Bonn Econ Discussion Papers

Discussion Paper 12/2006

Asymptotics and Consistent Bootstraps for DEA Estimators in Non-parametric Frontier Models

by

Alois Kneip, Leopold Simar, Paul W. Wilson

April 2006

Bonn Graduate School of Economics
Department of Economics
University of Bonn
Adenauerallee 24 - 42
D-53113 Bonn

The Bonn Graduate School of Economics is sponsored by the

Deutsche Post World Net

MAIL EXPRESS LOGISTICS FINANCE

Asymptotics and Consistent Bootstraps for DEA Estimators in Non-parametric Frontier Models

Alois Kneip Léopold Simar Paul W. Wilson*

April 2006

Abstract

Non-parametric data envelopment analysis (DEA) estimators based on linear programming methods have been widely applied in analyses of productive efficiency. The distributions of these estimators remain unknown except in the simple case of one input and one output, and previous bootstrap methods proposed for inference have not been proven consistent, making inference doubtful. This paper derives the asymptotic distribution of DEA estimators under variable returns-to-scale. This result is then used to prove that two different bootstrap procedures (one based on sub-sampling, the other based on smoothing) provide consistent inference. The smooth bootstrap requires smoothing the irregularly-bounded density of inputs and outputs as well as smoothing of the DEA frontier estimate. Both bootstrap procedures allow for dependence of the inefficiency process on output levels and the mix of inputs in the case of input-oriented measures, or on inputs levels and the mix of outputs in the case of output-oriented measures.

*Kneip: Institut für Gessellschafts- und Wirtschaftswissenschaften, Statistische Abteilung, Universität Bonn, Adenauerallee 24-26, 53113 Bonn, Germany; email kneip@uni-bonn.de. Simar: Institut de Statistique, Université Catholique de Louvain, Voie du Roman Pays 20, Louvain-la-Neuve, Belgium; email simar@stat.ucl.ac.be. Wilson: The John E. Walker Department of Economics, 222 Sirrine Hall, Clemson University, Clemson, South Carolina 29634–1309, USA; email wilson@eco.utexas.edu. Key Words: bootstrap, frontier, efficiency, data envelopment analysis, DEA. AMS Classification: Primary—62G09, 62G20; Secondary—62P20, 90B30, 91B38. *Journal of Economic Literature* Classification: C12, C14, C15.

1 Introduction

To date, non-parametric Data Envelopment Analysis (DEA) estimators have been discussed or applied in more than 1,800 articles published in more than 400 journals (see Gattoufi et al., 2004, for a comprehensive bibliography). DEA estimators are used to estimate various types of productive efficiency of firms in a wide variety of industries, as well as of governmental agencies, national economies, and other decision-making units. The estimators employ linear programming methods along the lines of Charnes et al. (1978, 1979) and Färe et al. (1985), and are based on the original ideas of Debreu (1951), Farrell (1957), and Shephard (1970).

DEA estimators measure efficiency relative to an ***estimate*** of an unobserved ***true*** frontier, conditional on observed data resulting from an underlying data-generating process (DGP). Although DEA estimators have been widely applied for more than 25 years, until recently, little was known about their statistical properties. It is now understood, however, that under certain assumptions the DEA ***frontier*** estimator is a consistent, maximum likelihood estimator (Banker, 1993), with a known rate of convergence (Korostelev et al., 1995). In addition, consistency and convergence rates of DEA ***efficiency*** estimators has been established (Kneip et al., 1998; see Simar and Wilson, 2000b, for a survey of recent developments regarding statistical properties of DEA estimators). Until now, however, the asymptotic distribution of DEA efficiency estimators has remained unknown except for the limited case of one input, one output derived by Gijbels et al. (1999); there have been no results that would allow one to perform classical inference regarding efficiency in more general (and more realistic) cases with multiple inputs and outputs. Moreover, the bootstrap methods proposed by Simar and Wilson (1998, 2000a) have been the only means for inferences about efficiency based on DEA estimators in a multivariate framework, but consistency for these procedures has not been proved.

This paper addresses these shortcomings by first deriving (in Theorem 2) the asymptotic distribution of DEA estimators under variable returns to scale, with arbitrary numbers of inputs and outputs. This is accomplished by characterizing DEA efficiency scores in a new way, and then localizing the problem in Theorem 1, which establishes that the DEA estimator for a given point is determined by observations in a small neighborhood of the projection of the given point onto the frontier. The asymptotic distribution derived in Theorem 2 is then used to prove that two different bootstrap methods yield consistent inference. The analysis that follows is a substantial departure from Gijbels et al. (1999), where the simple case of a single input and a single output allowed the frontier to be described as a functional relationship. In our general framework, the problem is more complicated due to the increased dimensionality of outputs as well as inputs, making it more difficult to characterize the frontier.

The first bootstrap method for which we prove consistency is based on sub-sampling, where bootstrap samples of size $m < n$ are drawn (independently, with replacement) from the empirical distribution of the n sample observations. That such an approach should work is not surprising; Swanepoel (1986) discussed this approach for inference about the boundary of support for a univariate distribution. The difficulty lies in the choice of m; our simulation results indicate that the choice of the subsample size is critical for obtaining confidence intervals with the desired coverage in finite samples. Unfortunately, there seems to be no reliable way of determining a reasonable value of m in applied settings. Experimentation with an iterated sub-sampling bootstrap has proved almost useless; for any realistic (original) sample size n, the inner bootstrap loops contain too few observations to provide useful information on the "optimal" size of m. Moreover, our simulations also suggest that suboptimal choices of m can be catastrophic for realized coverages of estimated confidence intervals.

The second bootstrap approach provides a tractable approach to inference with DEA estimators, but at a cost of increased complexity over the sub-sampling approach. Our second approach involves smoothing not only the distribution of the observations as proposed in Simar and Wilson (1998, 2000a), but also the initial estimate of the frontier itself. This necessitates choosing values for two smoothing parameters. One of these can be optimized using existing methods from kernel density estimation; in the second case, we provide a simple approach for selecting the bandwidth used to smooth the frontier estimate. We provide simulation results demonstrating that the method works well, provided the sample size n is sufficiently large for the given dimensionality of the problem (this caveat should of no surprise, since it is now well-known that the curse-of-dimensionality affects the quality of the initial DEA point-estimates; again, see Simar and Wilson, 2000b, for discussion).

The paper unfolds as follows. Section 2 defines notation and the statistical model, and briefly describe the DEA estimator. The local nature of the DEA estimator is described, and its asymptotic distribution is derived in Section 3, while results for the bootstrap procedures are proved in Section 4. Simulation results are presented in Section 5, and concluding remarks appear in the final section.

2 A Statistical Model for DEA Estimators

To establish notation for the rest of the paper, suppose that firms use input quantities $x \in \mathbb{R}_+^p$ to produce output quantities $y \in \mathbb{R}_+^q$. Standard microeconomic theory of the firm posits a production set

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

The production set Ψ is sometimes described in terms of its sections

$$\mathcal{Y}(x) \equiv \{y \mid (x, y) \in \Psi\} \tag{2.2}$$

and

$$\mathcal{X}(y) \equiv \{x \mid (x, y) \in \Psi\}, \tag{2.3}$$

which form the output feasibility and input requirement sets, respectively. Knowledge of either $\mathcal{Y}(x)$ for all x or $\mathcal{X}(y)$ for all y is equivalent to knowledge of Ψ; thus, both $\mathcal{Y}(x)$ and $\mathcal{X}(y)$ inherit the properties of Ψ. We denote the boundary of $\mathcal{X}(y)$ by

$$\mathcal{X}^{\partial}(y) = \{x \mid (x, y) \in \Psi, \ (\delta x, y) \notin \Psi \ \forall \ \delta < 1\} \tag{2.4}$$

Various economic assumptions regarding Ψ are possible; we adopt those of Shephard (1970) and Färe (1988):

Assumption 1. Ψ *is closed and strictly convex.*

Note that Assumption 1 implies that $\mathcal{Y}(x)$ is closed, strictly convex, and bounded for all $x \in \mathbb{R}_+^p$, and that $\mathcal{X}(y)$ is closed and strictly convex for all $y \in \mathbb{R}_+^q$. The boundary Ψ^{∂} of Ψ constitutes the ***technology***. Microeconomic theory of the firm suggests that in perfectly competitive markets, firms operating in the interior of Ψ will be driven from the market, but makes no prediction of how long this might take.

Assumption 2. $(x, y) \notin \Psi$ *if* $x = 0,\ y \geq 0, y \neq 0$, i.e., *all production requires use of some inputs.*

Assumption 3. *for* $\widetilde{x} \geq x$, $\widetilde{y} \leq y$, *if* $(x, y) \in \Psi$ *then* $(\widetilde{x}, y) \in \Psi$ *and* $(x, \widetilde{y}) \in \Psi$, 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 $\widetilde{x}$, $x \in \mathbb{R}_+^p$, $\widetilde{x} \geq x$ means that some number $\ell \in \{0,\ 1,\ \ldots,\ p\}$ of the corresponding elements of $\widetilde{x}$ and x are equal, while $(p - \ell)$ of the elements of $\widetilde{x}$ are greater than the corresponding elements of x. Note that Assumption 3 is equivalent to an assumption of monotonicity of the technology.

Various measures of technical efficiency are possible. We use the Farrell (1957) measure of input technical efficiency, defined by

$$\theta(x, y) \equiv \inf\{\delta \mid (\delta x, y) \in \Psi,\ \delta > 0\} \tag{2.5}$$

for an arbitrary point $(x, y) \in \mathbb{R}_+^{p+q}$. This is the the reciprocal of the Shephard (1970) input distance function. For $(x, y) \in \Psi$, $0 < \theta(x, y) \leq 1$. Note that θ provides a measure of Euclidean distance from the point $(x, y) \in \mathbb{R}_+^{p+q}$ to the boundary of Ψ in a direction parallel to the input axes and orthogonal to the output axes. One can also define output-oriented measures; we consider only the input orientation to conserve space. All of our results extend to output-oriented measures via straightforward, although perhaps tedious, changes in notation.

Of course, Ψ and hence $\theta(x, y)$ are unknown and must be estimated from a sample of observations $\mathcal{S}_n = \{(X_i, Y_i)\}_{i=1}^n$. The next three assumptions define a DGP; the framework here is similar to that in Simar (1996), Kneip et al. (1998), and Simar and Wilson (1998, 2000a).

Assumption 4. *The n observations in $\mathcal{S}_n$ are identically, independently distributed (iid) random variables on the convex attainable set Ψ.*

Assumption 5. *(a) The (X, Y) possess a joint density f with support $\mathcal{D} \subset \Psi$; (b) f is continuous on $\mathcal{D}$; and (c) $f(\theta(x, y)x, y) > 0$ for all $(x, y) \in \mathcal{D}$.*

Clearly, Assumption 5(c) imposes a discontinuity in f at frontier points where $\theta(x, y) = 1$, ensuring a significant, non-negligible probability of observing production units close to the production frontier. For technically non-attainable points which lie outside Ψ, $f \equiv 0$.

Assumption 6. *For (x, y) in the interior of $\mathcal{D}$, the function $\theta(v, w)$ is twice continuously differentiable for all (v, w) in a sufficiently small neighborhood of (x, y).*

Assumptions 1–6 describe the statistical model. In the analysis that follows, we concentrate on a fixed point $(x, y) \in \Psi$; interest lies in making inference about the distance measure $\theta(x, y)$.

