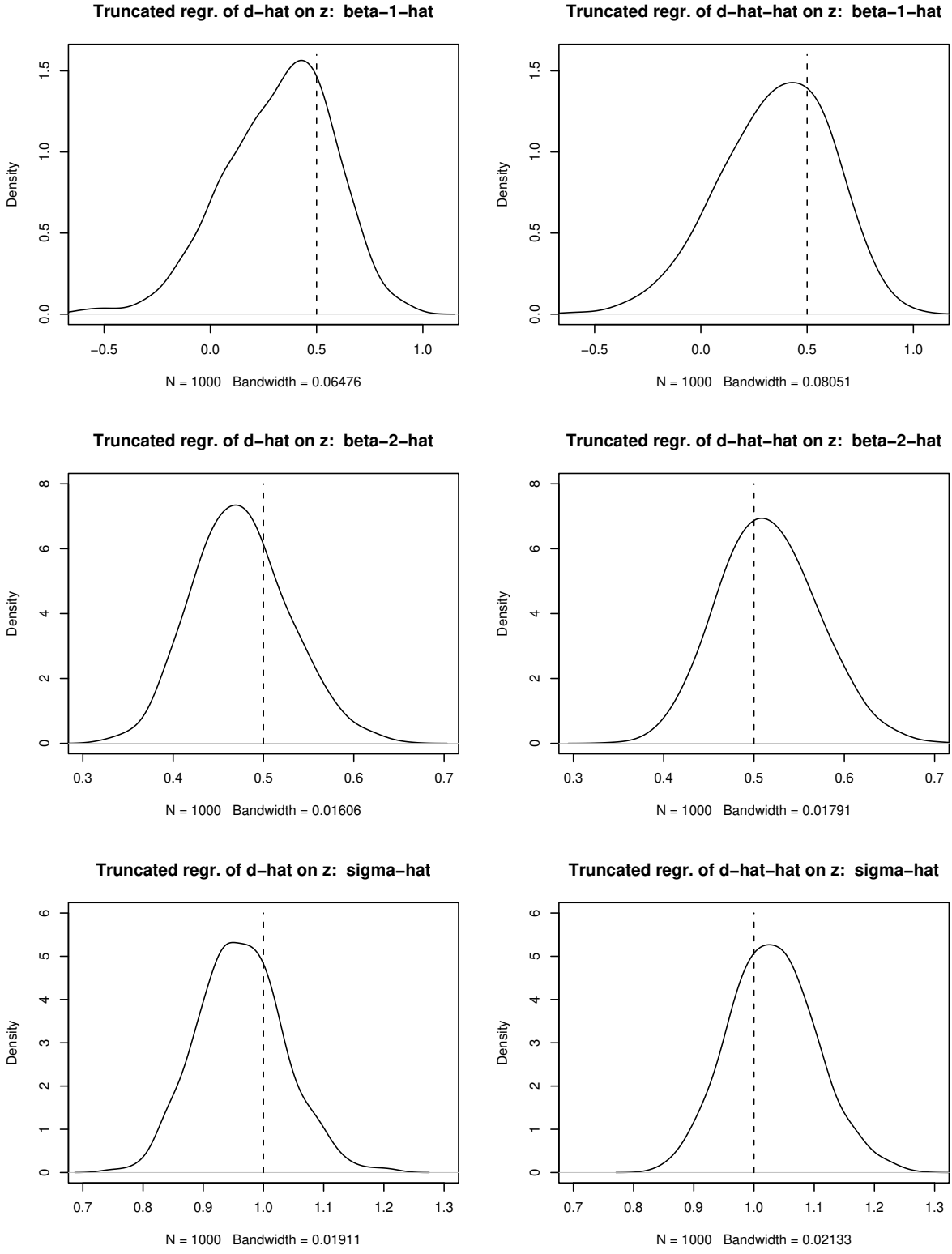



Figure 2
Estimates of Sampling Distributions, Algorithm #2
($p = q = 2$, $n = 400$)