The DEA estimator of Ψ is merely the convex hull of the free disposal hull of $\mathcal{S}_n$ given by

$$\widehat{\Psi} = \left\{(x, y) \mid y \leq Yq,\ x \geq Xq,\ i'q = 1,\ q \in \mathbb{R}_+^n\right\}, \tag{2.6}$$

where $Y = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}$, $X = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}$, i denotes an $(n \times 1)$ vector of ones, and q is an $(n \times 1)$ vector of intensity variables. The corresponding DEA estimator of $\theta(x, y)$ is obtained by replacing Ψ with $\widehat{\Psi}$ in (2.5); *i.e.,*

$$\widehat{\theta}(x, y) = \min\left\{\delta > 0 \mid y \leq Yq,\ \delta x \geq Xq,\ i'q = 1,\ q \in \mathbb{R}_+^n\right\}. \tag{2.7}$$

Minimization of the linear program in (2.7) provides a solution for both δ and q. Whereas $\theta(x, y)$ defined in (2.5) gives a measure of distance from a point $(x, y) \in \mathbb{R}_+^{p+q}$ to the boundary of Ψ, $\widehat{\theta}(x, y)$ measures distance from the same point to the boundary of the convex hull of the free-disposal hull of the n sample observations. The statistical performance of the DEA estimator $\widehat{\theta}(x, y)$ of $\theta(x, y)$ depends on the smoothness of the frontier. Kneip et al. (1998) derive different rates of convergence depending of the degree of smoothness. Per Assumption 6 above, we consider only the case where $\theta(x, y)$ is twice-differentiable. For this case, Kneip et al. (1998) prove that $\widehat{\theta}(x, y) = \theta(x, y) + O_p(n^{\frac{2}{p+q+1}})$; As with many non-parametric estimators, DEA estimators suffer from the curse of dimensionality.

3 Asymptotic Distribution of DEA Estimators

In this section we derive the (previously unknown) asymptotic distribution of DEA estimators for the general case with arbitrary numbers of inputs p and outputs q. Along the way, Theorem 1 characterizes the "local" nature of the estimation problem. Theorem 2 establishes the asymptotic distribution as well as its continuity. Continuity is needed to prove consistency of the bootstrap methods that are given in Section 4 below. The analysis in this section re-characterizes the problem by defining a new coordinate system. This in turn allows the efficient frontier to be described by a function; the efficiency score $\theta(x,y)$ can then be related to a particular value of this function.

To begin, consider a decomposition of the vectors X_i of inputs that is specific for the particular point of interest, x. Let $\mathcal{V}(x)$ denote the $(p-1)$-dimensional linear space of all vectors $z \in \mathbb{R}^p$ such that $z^T x = 0$. Any input vector X_i adopts a unique decomposition of the form

$$X_i = \gamma_i \frac{x}{||x||} + Z_i \quad \text{for some } Z_i \in \mathcal{V}(x) \text{ and } \gamma_i = \frac{x^T X_i}{||x||}, \tag{3.1}$$

where $||\cdot||$ denotes the Euclidean norm. In this new coordinate system (z,y), the attainable set Ψ can be re-expressed as

$$\Psi^*(x) = \left\{ (z,y) \in \mathcal{V}(x) \times \mathbb{R}_+^q \mid \left(\gamma \frac{x}{||x||} + z, y \right) \in \Psi \text{ for some } \gamma > 0 \right\}. \tag{3.2}$$

Note that the point of interest $(x,y) \in \Psi$ has coordinates $(0,y)$ in $\Psi^*(x)$. In addition, the boundary of Ψ can be described through the following function defined for any $(z,y) \in \Psi^*(x)$:

$$g_x(z,y) = \inf \left\{ \gamma \mid \left(\gamma \frac{x}{||x||} + z, y \right) \in \Psi \right\}. \tag{3.3}$$

The quantity of interest $\theta(x,y)$ can be expressed as

$$\theta(x,y) = \frac{g_x(0,y)}{||x||}. \tag{3.4}$$

Moreover, the DEA estimator of the frontier and of $\theta(x,y)$ can be similarly transformed by writing

$$\widehat{g}_x(z,y) = \inf \left\{ \gamma \mid \left(\gamma \frac{x}{||x||} + z, y \right) \in \widehat{\Psi} \right\} \tag{3.5}$$

and

$$\widehat{\theta}(x,y) = \frac{\widehat{g}_x(0,y)}{||x||}. \tag{3.6}$$

Finally, with only a small abuse of notation, one may extend the definition of g_x to all (v,y) with $\left(v - \frac{x^T v}{||x||^2} x, y \right) \in \Psi^*(x)$ by taking $g_x(v,y) = g_x\left(v - \frac{x^T v}{||x||^2} x, y \right)$.

In the case of one input $(p = 1)$, the function g_x is simply the "frontier function" and does not depend on x. Then $\mathcal{V} = \{0\}$ and $g_x(0,y) \equiv g(y) = \theta(x,y)x$ for all x.

We are interested only in analyzing $g_x(z,y)$ as a function of z and y. However, we have adopted the notation g_x to emphasize that for $p > 1$, the structure of this function depends on the vector $\frac{x}{||x||}$. Note that whenever (x,y) lies in the interior of Ψ, $(z,y) \in \Psi^*(x)\ \forall\ z \in \mathcal{V}(x)$.

Figure 1 illustrates the definition of g_x for the case $p = 2$. For a given output vector y, the input requirement set $\mathcal{X}(y)$ is a convex subset of $\mathbb{R}^2_+$ with efficiency boundary $\mathcal{X}^\partial(y)$, shown by the solid black line. We now consider an input vector x with $||x|| = 1$. The ray γx, $\gamma \geq 0$, is represented by the solid gray line passing through the origin. For a vector z with $z^T x = 0$, the dashed gray line $\gamma x + z$ is parallel to γx. The intersection between $\gamma x + z$ and $\mathcal{X}^\partial(y)$ then determines the point $g_x(z, y)x + z$.

The following lemma summarizes the most important properties of g_x.

Lemma 1. *By Assumption 1,*

(a) g_x is convex, and for all $(v, \widetilde{y}) \in \Psi$ and $z = v - \frac{x^T v}{||x||^2} x$,

$$\theta(v, \widetilde{y}) \frac{x^T v}{||x||} = g_x(\theta(v, \widetilde{y}) z, \widetilde{y}) \quad \text{and} \quad \widehat{\theta}(v, \widetilde{y}) \frac{x^T v}{||x||} = \widehat{g}_x(\widehat{\theta}(v, \widetilde{y}) z, \widetilde{y}).$$

(b) Let (x, y) be in the interior of $\mathcal{D}$. By Assumption 6,

- *the function $g_x(\cdot, \cdot)$ is twice continuously differentiable for all points in a sufficiently small neighborhood of $(0, y)$;*
- *The matrix $g''_x(0, y)$ of second derivatives at $(0, y)$ is positive semidefinite, and there exists a constant $c_0 > 0$ such that $w^T g''_x(0, y) w \geq c_0 \; \forall \; w \in \mathcal{V}(x) \times \mathbb{R}^q$ with $||w|| = 1$.*

Proof. For all (z_1, y_1), $(z_2, y_2) \in \Psi^*(x)$ and every $\alpha \in [0, 1]$, the definition of g_x implies that $[\alpha g_x(z_1, y_1) + (1 - \alpha) g_x(z_2, y_2)] \frac{x}{||x||} + \widetilde{z}_\alpha \geq g_x(\widetilde{z}_\alpha, \widetilde{y}_\alpha) \frac{x}{||x||} + \widetilde{z}_\alpha$ with $(\widetilde{z}_\alpha, \widetilde{y}_\alpha) = (\alpha z_1 + (1 - \alpha) z_2, \alpha y_1 + (1 - \alpha) y_2) \in \Psi^*(x)$. Consequently, g_x is a convex function. Moreover, for any $v \in \mathcal{X}^\partial(\widetilde{y})$ we necessarily have $v = g_x(z, \widetilde{y}) \frac{x}{||x||} + z$ for $z = v - \frac{x^T v}{||x||^2} x$. Assertion (a) then follows from $\theta(v, \widetilde{y}) v \in \mathcal{X}^\partial(\widetilde{y})$. In view of Assumption 6 twice-differentiability of g_x at $(0, y)$ follows directly.

Assumption 1 implies that

$$\begin{aligned} 1 &= \alpha \theta(g_x(z_1, y_1) \frac{x}{||x||} + z_1, y_1) + (1 - \alpha) \theta(g_x(z_2, y_2) \frac{x}{||x||} + z_2, y_2) \\ &> \theta\left((\alpha g_x(z_1, y_1) + (1 - \alpha) g_x(z_2, y_2)) \frac{x}{||x||}, y \right) \end{aligned}$$

holds for all $(z_1, y_1), (z_2, y_2) \in \Psi^*(x)$, $(z_1, y_1) \neq (z_2, y_2)$ and every $\alpha \in [0, 1]$ with $\alpha z_1 + (1-\alpha) z_2 = 0$ and $\alpha y_1 + (1-\alpha) y_2 = y$. Since $\theta(g_x(0, y)\frac{x}{||x||}, y) = 1$, we can conclude that $\alpha g_x(z_1, y_1) + (1-\alpha) g_x(z_2, y_2) > g_x(0, y)$, which leads to the asserted structure of g_x''. ■

As noted earlier, Kneip et al. (1998) showed that the rate of convergence of the input inefficiency estimator is $O_p(n^{-2/(p+q+1)})$. The following lemma shows that the problem of specifying the distribution of $\frac{\widehat{\theta}(x,y)}{\theta(x,y)}$ can be reformulated in terms of g_x and of the distribution of $\theta(X_i, Y_i)$, Z_i and Y_i.

Lemma 2. *Let (x, y) be in the interior of $\mathcal{D}$. Under Assumptions 1–6 we obtain for any $\delta > 0$*

$$\text{Prob}\left(\frac{\widehat{\theta}(x,y)}{\theta(x,y)} - 1 \leq \delta n^{-\frac{2}{p+q+1}}\right) = \text{Prob}(A[\delta, n]), \tag{3.7}$$

where $A[\delta, n]$ denotes the following event: There exist some $\alpha_1 \geq 0, \ \ldots, \ \alpha_n \geq 0$ with $\sum_{j=1}^n \alpha_j = 1$ such that

$$\sum_{i=1}^n \alpha_i Z_i = 0, \quad \textit{and} \quad \sum_{i=1}^n \alpha_i Y_i = y \tag{3.8}$$

and

$$\sum_{i=1}^n \alpha_i \frac{g_x(\theta_i Z_i, Y_i)}{\theta_i g_x(0, y)} - 1 \leq \delta n^{-\frac{2}{p+q+1}},$$

where $\theta_i = \theta(X_i, Y_i)$ and $Z_i = X_i - \frac{x^T X_i}{||x||^2} x$.

Proof. By definition of a DEA frontier we have $\frac{\widehat{\theta}(x,y)}{\theta(x,y)} - 1 \leq \delta n^{-\frac{2}{p+q+1}}$ if and only if there exists a $\beta > 0$ with $\frac{\beta}{\theta(x,y)} - 1 \leq \delta n^{-\frac{2}{p+q+1}}$ such that

$$\sum_{i=1}^k \alpha_i Y_i = y, \quad \text{and} \quad \sum_{i=1}^k \alpha_i X_i = \beta x \tag{3.9}$$

hold for some $\alpha_1 \geq 0, \ \ldots, \ \alpha_n \geq 0$ with $\sum_{j=1}^n \alpha_j = 1$. The relations in (3.1) and Lemma 1(a) imply $X_i = \frac{g_x(\theta_i Z_i, Y_i)}{\theta_i ||x||} x + Z_i$. Since all Z_i are orthogonal to x, (3.9) holds if and only if (3.8) is satisfied and $\sum_{i=1}^n \alpha_i \frac{g_x(\theta_i Z_i, Y_i)}{\theta_i ||x||} = \beta$. The lemma now follows from $g_x(0, y) = ||x|| \theta(x, y)$. ■

Now consider an orthonormal basis $z^{(1)}, \ \ldots, \ z^{(p-1)}$ of $\mathcal{V}(x)$. Every vector $Z_i \in \mathcal{V}(x)$ can be expressed in the form $Z_i = \sum_{j=1}^{p-1} \zeta_{ij} z^{(j)}$. Let $\zeta_i = (\zeta_{i1}, \ \ldots, \ \zeta_{i,p-1})$. Since $\theta_i = \theta(X_i, Y_i)$ and $Z_i = X_i - \frac{x^T X_i}{||x||^2} x$ are smooth functions of (X_i, Y_i), Assumption 5 implies that (θ_i, ζ_i, Y_i)

has a density $\bar{f}_x$ on $[0,1] \times \mathbb{R}^{p-1} \times \mathbb{R}^q_+$. Let $\bar{\mathcal{D}}$ denote the support of this density. By Assumption 5(a)–(c), it is easily seen that $\bar{f}_x(\cdot,\cdot,\cdot)$ is continuous on $(0,1) \times \mathbb{R}^{p-1} \times \mathbb{R}^q_+$, and $\bar{f}_x(1,0,y) > 0$.

Theorem 1, given below, plays an important role by "localizing" the frontier problem. The value of $\hat{\theta}(x,y)$ is essentially determined by those observations which fall into a small neighborhood of $(\theta(x,y)x, y)$. Note that for the proof of the theorem, Assumption 1 is crucial. The theorem does not apply if, for example, the frontier is linear or conical, since in such cases $\hat{\theta}(x,y)$ may be determined by points very far from the point of interest (x,y).

Before proceeding, some additional notation is needed. Note that the sample of observations $\mathcal{S}_n$ can be represented equivalently be the corresponding samples $\widetilde{\mathcal{S}}_n = \{(\theta_i, Z_i, Y_i)\}_{i=1}^n$ or $\bar{\mathcal{S}}_n = \{(\theta_i, \zeta_i, Y_i)\}_{i=1}^n$, where ζ_i is determined by $Z_i = \sum_{j=1}^{p-1} \zeta_{ij} z^{(j)}$. Next, define a set $C(x,y;h)$ by

$$\begin{aligned} C(x,y;h) \;=\; \Big\{ (\theta, \widetilde{z}, \widetilde{y}) \in (0,1) \times \Psi^*(x) \,\Big|\, & 1-\theta \le h^2, \\ & z = \sum_j \zeta_j z^{(j)} \text{ with } |\zeta_j| \le h \;\forall\, j = 1, \ldots, p-1, \\ & |y_r - \widetilde{y}_r| \le h \;\forall\, r = 1, \ldots, q \Big\}. \end{aligned}$$

The point $(1,0,y)$ in the transformed space $\{(\theta(v,\widetilde{y}), v - (x^T v/||x||^2)x, \widetilde{y}) \mid (v,\widetilde{y}) \in \Psi\}$ corresponds to the boundary point $(\theta(x,y)x, y)$ in the original space Ψ. The set $C(x,y;h)$ is a neighborhood of the transformed boundary point (1,0,y). Then let $A[\delta, n; h]$ denote the following event: for some $k \le n$ and $i_1, \ldots, i_k \in \{1, \ldots, n\}$, there exist some $(X_{i_1}, Y_{i_1}), \ldots, (X_{i_k}, Y_{i_k})$ with $(\theta_{i_1}, Z_{i_1}, Y_{i_1}), \ldots, (\theta_{i_k}, Z_{i_k}, Y_{i_k}) \in \widetilde{\mathcal{S}}_n \cap C(x,y; h \cdot n^{-\frac{1}{p+q+1}})$, as well as some $\alpha_1 \ge 0, \ldots, \alpha_k \ge 0$ with $\sum_{j=1}^k \alpha_j = 1$ such that $\sum_{j=1}^k \alpha_j Y_{i_j} = y$, $\sum_{j=1}^k \alpha_j Z_{i_j} = 0$, and

$$\sum_{j=1}^k \alpha_j \frac{g_x(\theta_{i_j} Z_{i_j}, Y_{i_j})}{\theta_{i_j} g_x(0,y)} - 1 \le \delta n^{-\frac{2}{p+q+1}}. \tag{3.10}$$

Again, $\theta_{i_j} = \theta(X_{i_j}, Y_{i_j})$ and $Z_{i_j} = X_{i_j} - \frac{x^T X_{i_j}}{||x||^2} x$.

Theorem 1. *Let (x,y) be in the interior of $\mathcal{D}$. Then under Assumptions 1–6,*

(a) for any $\epsilon > 0$ there exists an $h_\epsilon < \infty$ such that for all $h \ge h_\epsilon$, every $\delta > 0$ and all

sufficiently large n,

$$|\text{Prob}(A[\delta, n] - \text{Prob}(A[\delta, n; h])| \leq \epsilon; \tag{3.11}$$

(b) there exists an open neighborhood $N(x, y)$ of (x, y) such that

$$\text{Prob}\left(\sup_{(\widetilde{x},\widetilde{y})\in N(x,y)} \left|\frac{\widehat{\theta}(\widetilde{x}, \widetilde{y})}{\theta(\widetilde{x}, \widetilde{y})} - 1\right| \leq n^{-\frac{2}{p+q+1}} \log n\right) \to 1 \quad \textit{as } n \to \infty$$

and

$$\text{Prob}\left(\sup_{(\widetilde{x},\widetilde{y})\in N(x,y)} \left|\frac{\widehat{g}_x(\widetilde{x} - \frac{x^T\widetilde{x}}{||x||^2}x, \widetilde{y})}{g_x(\widetilde{x} - \frac{x^T\widetilde{x}}{||x||^2}x, \widetilde{y})} - 1\right| \leq n^{-\frac{2}{p+q+1}} \log n\right) \to 1 \quad \textit{as } n \to \infty.$$

A proof is given in the appendix.

In order to examine the probabilities $P(A[\delta, n; h])$, still more notation is required. Let $(\widetilde{\vartheta}_1, \widetilde{\zeta}_1, \widetilde{y}_1), (\widetilde{\vartheta}_2, \widetilde{\zeta}_2, \widetilde{y}_2), \ldots$ denote a sequence of iid random variables uniformly distributed on $[0, 1] \times [-1, 1]^{p-1} \times [-1, 1]^q$. For $k \in \mathbb{N}$, let $U[\gamma, k]$ denote the following event: there exist some $\alpha_1 \geq 0, \ldots, \alpha_k \geq 0$ with $\sum_{j=1}^k \alpha_j = 1$ such that

$$\sum_{j=1}^k \alpha_j \widetilde{y}_j = 0 \quad \text{and} \quad \sum_{j=1}^k \alpha_j \widetilde{z}^{(j)} = 0, \tag{3.12}$$

where $\widetilde{z}_j = \sum_{r=1}^{p-1} \widetilde{\zeta}_{jr} z^{(r)}$, and

$$\sum_{j=1}^k \alpha_j \frac{1}{2g_x(0, y)} \left[\widetilde{z}_j^T g''_{x;zz}(0, y)\widetilde{z}_j \;+\; 2\widetilde{z}_j^T g''_{x;zy}(0, y)\widetilde{y}_j \;+\; \widetilde{y}_j^T g''_{x;yy}(0, y)\widetilde{y}_j\right] + \sum_{j=1}^k \alpha_j \widetilde{\vartheta}_j \leq \gamma. \tag{3.13}$$

Here we use

$$g''(x; 0, y) = \begin{bmatrix} g''_{x;zz}(0, y) & g''_{x;zy}(0, y)^T \\ g''_{x;zy}(0, y) & g''_{x;yy}(0, y) \end{bmatrix}$$

to denote the matrix of second derivatives of g_x at $(0, y)$. Finally, let $\tau(h) = 2^{(p+q-1)} h^{(p+q+1)}$.

Proposition 1. *Under the conditions of Theorem 1,*

$$\left|\text{Prob}(A[\delta, n; h]) - \sum_{k=1}^\infty \text{Prob}\left(U\left[\frac{\delta}{h^2}, k\right]\right) \frac{\tau(h)^k \bar{f}_x(1, 0, y)^k}{k!} e^{-\tau(h)\bar{f}_x(1,0,y)}\right| \to 0 \tag{3.14}$$

as $n \to \infty$ for any $h > 0$.

Proof. Recall the definition of $A[\delta, n; h]$. Since $Z_{i_j} = O_p(n^{-\frac{1}{p+q+1}})$, $|y - Y_{i_j}| = O_p(n^{-\frac{1}{p+q+1}})$ and $1 - \theta_{i_j} = O_p(n^{-\frac{2}{p+q+1}})$, Taylor expansions of g_x yield

$$\begin{aligned}
\sum_{j=1}^{k} \alpha_j \frac{g_x(\theta_{i_j} Z_{i_j}, Y_{i_j})}{\theta_{i_j} g_x(0,y)} - 1 &= \sum_{j=1}^{k} \alpha_j \frac{g_x(\theta_{i_j} Z_{i_j}, Y_{i_j}) - g_x(0,y)}{g_x(0,y)} + \sum_{j=1}^{k} \alpha_j (1 - \theta_{i_j}) + o_p(n^{-\frac{2}{p+q+1}}) \\
&= \sum_{j=1}^{k} \alpha_j \frac{1}{2 g_x(0,y)} \Big[Z_{i_j}^T g''_{x;zz}(0,y) Z_{i_j} + 2 Z_{i_j}^T g''_{x;zy}(0,y)(Y_{i_j} - y) \\
&\qquad + (Y_{i_j} - y)^T g''_{x;yy}(0,y)(Y_{i_j} - y) \Big] \\
&\qquad + \sum_{j=1}^{k} \alpha_j (1 - \theta_{i_j}) + o_p(n^{-\frac{2}{p+q+1}})
\end{aligned} \tag{3.15}$$

where the convergence is uniform for all possible $(X_{i_j}, Y_{i_j}) \in C(x, y; hn^{-\frac{1}{p+q+1}})$. Note that necessarily $\sum_{j=1}^{k} \alpha_j \left[g_{x;z}(0,y)' \cdot Z_{i_j} + g'_{x;y}(0,y) \cdot (Y_{i_j} - y) \right] = 0$, where $g'_x(0,y) = (g_{x;z}(0,y)', g_{x;y}(0,y)')^T$ denotes the vector of first derivatives of g_x at $(0, y)$.

The density $\bar{f}_x$ is continuous at $(1, 0, y)$. Hence, the probability that there is an observation in $C(x, y; h \cdot n^{-\frac{1}{p+q+1}})$ is asymptotically equivalent to $\tau(h) \bar{f}_x(1, 0, y) \cdot n^{-1}$. Thus for large n, the distribution of the number k of points in $C(x, y; h \cdot n^{-\frac{1}{p+q+1}})$ follows approximately a Poisson distribution with parameter $\tau(h) \bar{f}_x(1, 0, y)$. Continuity of the densities implies that the conditional distribution of (θ_i, ζ_i, Y_i), given $(\theta_i, Z_i, Y_i) \in C(x, y; h \cdot n^{-\frac{1}{p+q+1}})$, is uniform on $\bar{C}(h \cdot n^{-\frac{1}{p+q+1}}) := [1 - h^2 n^{-\frac{2}{p+q+1}}, 1] \times [-hn^{-\frac{1}{p+q+1}}, hn^{-\frac{1}{p+q+1}}]^{p-1} \times [y_1 - hn^{-\frac{1}{p+q+1}}, y_1 + hn^{-\frac{1}{p+q+1}}] \times \cdots \times [y_q - hn^{-\frac{1}{p+q+1}}, y_q + hn^{-\frac{1}{p+q+1}}]$. Combining these arguments with (3.12) reveals that

$$\left| \text{Prob}(A[\delta, n; h] - \sum_{k=1}^{\infty} \text{Prob}(\bar{A}[\delta, n; h; k]) \frac{\tau(h)^k \bar{f}_x(1,0,y)^k}{k!} e^{-\tau(h) \bar{f}_x(1,0,y)} \right| \to 0$$

as $n \to \infty$, where for a sequence $(\widetilde{\theta}_{1,n}, \widetilde{\zeta}_{1,n}, \widetilde{Y}_{1,n})$, $\ldots$, $(\widetilde{\theta}_{k,n}, \widetilde{\zeta}_{k,n}, \widetilde{Y}_{k,n})$ of k iid random variables uniformly distributed on $\bar{C}(h \cdot n^{-\frac{1}{p+q+1}})$, we use $\bar{A}[\delta, n; h; k]$ to describe the following event: there exist some $\alpha_1 \geq 0$, $\ldots$, $\alpha_k \geq 0$ with $\sum_{j=1}^{k} \alpha_j = 1$ such that $\sum_{j=1}^{k} \alpha_j \widetilde{Y}_{j,n} = y$ and

$$\sum_{j=1}^{k} \alpha_j \widetilde{Z}_{j,n} = 0 \text{ for } \widetilde{Z}_{j,n} = \sum_{r=1}^{p-1} \zeta_{j,n,r} z^{(r)} \tag{3.16}$$

and

$$\sum_{j=1}^{k} \alpha_j \frac{1}{2g_x(0,y)} \Big[\widetilde{Z}_{j,n}^T g''_{x;zz}(0,y)\widetilde{Z}_{j,n} + 2\widetilde{Z}_{j,n}^T g''_{x;zy}(0,y)(\widetilde{Y}_{j,n} - y) + (\widetilde{Y}_{j,n} - y)^T g''_{x;yy}(0,y)(\widetilde{Y}_{j,n} - y)\Big] + \sum_{j=1}^{k} \alpha_j (1 - \widetilde{\theta}_{j,n}) \leq \delta \cdot n^{-\frac{2}{p+q+1}}. \tag{3.17}$$

The assertion of the proposition now follows from the fact that $\bar{A}[\delta, n; h; k]$ is realized iff the event $U[\frac{\delta}{h^2}, k]$ is realized for $\widetilde{\vartheta}_j = \frac{1}{h^2 n^{-\frac{2}{p+q+1}}}(1 - \widetilde{\theta}_{j,n})$, $\widetilde{\zeta}_j = \frac{1}{hn^{-\frac{1}{p+q+1}}}\widetilde{\zeta}_{j,n}$ and $\widetilde{y}_j = \frac{1}{hn^{-\frac{1}{p+q+1}}}(\widetilde{Y}_{j,n} - y)$. It then follows that uniformity of $(\widetilde{\theta}_{j,n}, \widetilde{\zeta}_{j,n}, \widetilde{Y}_{j,n})$ on $\bar{C}(h \cdot n^{-\frac{1}{p+q+1}})$ is equivalent to uniformity of $(\widetilde{\vartheta}_j, \widetilde{\zeta}_j, \widetilde{y}_j)$ on $[0,1] \times [-1,1]^{p-1} \times [-1,1]^q$, and that (3.13) corresponds to (3.12). Finally, (3.17) implies (3.13) holds when γ is replaced by δ/h^2. ■

We are now ready to state a theorem about the asymptotic distribution of $n^{\frac{2}{p+q+1}}(\frac{\widehat{\theta}(x,y)}{\theta(x,y)} - 1)$.

Theorem 2. *Under the conditions of Theorem 1, let*

$$F_x(\delta) = \lim_{k\to\infty} \text{Prob}\left(U\left[\delta \frac{\bar{f}_x(1,0,y)^{2/(p+q+1)}}{k^{2/(p+q+1)}}, k\right]\right) \tag{3.18}$$

for $-\infty < \delta < \infty$. *Then* F_x *is a continuous distribution function with* $F_x(0) = 0$, $0 \leq F_x(\delta) < 1$, *and*

$$\begin{aligned} F_x(\delta) &= \lim_{n\to\infty} \text{Prob}\left[n^{\frac{2}{p+q+1}}\left(\frac{\widehat{\theta}(x,y)}{\theta(x,y)} - 1\right) \leq \delta\right] = \lim_{n\to\infty} \text{Prob}(A[\delta, n]) \\ &= \lim_{h\to\infty} \sum_{k=1}^{\infty} \text{Prob}\left(U\left[\frac{\delta}{h^2}, k\right]\right) \frac{\tau(h)^k \bar{f}_x(1,0,y)^k}{k!} e^{-\tau(h)\bar{f}_x(1,0,y)}. \end{aligned}$$

A proof is given in the appendix.

Although the asymptotic distribution in Theorem 2 possesses a non-standard structure, it nevertheless is a well-defined, continuous probability distribution. Recalling the definition of the event $U(\cdot,\cdot)$, it is clear that the shape of the distribution function F_x is determined by $\frac{(p+q)(p+q+1)}{2} + 2$ parameters determined by (i) the value $\bar{f}_x(1,0,y)$ of the density $\bar{f}_x$, (ii) the value $g_x(0,y)$ of the function g_x at the corresponding frontier point, and (iii) the matrix

$g_x''(0, y)$ of second derivatives of g_x at $(0, y)$. If these parameters were known, quantiles of the asymptotic distribution could be estimated by Monte Carlo simulations. Unfortunately, however, obtaining reliable estimates of the matrix $g_x''(0, y)$ necessary for this approach to work well seems particularly difficult. Fortunately, the bootstrap, when bootstrap samples are drawn appropriately, provides a way out of this difficulty.

4 Bootstrapping DEA Estimators

Two bootstrap methods are presented in this section, and their consistency for inference-making purposes are established in Theorems 3 and 4 using the results from Section 3. The first bootstrap method is, in principle, easy to apply, but depends critically on a tuning parameter for which to date no reliable method exists for choosing its value. The second method depends on two tuning parameters for which we offer data-based methods for selecting values in real-world applications.

As in Section 3, we consider a fixed point (x, y) in the interior of $\mathcal{D}$ satisfying Assumption 6. In this section, we consider suitable bootstrap procedures for estimating confidence intervals for $\theta(x, y)$.

The simplest bootstrap would, on each replication, take n independent draws from the empirical distribution of the observations in $\mathcal{S}_n$ to construct a pseudo-sample $\mathcal{S}_n^*$, and then apply (2.7) to obtain a bootstrap estimate $\widehat{\theta}^*(x, y)$ (note that $\widehat{\theta}^*(x, y)$ measures distance from the original point of interest, (x, y), to the boundary of the convex hull of the free-disposal hull of the pseudo-observations in $\mathcal{S}_n^*$). However, this naive bootstrap does not provide consistent inference as discussed by Simar and Wilson (1999b, 1999a). From Theorem 1 it is clear that as $n \to \infty$, the distribution of $n^{\frac{2}{p+q+1}} \left(\frac{\hat{\theta}^*}{\widehat{\theta}} - 1 \right)$ does not tend to the true distribution F. The empirical distribution of (θ_i, Z_i, Y_i) does not converge sufficiently fast to mimic the true probabilities on the sets $C(x, y; hn^{-\frac{1}{p+q+1}})$ which are proportional to $\frac{1}{n}$. This result is not surprising; it is well-known that the naive bootstrap does not work in the case of estimating the boundary of support for a univariate distribution (*e.g.,* see Bickel and Freedman, 1981).

We consider two different bootstrap approaches; the first is based on sub-sampling, while the second is based on smoothing.

4.1 Bootstrap with Sub-sampling

Let $m = n^\kappa$ for some $\kappa \in (0,1)$, and consider the following bootstrap scheme:

Algorithm #1:

[1] Generate a bootstrap sample $\mathcal{S}_m^* = \{(X_i^*, Y_i^*)\}_{i=1}^m$ by randomly drawing (independently, uniformly, and with replacement) m observations from the original sample, $\mathcal{S}_n$.

[2] Apply the DEA estimator in (2.7) to construct bootstrap estimates $\widehat{\theta}^*(x,y)$.

[3] Repeat steps [1]–[2] B times; use the resulting bootstrap values to approximate the conditional distribution of $m^{\frac{2}{p+q+1}}(\frac{\hat{\theta}^*(x,y)}{\hat{\theta}(x,y)} - 1)$ given $\mathcal{S}_n$, and use this approximation to approximate the unknown distribution of $n^{\frac{2}{p+q+1}}(\frac{\hat{\theta}(x,y)}{\theta(x,y)} - 1)$. For a given $\alpha \in (0,1)$, use the bootstrap values to estimate the quantiles $\delta_{\alpha/2,m}$, $\delta_{1-\alpha/2,m}$ where

$$\begin{aligned}
\text{Prob}\left[m^{\frac{2}{p+q+1}}\left(\frac{\hat{\theta}^*(x,y)}{\widehat{\theta}(x,y)} - 1\right) \leq \delta_{\alpha/2,m} \mid \mathcal{S}_n\right] &= \frac{\alpha}{2}, \\
\text{Prob}\left[m^{\frac{2}{p+q+1}}\left(\frac{\hat{\theta}^*(x,y)}{\widehat{\theta}(x,y)} - 1\right) \leq \delta_{1-\alpha/2,m} \mid \mathcal{S}_n\right] &= 1 - \frac{\alpha}{2}.
\end{aligned}$$

[4] Compute $\left[\frac{\widehat{\theta}(x,y)}{1+n^{-\frac{2}{p+q+1}}\delta_{1-\alpha/2,m}}, \frac{\widehat{\theta}(x,y)}{1+n^{-\frac{2}{p+q+1}}\delta_{\alpha/2,m}}\right]$, a symmetric $1-\alpha$ confidence interval estimate for $\theta(x,y)$.

Consistency of this bootstrap is easily demonstrated by the following theorem.

Theorem 3. *Under the conditions of Theorem 1, let $m \equiv m(n) = n^\kappa$ for some $\kappa \in (0,1)$. Then*

$$\sup_{\delta>0}\left|F_x(\delta) - \text{Prob}\left[m^{\frac{2}{p+q+1}}\left(\frac{\widehat{\theta}^*(x,y)}{\widehat{\theta}(x,y)} - 1\right) \leq \delta \mid \mathcal{S}_n\right]\right| \xrightarrow{p} 0 \quad \text{as } n \to \infty. \tag{4.1}$$

Proof. The bootstrap samples $\mathcal{S}_m^*$ can be represented equivalently by the samples $\widetilde{\mathcal{S}}_m^* = \{(\theta_i^*, Z_i^*, Y_i^*)\}_{i=1}^m$ or $\bar{\mathcal{S}}_m^* = \{(\theta_i^*, \zeta_i^*, Y_i^*)\}_{i=1}^m$. Recall the definitions of the events $A[\delta, n; h]$ and $A[\delta, n]$; replace n by m and (θ_i, Z_i, Y_i) by $(\theta_i^*, Z_i^*, Y_i^*)$ to define events $A[\delta, m; h]^*$ and $A[\delta, m]^*$, and note that $\text{Prob}\left[m^{\frac{2}{p+q+1}}\left(\frac{\widehat{\theta}^*(x,y)}{\widehat{\theta}(x,y)} - 1\right) \leq \delta \mid \mathcal{S}_n\right] =$

$\text{Prob}\,(A[\delta,m]^* \mid \mathcal{S}_n)$ holds for all $m,\ \delta$. Theorem 2 implies $|m^{\frac{2}{p+q+1}}(\frac{\widehat{\theta}(x,y)}{\theta(x,y)}-1)|\xrightarrow{p}0$ as $n\to\infty$, and hence

$$\sup_{\delta}\left|\text{Prob}\left[m^{\frac{2}{p+q+1}}\left(\frac{\widehat{\theta}^*(x,y)}{\widehat{\theta}(x,y)}-1\right)\le\delta\mid\mathcal{S}_n\right]-\text{Prob}\,(A[\delta,m]^*\mid\mathcal{S}_n)\right|=o_p(1). \tag{4.2}$$

Now consider the sets $C(x,y;hm^{-\frac{1}{p+q+1}})$, and note $\text{Prob}((\theta_i^*,Z_i^*,Y_i^*)\in C(x,y;hm^{-\frac{1}{p+q+1}})\mid\mathcal{S}_n)$ is equivalent to the relative frequency of points in $\widetilde{\mathcal{S}}_n$ falling into $C(x,y;hm^{-\frac{1}{p+q+1}})$. Consequently,

$$\left|\frac{\text{Prob}((\theta_i^*,Z_i^*,Y_i^*)\in C(x,y;hm^{-\frac{1}{p+q+1}})\mid\mathcal{S}_n)}{\text{Prob}((\theta_i,Z_i,Y_i)\in C(x,y;hm^{-\frac{1}{p+q+1}}))}-1\right|=O_p\left(n^{(\kappa-1)/2}\right).$$

Standard results on the convergence of the empirical distribution now can be used to show that also the conditional distributions of the points falling into $C(x,y;hn^{-\frac{1}{p+q+1}})$ asymptotically coincide:

$$\sup_{\mathcal{C}}\left|\frac{\text{Prob}[(\theta_i^*,Z_i^*,Y_i^*)\in\mathcal{C}\mid\mathcal{S}_n]}{\text{Prob}[(\theta_i^*,Z_i^*,Y_i^*)\in C(x,y;hm^{-\frac{1}{p+q+1}})\mid\mathcal{S}_n]}-\frac{\text{Prob}[(\theta_i,Z_i,Y_i)\in\mathcal{C}]}{\text{Prob}[(\theta_i,Z_i,Y_i)\in C(x,y;hm^{-\frac{1}{p+q+1}})]}\right|=o_p(1)$$

where the supremum refers to all $(p+q)$-dimensional subintervals $\mathcal{C}$ of $C(x,y;hm^{-\frac{1}{p+q+1}})$.

This leads to $\sup_\delta|\text{Prob}(A[\delta,m;h]^*\mid\mathcal{S}_n)-\text{Prob}(A[\delta,m;h])|\xrightarrow{p}0$ as $n\to\infty$. By arguments similar to those used to prove Theorem 1, it follows that for all $\epsilon>0$ there exists a h_ϵ such that for every $h\ge h_\epsilon$, $\text{Prob}\,(\sup_\delta|\text{Prob}(A[\delta,m;h]^*\mid\mathcal{S}_n)-\text{Prob}(A[\delta,m])|\ge\epsilon)\to0$ and $\text{Prob}\,(\sup_\delta|P(A[\delta,m;h]^*\mid\mathcal{S}_n)-P(A[\delta,m]^*\mid\mathcal{S}_n)|\ge\epsilon)\to0$ as $n\to\infty$. The assertion of the theorem now follows from (4.2) and Theorems 1 and 2. ■

4.2 Bootstrap with Smoothing

Alternatively, a bootstrap procedure that generates pseudo-samples based on a smoothed empirical distribution and a smoothed estimate of g_x allows consistent inference about $\theta(x,y)$. This bootstrap procedure consists of the following steps (details of the smoothing procedures will be discussed in a sequel):

Algorithm #2:

[1] Compute a ***smooth*** analog $\widehat{g}_x^*(z,\widetilde{y})$ of the frontier function $\widehat{g}_x(z,\widetilde{y})$; details are given below.

[2] Draw a bootstrap sample $\bar{\mathcal{S}}_n^* = \{(\theta_i^*, \zeta_i^*, Y_i^*)\}_{i=1}^n$ by iid sampling from a smooth, nonparametric estimate $\widehat{f}_x$ of the density $\bar{f}_x$. Then determine $\widetilde{\mathcal{S}}_n^* = \{(\theta_i^*, Z_i^*, Y_i^*)\}_{i=1}^n$ using $Z_i^* = \sum_{j=1}^{p-1} \zeta_{ij}^* z^{(j)}$.

[3] Define a bootstrap sample $\mathcal{S}_n^* = \{(X_i^*, Y_i^*)\}_{i=1}^n$ of size n by setting

$$X_i^* = \frac{\widehat{g}_x^*(\theta_i^* Z_i^*, Y_i^*)}{\theta_i^*} \frac{x}{||x||} + Z_i^*.$$

[4] Apply the original DEA estimator in (2.7) to obtain a bootstrap estimate $\widehat{\theta}^*(x, y)$.

[5] Repeat steps [2]–[4] B times; use the resulting bootstrap values to approximate the conditional distribution of $(\frac{\widehat{\theta}^*(x,y)}{\widehat{\theta}(x,y)} - 1)$ given $\mathcal{S}_n$, and use this to approximate the unknown distribution of $(\frac{\hat{\theta}(x,y)}{\theta(x,y)} - 1)$. For a given $\alpha \in (0, 1)$, use the bootstrap values to estimate the quantiles $\delta_{\alpha/2},\ \delta_{1-\alpha/2}$ where

$$\begin{aligned} \text{Prob}\left[\left(\frac{\hat{\theta}^*(x,y)}{\widehat{\theta}(x,y)} - 1\right) \le \delta_{\alpha/2} \mid \mathcal{S}_n\right] &= \frac{\alpha}{2}, \\ \text{Prob}\left[\left(\frac{\hat{\theta}^*(x,y)}{\widehat{\theta}(x,y)} - 1\right) \le \delta_{1-\alpha/2} \mid \mathcal{S}_n\right] &= 1 - \frac{\alpha}{2}. \end{aligned}$$

[6] Compute $\left[\frac{\widehat{\theta}(x,y)}{1+\delta_{1-\alpha/2}},\ \frac{\widehat{\theta}(x,y)}{1+\delta_{\alpha/2}}\right]$, a symmetric $(1-\alpha)$ confidence interval estimate for $\theta(x, y)$.

Recall that if $p = 1$, then g_x is the "frontier function" and does not depend on x. Moreover, in this case, $Z_i \equiv 0$ and $\widehat{f}_x$ as well as g_x only depend on y. However, for $p > 1$ the above steps define g_x and $\widehat{f}_x$ specifically for the point (x, y) that is of interest. Consequently, if confidence intervals are to be constructed for the efficiency measure defined in (2.5) evaluated at different points in $\mathbb{R}_+^{p+q}$, separate bootstraps must be performed for each of these points.

In the simulations described in the next section, we use kernel estimators to approximate $\bar{f}_x$. The only particular difficulty is the discontinuity of $\bar{f}_x(\theta, \zeta, \widetilde{y})$ at points $(\theta, \zeta, \widetilde{y})$ with $\theta = 1$. This problem is handled by reflecting observations $(\widehat{\theta}_i, \zeta_i, Y_i)$ to obtain $(2 - \widehat{\theta}_i, \zeta_i, Y_i)$ (where $\widehat{\theta}_i$ denotes the efficiency estimate computed from the smoothed frontier $\widehat{g}_x^*$ for the ith observation), and incorporating the resulting $2n$ points in the estimation. We use a Gaussian product kernel, with separate bandwidths for each marginal dimension chosen using the univariate two-stage plug-in method described by Sheather and Jones (1991). Alternatively, one could use least-squares cross-validation as described by Simar and Wilson (2000a), but the approach employed here imposes much less computational burden.

The specification of the function $\widehat{g}_x^*$ in step [1] of Algorithm #2 is crucial for validity of the bootstrap procedure. Unfortunately, it is not possible to rely on the estimated DEA frontier. The difference between $\widehat{g}_x$ and g_x is of order $n^{-\frac{2}{p+q+1}}$; even more importantly, $\widehat{g}_x$ is not differentiable and hence does not possess the same degree of smoothness as g_x. Setting $\widehat{g}_x^* = \hat{g}_x$ therefore does not seem to lead to a consistent bootstrap. Even if the distributions of (θ_i, Z_i, Y_i) and $(\theta_i^*, Z_i^*, Y_i^*)$ were identical, the asymptotic distributions of $\sum_{j=1}^k \alpha_j \frac{g_x(\theta_j Z_j, Y_j)}{\theta_j g_x(0,y)} - 1$ and $\sum_{j=1}^k \alpha_j \frac{\hat{g}_x(\theta_j^* Z_j^*, Y_j^*)}{\theta_j^* \hat{g}_x(0,y)} - 1$ will not in general coincide.

It is important to understand the purpose of smoothing the DEA frontier estimate. We do not require that $\widehat{g}_x^*$ be closer to g_x than $\widehat{g}_x$. It suffices completely if the relative distances $\frac{\tilde{g}_x(z,\widetilde{y})}{g_x(z,\widetilde{y})}$ do not change very much with $(z, \widetilde{y})$. If, for some $\beta > 0$, we have $\beta g_x(z, \widetilde{y}) = \widetilde{g}_x(z, \widetilde{y})$ for all $(z, \widetilde{y})$, then $\frac{g_x(\theta_i Z_i, Y_i)}{g_x(0,y)} = \frac{\tilde{g}_x(\theta_i Z_i, Y_i)}{\tilde{g}_x(0,y)}$, and by Lemma 2 the errors of the resulting DEA estimators are identical. In effect, proportionality is not necessary. We can infer from Proposition 1 that even if the first derivatives of g_x and $\widetilde{g}_x^*$ are completely different, the limiting distributions will be close as long as the second derivatives approximately coincide. In smoothing the DEA frontier function in step [1], it is therefore essential to preserve convexity.

One possibility would be to employ convolution smoothing of $\widehat{g}_x$. This approach, however, presents a formidable integration problem in $(p + q - 1)$-dimensions, and it seems unlikely that such an approach could be successfully implemented with real data. Alternatively, one may use a bandwidth $b \in (0, 1)$ to define a smooth *bootstrap frontier* $\widehat{g}_x^*$ by

$$\widehat{g}_x^*(z, \widetilde{y}) = \widehat{g}_x(0, y) + b^2 \left[\widehat{g}_x \left(\frac{z}{b},\ y + \frac{\widetilde{y} - y}{b} \right) - \widehat{g}_x(0, y) \right] \tag{4.3}$$

Note that setting $b = 1$ in (4.3) results in no smoothing of the frontier; in this case, the resulting procedure is similar to the "single-smooth" algorithm proposed by Simar and Wilson (2000a).

To understand the motivation for the smoothing in (4.3), let $b < 1$ and define

$$g_x^*(z, \widetilde{y}) = g_x(0, y) + b^2 \left[g_x \left(\frac{z}{b},\ y + \frac{\widetilde{y} - y}{b} \right) - g_x(0, y) \right]. \tag{4.4}$$

The following properties are easily verified: (i) $\widehat{g}_x^*$ as well as g_x^* are convex functions; (ii) $\widehat{g}_x^*(0, y) = \widehat{g}_x(0, y) = \widehat{\theta}(x, y)||x||$ as well as $g_x^*(0, y) = g_x(0, y) = \theta(x, y)||x||$; (iii) the second derivatives of g_x^* and of g_x at the point $(0,y)$ are identical, i.e. $g_x''(0,y) = g_x^{*\prime\prime}(0,y)$; and (iv) by Theorem 1(b),

$$\left|\frac{\widehat{g}_x^*(z,\widetilde{y})}{\widehat{g}_x^*(0,y)} - \frac{g_x^*(z,\widetilde{y})}{g_x^*(0,y)}\right| = \left| b^2 \frac{\widehat{g}_x\left(\frac{z}{b}, y + \frac{\widetilde{y}-y}{b}\right)}{\widehat{g}_x(0,y)} - b^2 \frac{g_x\left(\frac{z}{b}, y + \frac{\widetilde{y}-y}{b}\right)}{g_x(0,y)} \right| = b^2 n^{-\frac{2}{p+q+1}} \log n \qquad (4.5)$$

for all $(\frac{z}{b}, y + \frac{\tilde{y}-y}{b})$ in a sufficiently small neighborhood of $(0,y)$.

Property (iv) implies that if $b^2 \log n \to 0$ as $n \to \infty$, the difference between $\widehat{g}_x^*$ and g_x^* is of ***smaller*** order than $n^{-\frac{2}{p+q+1}}$. Asymptotically, a bootstrap based on $\widehat{g}_x^*$ will thus provide the same results as a bootstrap directly relying on g_x^*. On the other hand, it follows from properties (i)–(iii) that the parameters determining the asymptotic distribution of efficiency estimates from g_x^* coincide with those from g_x.

It is possible to determine a suitable order of magnitude of b. For purposes of establishing consistency of the bootstrap, g_x need only be twice continuously differentiable (see Assumption 7 below). Here, we assume that g_x is three-times continuously differentiable only for selecting a suitable order of magnitude for b. Of course, one might exploit this assumption to develop an inefficiency estimator different from the DEA estimator; such a method would be based on further smoothing of the frontier, but would likely be rather more complicated for practitioners than the DEA estimator which is the focus of this paper. If g_x is replaced by g_x^*, then (3.15) becomes

$$\begin{aligned}
\sum_{j=1}^{k} \alpha_j \frac{g_x^*(\theta_{i_j} Z_{i_j}, Y_{i_j})}{\theta_{i_j} g_x^*(0,y)} - 1 &= \sum_{j=1}^{k} \alpha_j \frac{g_x^*(Z_{i_j}, Y_{i_j}) - g_x^*(0,y)}{g_x^*(0,y)} + \sum_{j=1}^{k} \alpha_j (1-\theta_{i_j}) + O_p(n^{-\frac{3}{p+q+1}}) \\
&= \sum_{j=1}^{k} \alpha_j \frac{1}{2 g_x^*(0,y)} \Big[Z_{i_j}^T g_{x;zz}^{*}{}''(0,y) Z_{i_j} + 2 Z_{i_j}^T g_{x;zy}^{*}{}''(0,y)(Y_{i_j} - y) \\
&\qquad + (Y_{i_j} - y)^T g_{x;yy}^{*}{}''(0,y)(Y_{i_j} - y) \Big] \\
&\qquad + \sum_{j=1}^{k} \alpha_j (1-\theta_{i_j}) + O_p\left(b^{-1} n^{-\frac{3}{p+q+1}}\right).
\end{aligned} \qquad (4.6)$$

Thus, the bootstrap analog of the assertion in Proposition 1 holds provided $n^{-\frac{1}{p+q+1}}/b \to 0$. The approximation error in (4.6) is the smaller the larger is b. On the other hand, the estimation error (4.5) decreases with b. The remainder terms in (4.5) and (4.6) are of the same order of magnitude (up to a $\log n$ term); summing the remainder terms and then minimizing with respect to b suggests that b should be chosen proportional to $n^{-\frac{1}{3(p+q+1)}}$.

An obvious difficulty of the above bootstrap consists in the fact that in most bootstrap samples there will exist points (Z_i^*, Y_i^*) with $(\frac{Z_i^*}{b}, y+\frac{Y_i^*-y}{b}) \notin \widehat{\Psi}^*$, where $\widehat{\Psi}^*$ denotes the convex hull of the free-disposal hull of the bootstrap observations in $\mathcal{S}_n^*$. This phenomenon is not very important in terms of asymptotic theory since by Theorem 1, the DEA estimator is essentially only determined by points in a neighborhood of $(\theta(x,y)x, y)$. However, any implementation of the algorithm requires that one must deal with such points. Two possibilities exist:

Elimination: Suppose that in the bootstrap sample there are $\ell < n$ points with $\left(\frac{Z_{i_j}^*}{b}, y + \frac{Y_{i_j}^* - y}{b}\right) \notin \widehat{\Psi}^*$, $i_j \in \{1, \ldots, n\}$, $j = 1, \ldots, l$. Eliminate these points from the bootstrap samples and calculate $\widehat{\theta}^*(x,y)$ from the remaining $(n-\ell)$ bootstrap observations.

Extrapolation: Suppose that for some $i \in \{1, \ldots, n\}$ we have $\left(\frac{Z_i^*}{b}, y + \frac{Y_i^*-y}{b}\right) \notin \widehat{\Psi}^*$. Let b^* denote the smallest possible $\widetilde{b}$ such that $\left(\frac{Z_i^*}{b}, y + \frac{Y_i^*-y}{\widetilde{b}}\right) \in \widehat{\Psi}^*$. Clearly, $b^* > b$. The structure of the DEA estimator implies that for all $\widetilde{b} > b^*$ sufficiently close to b^*, there exist some β_0, β_1 such that $\widehat{g}_x\left(\frac{Z_i^*}{b}, y + \frac{Y_i^*-y}{\widetilde{b}}\right) = \beta_0 + \beta_1 \frac{1}{\widetilde{b}}$. Then "define" $\widehat{g}_x\left(\frac{Z_i^*}{b}, y + \frac{Y_i^*-y}{b}\right) := \beta_0 + \beta_1 \frac{1}{b}$ and calculate the corresponding value of $\widehat{g}_x^*(Z_i^*, Y_i^*)$.

In the simulations described in Section 5, we use the elimination option.

We now consider the asymptotic behavior of the double-smooth bootstrap proposed above. Our analysis rests upon the following additional assumption:

Assumption 7. *The density estimate $\widehat{f}_x$ satisfies*

$$\sup_{(\theta,z,\widetilde{y}) \in C(x,y;h)} \left| \widehat{f}_x(\theta, z, \widetilde{y}) - \bar{f}_x(\theta, z, \widetilde{y}) \right| = o_p(1) \quad \text{as } n \to \infty \tag{4.7}$$

if h is sufficiently small. Furthermore, g_x is two times continuously differentiable and $b \to 0$ as well as $n^{-\frac{1}{p+q+1}}/b \to 0$ as $n \to \infty$.

The next theorem ensures consistency of our double-smooth bootstrap.

Theorem 4. *Given Assumptions 1–7,*

$$\sup_{\delta > 0} \left| F_x(\delta) - \text{Prob}\left(n^{\frac{2}{p+q+1}} \left(\frac{\widehat{\theta}^*(x,y)}{\widehat{\theta}(x,y)} - 1 \right) \leq \delta \mid \mathcal{S}_n \right) \right| \xrightarrow{p} 0 \quad \text{as } n \to \infty.$$

A proof is given in the appendix.

5 Monte Carlo Evidence

We conducted two sets of experiments, with $p = q = 1$ and $p = q = 2$. All experiments consist of 1,000 Monte Carlo trials, with 2,000 bootstrap replications on each trial. Within either set of experiments, we examined 7 sample sizes, with $n \in \{25,\ 50,\ 100,\ 200,\ 400,\ 800\}$. For the case with one output and one input ($p = q = 1$), we simulated a DGP by drawing an "efficient" input observation x_e distributed uniformly on $[10, 20]$, and setting the output level $y = x_e^{0.8}$. We then computed the "observed" input observation $x = x_e e^{0.2|\varepsilon|}$, where $\varepsilon \sim N(0, 1)$ and is independent. The DGP for this case can therefore be written as

$$y = x^{0.8} e^{-0.16|\varepsilon|}. \tag{5.1}$$

We take the point $(x, y) = (20.69, 7.5)$ as the fixed point for which efficiency is estimated on each Monte Carlo trial; the true efficiency for this point is $\theta(x, y) = 0.6$.

For the two-input, two-output ($p = q = 2$) case, we again generated efficient input levels $x_{1e},\ x_{2e}$ from the uniform distribution on $[10, 20]$. Next, we computed output levels by generating ω uniform on $\left[\frac{1}{9}\frac{\pi}{2},\ \frac{8}{9}\frac{\pi}{2}\right]$ and setting $y_1 = x_{1e}^{0.4} x_{2e}^{0.4} \times \cos(\omega)$ and $y_2 = x_{1e}^{0.4} x_{2e}^{0.4} \times \sin(\omega)$. We then generated the observed output levels by setting $x_1 = x_{1e} e^{0.2|\varepsilon|}$ and $x_2 = x_{2e} e^{0.2|\varepsilon|}$ and where $\varepsilon \sim N(0, 1)$ as before. Efficiency is estimated for the fixed point $x = (22.07, 22.07)$, $y = (5.59, 5.59)$ on each Monte Carlo trial. The true efficiency for this point is $\theta(x, y) = 0.6$, as in the previous case.

In both cases, the fixed points of interest were chosen to lie roughly in the middle of the range of the output data. In the case where $p = q = 2$, the output quantities, for a given level of inputs, are generated to lie on an arc between $\pi/18$ and $8\pi/18$ radians.

Table 1 shows results for coverages of confidence intervals estimated by the bootstrap-with-sub-sampling using Algorithm #1 as described in Section 4.1. For each sample size n, we examined bootstrap sample sizes $m = n^\kappa$ with $\kappa \in \{0.50,\ 0.55,\ \ldots,\ 0.95,\ 1.00\}$ When $\kappa = 1$ Algorithm #1 is identical to the naive bootstrap, which is known to provide inconsistent inference. For the case where $p = q = 1$ shown in columns 3–5, the results in Table 1 reveal good coverages for the ratio-based confidence intervals at the three significance levels considered when κ is in the neighborhood of 0.80. The optimal value of κ apparently remains about the same as sample size is increased from 25 to 800.

The results for the case where $p = q = 2$, shown in columns 6–8 of Table 1, reveal reduced coverage relative to the results for $p = q = 1$ for given values of n and κ, due to the curse of dimensionality. However, with $p = q = 2$, the coverages of confidence intervals are consistently good across the various sample sizes when κ lies in the neighborhood of 0.65–0.70. Not surprisingly, the optimal value of κ appears to depend on the dimensionality of the problem. The results also indicate that, as a practical matter, the wrong choice of κ, which determines the size of the subsamples, can lead to very poor coverages.

Results from the double bootstrap using Algorithm #2 are shown in Table 2, again for the cases $p = q = 1$ (shown in columns 3–5) and $p = q = 2$ (shown in columns 6–8). In either case, bandwidths $b \in \{0.4,\ 0.6,\ 0.8,\ 1.0\}$ were used to smooth $\widehat{g}_x$ in step [1] of the algorithm, using (4.3). As discussed previously, this bootstrap is inconsistent when $b = 1$; we include this case only for comparison. The results in Table 2 indicate some gains in terms of coverage of estimated confidence intervals as b is reduced below 1.0. In both cases, $b = 0.4$ appears too small, and indeed for $p = q = 2$ results could not be computed due to numerical problems when $n = 25$ or $n = 50$ (see the discussion preceding Assumption 7).

Recall from the discussion surrounding (4.6) that our theoretical results imply that the optimal value of b should be proportional to $n^{-\frac{1}{3(p+q+1)}}$. Since b is necessarily bounded between 0 and 1 (as opposed to bandwidths in ordinary kernel estimators), it is independent of the units of measurement for x and y. Clearly, b should be close to 1 for small n, and should become smaller as n increases. Using $b = n^{-\frac{1}{3(p+q+1)}}$ as a rule-of-thumb implies $b = n^{-1/9}$ for the case where $p = q = 1$, and $b = n^{-1/15}$ for $p = q = 2$. Hence, for $p = q = 1$, the rule-of-thumb criterion yields $b = 0.70,\ 0.65,\ 0.60,\ 0.56,\ 0.51$ and 0.48 corresponding to $n = 25,\ 50,\ 100,\ 200,\ 400$ and 800, respectively; for $p = q = 2$, we have $b = 0.81,\ 0.77,\ 0.74,\ 0.70,\ 0.67$ and 0.64, respectively. The results in Table 2 indicate that the rule-of-thumb gives rather reasonable choices for b. It is also interesting to note that, for sample sizes of 50 or greater, the estimated coverages in Table 2 vary little across $b = 0.4$ and $b = 0.6$ when $p = q = 1$, and $b = 0.6$ and $b = 0.8$ when $p = q = 2$.

The estimated coverages shown in Table 2 reveal that, for the case $p = q = 1$ and when $b = 0.4$ and $n = 200$ or 400 or when $b = 0.6$ and $n = 800$, the estimated coverages obtained with the double-smooth bootstrap are similar to the best coverages obtained with the sub-sampling bootstrap and shown in Table 1 when $p = q = 1$ and $n = 200$, 400, or 800. With $p = q = 2$, Table 2 reveals that coverages obtained with the double-smooth bootstrap are smaller than the best covereges for $p = q = 2$ shown in Table 1 for the sub-sampling bootstrap. However, Table 1 also reveals that sub-optimal choices of the tuning parameter κ required for the sub-sampling method can easily result in coverages worse than those shown in Table 2 when b is chosen according to the rule-of-thumb discussed above. Moreover, the coverages in Table 2 are typically too small, whereas coverages shown in Table 1 are either too large or too small, depending on whether κ is chosen too small or too large.

6 Conclusions

The analysis in Section 3 establishes the asymptotic distribution of the DEA efficiency estimator for the variable returns to scale case under rather weak assumptions on the DGP, while the analysis in Section 4 establishes consistency of two bootstrap procedures. The bootstrap procedures are necessary for any practical application since the asymptotic distribution in Theorem 2 contains unknown terms and would be difficult to either estimate or simulate. As noted in Sections 1 and 5, there is at present no reliable way to choose the size of subsamples in Algorithm #1, and hence we do not recommend the sub-sampling bootstrap. While Tables 1 and 2 indicate that in the ***best*** cases, the subsampling bootstrap performs better than the double-smooth bootstrap in terms of realized coverages, the practitioner—operating outside a Monte Carlo framework—is unlikely to achieve such performance, and is rather likely to do worse than he would using the double-smooth bootstrap. The second bootstrap procedure—based on smoothing—is, by contrast, readily implementable, and provides better coverage properties than the subsampling bootstrap is likely to provide without more guidance on choice of the tuning parameter κ. For finite samples in applications, one might optimize the choice of the bandwidth b in Algorithm #2. This could be accomplished by iterating the bootstrap procedures along the lines of Hall (1992).

A Appendix

Lemma A1: *Suppose that Assumptions 1–6 hold for a given* $(x, y) \in \mathcal{D}$ *and let* b, h *be real numbers with* $0 < b \leq h/2$. *Consider* $k \in \mathbb{N}$ *arbitrary points* $(\theta_1, z_1, y_1), \ldots, (\theta_k, z_k, y_k) \in \bar{\mathcal{D}}$

satisfying

$$\sum_{r=1}^{k} \alpha_r z_r = 0, \qquad \sum_{r=1}^{k} \alpha_r y_r = y \tag{A.1}$$

for some $\alpha_1, \ldots, \alpha_k \geq 0$ *with* $\sum_{r=1}^{k} \alpha_r = 1$. *If* $(\theta_k, z_k, y_k) \notin C(x, y; hn^{-\frac{1}{p+q+1}})$, *then for all sufficiently large* n *there exists some* $(\widetilde{z}, \widetilde{y}) \in \Psi^*(x)$ *with* $(1, \widetilde{z}, \widetilde{y}) \in C(x, y; bn^{-\frac{1}{p+q+1}})$ *such that*

$$\sum_{r=1}^{k-1} \widetilde{\alpha}_r z_r + \widetilde{\alpha}_k \widetilde{z} = 0, \qquad \sum_{r=1}^{k-1} \widetilde{\alpha}_r y_r + \widetilde{\alpha}_k \widetilde{y} = y \tag{A.2}$$

for some $\widetilde{\alpha}_1, \ldots, \widetilde{\alpha}_k \geq 0$ *with* $\sum_{r=1}^{k} \widetilde{\alpha}_r = 1$ *and such that*

$$\sum_{r=1}^{k} \alpha_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0, y)} \geq \sum_{r=1}^{k-1} \widetilde{\alpha}_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0, y)} + \widetilde{\alpha}_k \frac{g_x(\widetilde{z}, \widetilde{y})}{g_x(0, y)} + c_1 \cdot \widetilde{\alpha}_k hbn^{-\frac{2}{p+q+1}} \tag{A.3}$$

where $c_1 = \min\{\frac{1}{2}, \frac{c_0}{8g_x(0,y)}\}$ *and* c_0 *is defined as in Lemma 1(b).*

Proof: Assume (A.1) holds with $(\theta_k, z_k, y_k) \notin C(x, y; hn^{-\frac{1}{p+q+1}})$. Then either $\theta_k \leq 1 - h^2 n^{-\frac{2}{p+q+1}}$ and $(1, z_k, y_k) \in C(x, y; hn^{-\frac{1}{p+q+1}})$ or $(1, z_k, y_k) \notin C(x, y; hn^{-\frac{1}{p+q+1}})$.

First consider the case where $\theta_k \leq 1 - h^2 n^{-\frac{2}{p+q+1}}$ but $(1, z_k, y_k) \in C(x, y; hn^{-\frac{1}{p+q+1}})$. Since $\frac{1}{\theta_k} - 1 \geq 1 - \theta_k$ we obtain $\frac{g_x(\theta_k z_k, y_k)}{\theta_k g_x(0,y)} \geq \frac{g_x(\theta_k z_k, y_k)}{g_x(0,y)} + (1 - \theta_k)\frac{g_x(\theta_k z_k, y_k)}{g_x(0,y)}$. Straightforward Taylor expansions of g_x can be used to show that for all sufficiently large n,

$$\frac{g_x(\theta_k z_k, y_k)}{\theta_k g_x(0, y)} \geq \frac{g_x(z_k, y_k)}{g_x(0, y)} + \frac{1}{2}(1 - \theta_k) \geq \frac{g_x(z_k, y_k)}{g_x(0, y)} + \frac{1}{2} h^2 n^{-\frac{2}{p+q+1}}. \tag{A.4}$$

Note that $(1, z_k, y_k) \in C(x, y; hn^{-\frac{1}{p+q+1}})$ implies that $(1, \frac{b}{h} z_k, y + \frac{b}{h}(y_k - y)) \in C(x, y; bn^{-\frac{1}{p+q+1}})$. Relation (A.2) thus holds for $(\widetilde{z}, \widetilde{y}) := (\frac{b}{h} z_k, y + \frac{b}{h}(y_k - y))$ and $\widetilde{\alpha}_r = \alpha_r \frac{\frac{b}{h}}{\frac{b}{h} + \alpha_k(1 - \frac{b}{h})}$ as well as $\widetilde{\alpha}_k = \alpha_k \frac{1}{\frac{b}{h} + \alpha_k(1 - \frac{b}{h})}$. Then (A.4) and convexity of g_x lead to

$$\sum_{r=1}^{k} \alpha_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0, y)} \geq \frac{\frac{b}{h}}{\frac{b}{h} + \alpha_k(1 - \frac{b}{h})} \left(\sum_{r=1}^{k} \alpha_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0, y)} \right) + \frac{\alpha_k(1 - \frac{b}{h})}{\frac{b}{h} + \alpha_k(1 - \frac{b}{h})}$$

$$\geq \sum_{r=1}^{k-1} \widetilde{\alpha}_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0, y)} + \widetilde{\alpha}_k \left(\frac{\frac{b}{h} g_x(z_k, y_k)}{g_x(0, y)} + (1 - \frac{b}{h}) \frac{g_x(0, y)}{g_x(0, y)} \right) + \widetilde{\alpha}_k \frac{b}{h} \frac{1}{2} h^2 n^{-\frac{2}{p+q+1}}$$

$$\geq \sum_{r=1}^{k-1} \widetilde{\alpha}_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0, y)} + \widetilde{\alpha}_k \frac{g_x(\widetilde{z}, \widetilde{y})}{g_x(0, y)} + \widetilde{\alpha}_k \frac{1}{2} bhn^{-\frac{2}{p+q+1}}. \tag{A.5}$$

It now only remains to prove (A.3) for the case where $(1, z_k, y_k) \notin C(x, y; hn^{-\frac{1}{p+q+1}})$. Let $\gamma = \max\{\delta \mid (1, \delta z_k, y + \delta(y_k - y)) \in C(x, y; hn^{-\frac{1}{p+q+1}})\}$ as well as $\alpha_r^* = \alpha_r \frac{\gamma}{\gamma + \alpha_k(1-\gamma)}$ and $\alpha_k^* = \alpha_k \frac{1}{\gamma + \alpha_k(1-\gamma)}$. This yields

$$\sum_{r=1}^{k-1} \alpha_r^* z_r + \alpha_k^* \gamma z_k = 0, \qquad \sum_{r=1}^{k-1} \alpha_r^* y_r + \alpha_k^* (y + \gamma(y_k - y)) = y \tag{A.6}$$

By definition of g_x we have $g_x(\theta_k z_k, y_k)/\theta_k \geq g_x(z_k, y_k)$. Convexity of g_x and arguments similar to (A.5) then imply

$$\begin{aligned}\sum_{r=1}^{k} \alpha_r \frac{g_x(\theta_r z_r, Y_r)}{\theta_r g_x(0, y)} &\geq \sum_{r=1}^{k-1} \alpha_r^* \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0, y)} + \alpha_k^* \left(\frac{\gamma g_x(z_k, y_k)}{g_x(0, y)} + (1-\gamma) \frac{g_x(0, y)}{g_x(0, y)} \right) \\ &\geq \sum_{r=1}^{k-1} \alpha_r^* \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0, y)} + \alpha_k^* \frac{g_x(\gamma z_k, y + \gamma(y_k - y))}{g_x(0, y)}.\end{aligned} \tag{A.7}$$

Finally, define $(\widetilde{z}, \widetilde{y}) := \left(\frac{b}{h}\gamma z_k, y + \frac{b}{h}\gamma(y_k - y)\right)$ and $\widetilde{\alpha}_r = \alpha_r^* \frac{\frac{b}{h}}{\frac{b}{h} + \alpha_k^*(1 - \frac{b}{h})}$ as well as $\widetilde{\alpha}_k = \alpha_k^* \frac{1}{\frac{b}{h} + \alpha_k^*(1 - \frac{b}{h})}$. Clearly, then, $(1, \widetilde{z}, \widetilde{y}) \in C(x, y; bn^{-\frac{1}{p+q+1}})$, and relation (A.2) is a direct consequence of (A.6). Moreover, for sufficiently large n,

$$\begin{aligned}&\sum_{r=1}^{k-1} \alpha_r^* \frac{g_x(\theta_r z_r, Y_r)}{\theta_r g_x(0, y)} + \alpha_k^* \frac{g_x(\gamma z_k, y + \gamma(y_k - y))}{g_x(0, y)} \\ &\quad\geq \sum_{r=1}^{k-1} \widetilde{\alpha}_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0, y)} + \widetilde{\alpha}_k \left[\frac{\frac{b}{h} g_x(\gamma z_k, y + \gamma(y_k - y))}{g_x(0, y)} + (1 - \frac{b}{h}) \frac{g_x(0, y)}{g_x(0, y)} \right] \\ &\quad\geq \sum_{r=1}^{k-1} \widetilde{\alpha}_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0, y)} + \widetilde{\alpha}_k \frac{g_x(\widetilde{z}_k, \widetilde{y}_k)}{g_x(0, y)} + \widetilde{\alpha}_k \frac{b}{h} \frac{c_0 h^2 n^{-\frac{2}{p+q+1}}}{8 g_x(0, y)}.\end{aligned} \tag{A.8}$$

By using Lemma 1(b) the second inequality follows from Taylor expansions of $g_x(\gamma z_k, y + \gamma(y_k - y))$ as well as $g_x(0, y)$ at the point $(\widetilde{z}, \widetilde{y}) := (\frac{b}{h}\gamma z_k, y + \frac{b}{h}\gamma(y_k - y))$. Note that the first derivatives cancel out due to $\frac{b}{h}(\gamma z_k - \frac{b}{h}\gamma z_k) + (1 - \frac{b}{h}) \cdot (-\frac{b}{h}\gamma z_k) = 0$ and $\frac{b}{h}(\gamma(y_k - y) - \frac{b}{h}\gamma(y_k - y)) + (1 - \frac{b}{h}) \cdot (-\frac{b}{h}\gamma(y_k - y)) = 0$. The bound given in (A.8) is then obtained by an analysis of the second derivatives while taking into account that $1 - \frac{b}{h} \geq \frac{1}{2}$, $\left\| \begin{pmatrix} \gamma z_k \\ \gamma(y_k - y) \end{pmatrix} \right\|^2 \geq h^2$, and that $\inf_{(1,z,w) \in C(x,y;bn^{-\frac{1}{p+q+1}})} \inf_{||v||=1} v^T g_x''((z, w)v \geq \frac{c_0}{2}$ for all sufficiently large n, where c_0 is defined in Lemma 1(b). Combining (A.7) and (A.8) yields (A.3). ∎

Proof of Theorem 1: Let $z^{(1)},\ \ldots,\ z^{(p-1)}$ denote the orthonormal basis of $\mathcal{V}(x)$ used in the definition of $\bar{f}_x$. Note that the sample $\mathcal{S}_n$ of observations can be equivalently represented be the corresponding samples $\widetilde{\mathcal{S}}_n = \{(\theta_i, Z_i, Y_i)\}_{i=1}^n$ and $\bar{\mathcal{S}}_n = \{(\theta_i, \zeta_i, Y_i)\}_{i=1}^n$, where ζ_i is determined by $Z_i = \sum_{j=1}^{p-1} \zeta_{ij} z^{(j)}$.

Choose an arbitrary $b > 0$ and set $b_n = b \cdot n^{-\frac{1}{p+q+1}}$, $b_n^* = \frac{b_n}{2(p-1)+2q}$. For $i = 1,\ \ldots,\ p-1$ and $j = 1,\ \ldots,\ q$, define

$$\bar{B}_{2i-1} = \{(v, w) \in \mathbb{R}^{p-1} \times \mathbb{R}^q \mid \max_{r \neq i} |v_r| \leq b_n^*, |v_i - b_n| \leq b_n^*, \max_{s=1,\ \ldots,\ q} |y_s - w_s| \leq b_n^*\},$$

$$\bar{B}_{2i} = \{(v, w) \in \mathbb{R}^{p-1} \times \mathbb{R}^q \mid \max_{r \neq i} |v_r| \leq b_n^*, |v_i + b_n| \leq b_n^*, \max_{s=1,\ \ldots,\ q} |y_s - w_s| \leq b_n^*\},$$

$$\bar{B}_{2j-1+2(p-1)} = \{(v, w) \in \mathbb{R}^{p-1} \times \mathbb{R}^q \mid \max_{r=1,\ \ldots,\ p-1} |v_r| \leq b_n^*, \max_{s \neq j} |y_s - w_s| \leq b_n^*, |y_j + b_n - w_j| \leq b_n^*\},$$

$$\bar{B}_{2j+2(p-1)} = \{(v, w) \in \mathbb{R}^{p-1} \times \mathbb{R}^q \mid \max_{r=1,\ \ldots,\ p-1} |v_r| \leq b_n^*, \max_{s \neq j} |y_s - w_s| \leq b_n^*, |y_j - b_n - w_j| \leq b_n^*\}.$$

Finally, for $j = 1,\ \ldots,\ 2(p-1) + 2q$ let B_j denote the set of all $(z, w) \in \mathcal{V}(x) \times \mathbb{R}_+^q$ with $(z, w) = (\sum_j v_j z^{(j)}, w)$ for some $(v, w) \in \bar{B}_j$.

It follows from Assumptions 4–5 that if n is sufficiently large,

$$\bar{D}_{j,n} := [1 - b_n^2, 1] \times \bar{B}_j \subset \bar{\mathcal{D}} \tag{A.9}$$

for all $j = 1,\ \ldots,\ 2(p-1) + 2q$. Recall that $\bar{\mathcal{D}}$ denotes the support of $\bar{f}_x$.

For each $j = 1,\ \ldots,\ 2(p-1) + 2q$ the set $\bar{D}_{j,n}$ has Lebesgue measure proportional to $b^{p+q+1} \cdot \frac{1}{n}$, and our assumptions on the distribution of the random variables (θ_i, ζ_i, Y_i) thus imply $\text{Prob}\left[(\theta_i, \zeta_i, y_i) \in \bar{D}_{j,n}\right]$ is proportional to $b^{p+q+1} \cdot \frac{1}{n}$. It therefore follows from standard arguments that there exist some $0 < d_0,\ d_1 < \infty$ such that for all n sufficiently large,

$$\begin{aligned} 1 - (2(p-1) + 2q) \cdot \exp(-d_0 b^{p+q+1}) &\leq \text{Prob}\left(\bar{\mathcal{S}}_n \cap \bar{D}_{j,n} \neq \emptyset \ \forall\ j = 1,\ \ldots,\ 2(p-1) + 2q\right) \\ &\leq 1 - \exp(-d_1 b^{p+q+1}). \end{aligned} \tag{A.10}$$

Hence for every $\epsilon > 0$, there exits a $b_\epsilon < \infty$ such that for all $b \geq b_\epsilon$ and all n sufficiently large,

$$\text{Prob}\left(\bar{\mathcal{S}}_n \cap \bar{D}_{j,n} \neq \emptyset \ \forall\ j = 1,\ \ldots,\ 2(p-1) + 2q\right) \geq 1 - \epsilon. \tag{A.11}$$

By (A.11), assertion (a) of the theorem holds if there is a $h_\epsilon > 0$ such that for all $h > h_\epsilon$ the following conditional probabilities are equivalent for sufficiently large n:

$$\text{Prob}\left(A[\delta, n] \mid \bar{\mathbb{S}}_n \cap \bar{D}_{j,n} \neq \emptyset \;\forall\, j\right) = \text{Prob}\left(A[\delta, n; h \cdot n^{-\frac{1}{p+q+1}}] \mid \bar{\mathbb{S}}_n \cap \bar{D}_{j,n} \neq \emptyset \;\forall\, j\right). \quad \text{(A.12)}$$

Now we will demonstrate that (A.12) is satisfied for all $h \geq c_3 \cdot b$, where $c_3 < \infty$ denotes a suitable constant which will be specified in the sequel.

By construction of $\bar{B}_j$ and B_j, for any $(\widetilde{z}, \widetilde{y}) \in \Psi^*(x)$ with $(1, \widetilde{z}, \widetilde{y}) \in C(x, y; b_n^*)$ and ***arbitrary*** vectors $(\widetilde{\theta}_1, \widetilde{z}_1, \widetilde{w}_1) \in [1 - b_n^2, 1] \times B_1, \ldots, (\widetilde{\theta}_{2(p-1)+2q}, \widetilde{z}_{2(p-1)+2q}, \widetilde{w}_{2(p-1)+2q}) \in [1 - b_n^2, 1] \times B_{2(p-1)+2q}$, there exist some $\gamma_1, \ldots, \gamma_{2(p-1)+2q} \geq 0$ with $\sum_{j=1}^{2(p-1)+2q} \gamma_j = 1$ such that

$$\widetilde{z} = \sum_{j=1}^{2(p-1)+2q} \gamma_j \widetilde{z}_j, \qquad \widetilde{y} = \sum_{j=1}^{2(p-1)+2q} \gamma_j \widetilde{w}_j. \quad \text{(A.13)}$$

By definition of $(\widetilde{\theta}_j, \widetilde{z}_j, \widetilde{w}_j)$, for sufficiently large n $\frac{g_x(\widetilde{\theta}_j \widetilde{z}_j, \widetilde{w}_j)}{\widetilde{\theta}_j g_x(0,y)} \leq 1.5$, $\left|\left|\begin{pmatrix} \widetilde{\theta}_j \widetilde{z}_j - \widetilde{z} \\ \widetilde{w}_j - \widetilde{y} \end{pmatrix}\right|\right|^2 \leq (2(p-1) + 2q) b_n^2$, and

$$\sup_{(1,z,w) \in C(x,y;b_n^*)} \left[\sup_{||v||=1} v^T g_x''((z,w)v\right] \leq c_0^*$$

for some $c_0^* < \infty$. Therefore, for all n sufficiently large,

$$\begin{aligned} \frac{g_x(\widetilde{z}, \widetilde{y})}{g_x(0,y)} &\leq \sum_{j=1}^{2(p-1)+2q} \gamma_j \frac{g_x(\widetilde{\theta}_j \widetilde{z}_j, \widetilde{w}_j)}{\widetilde{\theta}_j g_x(0,y)} \\ &\leq \sum_{j=1}^{2(p-1)+2q} \gamma_j \left(\frac{g_x(\widetilde{\theta}_j \widetilde{z}_j, \widetilde{w}_j)}{g_x(0,y)} + 1.5(\frac{1}{\widetilde{\theta}_j} - 1)\right) \leq \frac{g_x(\widetilde{z}, \widetilde{y})}{g_x(0,y)} + c_2 b^2 n^{-\frac{2}{p+q+1}} \end{aligned} \quad \text{(A.14)}$$

where $c_2 = \frac{(2(p-1)+2q)c_0^*}{2g_x(0,y)} + 2$.

Using the continuity of g_x'', the second inequality can be derived from second order Taylor expansions of $g_x(\widetilde{\theta}_j \widetilde{z}_j, \widetilde{w}_j)$ at $(\widetilde{z}, \widetilde{y})$. Note that due to (A.13) all first order terms cancel out.

Set $c_3 = c_2(2(p-1) + 2q)/c_1$, where c_1 is defined by Lemma A1, and let $b \geq b_\epsilon$ as well as $h \geq c_3 b$. Consider an arbitrary $(\theta, z, w) \in \bar{\mathbb{S}}_n$ with $(\theta, z, w) \notin C(x, y; hn^{-\frac{1}{p+q+1}})$, and assume that for $k \leq n$ there exist some $(\theta_1, z_1, y_1), \ldots, (\theta_{k-1}, z_{k-1}, y_{k-1}) \in \bar{\mathbb{S}}_n$ such that (A.1) holds with $(\theta_k, z_k, y_k) = (\theta, z, w)$. Lemma A1 then implies that there is a $(\widetilde{z}, \widetilde{y})$ with

$(1, \widetilde{z}, \widetilde{y}) \in C(x, y; \frac{b}{2(p-1)+2q} n^{-\frac{1}{p+q+1}})$ such that relations (A.2)–(A.3) are satisfied when b is replaced by $\frac{b}{2(p-1)+2q}$.

On the other hand, $\bar{\mathbb{S}}_n \cap D_{j,n} \neq \emptyset \; \forall \; j = 1, \ldots, 2(p-1)+2q$ imposes the existence of $2(p-1)+2q$ points $(\widetilde{\theta}_1, \widetilde{z}_1, \widetilde{w}_1) \in \bar{\mathbb{S}}_n \cap [1-b_n^2, 1] \times B_1, \ldots, (\widetilde{\theta}_{2(p-1)+q}, \widetilde{z}_{2(p-1)+q}, \widetilde{w}_{2(p-1)+q}) \in \bar{\mathbb{S}}_n \cap [1-b_n^2, 1] \times B_{2(p-1)+q}$. For some suitable $\gamma_1, \ldots, \gamma_{2(p-1)+q} \geq 0$ with $\sum_{j=1}^{2(p-1)+q} \gamma_j = 1$, we then obtain (A.13)–A.14), and one can conclude from (A.3) that

$$\begin{aligned} \sum_{r=1}^{k-1} \alpha_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0, y)} \quad &+ \quad \alpha_k \frac{g_x(\theta z, w)}{\theta g_x(0, y)} \\ &\geq \sum_{r=1}^{k-1} \widetilde{\alpha}_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0, y)} + \widetilde{\alpha}_k \frac{g_x(\widetilde{z}, \widetilde{y})}{g_x(0, y)} + \alpha_k \frac{c_1 c_3}{2(p-1)+2q} b^2 n^{-\frac{2}{p+q+1}} \\ &\geq \sum_{r=1}^{k-1} \widetilde{\alpha}_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0, y)} + \sum_{j=1}^{2(p-1)+2q} \widetilde{\alpha}_k \gamma_j \frac{g_x(\widetilde{\theta}_j \widetilde{z}_j, \widetilde{w}_j)}{\widetilde{\theta}_j g_x(0, y)}, \end{aligned} \tag{A.15}$$

where α_r, $\widetilde{\alpha}_r$ are defined as in Lemma A1. Clearly, $\sum_{r=1}^{k-1} \widetilde{\alpha}_r + \sum_{j=1}^{2(p-1)+2q} \widetilde{\alpha}_k \gamma_j = 1$ as well as $\sum_{r=1}^{k-1} \widetilde{\alpha}_r z_r + \sum_{j=1}^{2(p-1)+2q} \widetilde{\alpha}_k \gamma_j \widetilde{z}_j = 0$ and $\sum_{r=1}^{k-1} \widetilde{\alpha}_r y_r + \sum_{j=1}^{2(p-1+2q} \widetilde{\alpha}_k \gamma_j \widetilde{w}_j = y$.

Note that $(\widetilde{\theta}_j, \widetilde{z}_j, \widetilde{w}_j) \in \bar{\mathbb{S}}_n \cap C(x, y; hn^{-\frac{1}{p+q+1}})$ for all j. From (A.15), if $\bar{\mathbb{S}}_n \cap D_{j,n} \neq \emptyset \; \forall \; j$, then the minimal value of $\sum_i \alpha_i \frac{g_x(\theta_i Z_i, Y_i)}{\theta_i g_x(0,y)}$ over all $\alpha_1, \ldots, \alpha_n \geq 0$ with $\sum \alpha_i = 1$ is achieved by those linear combinations which assign zero weight $\alpha_i = 0$ to all observations with $(\theta, z, w) := (\theta_i, Z_i, Y_i) \notin C(x, y; hn^{-\frac{1}{p+q+1}})$. This leads to (A.12) and thus completes the proof of part (a).

In order to prove part (b) first note that (A.9)–(A.15) remain valid when defining $b = [(2c_2)^{-1} \log n]^{1/2}$ and $(\widetilde{z}, \widetilde{y}) = (0, y)$. By (A.10) and A.14) we can then infer that there is a constant d_0^* such that

$$\text{Prob}\left(\frac{\widehat{\theta}(x, y)}{\theta(x, y)} - 1 \leq n^{-\frac{2}{p+q+1}} \frac{\log n}{2} \right) \geq 1 - (2(p-1)+2q) \cdot \exp[-d_0^* (\log n)^{(p+q+1)/2}] \tag{A.16}$$

By Lemma 1 the above arguments can also be used to show that (A.16) holds for any point in a sufficiently small neighborhood $N(x, y)$ of (x, y). Using the continuity and convexity of θ and $\widehat{\theta}$, the asserted property of $\widehat{\theta}$ now follows from standard arguments based on interpolating a sufficiently fine grid of n points in $N(x, y)$. In view of Lemma 1(a) the assertion on $\widehat{g}_x$ is an immediate consequence. ∎

Proof of Theorem 2: Let

$$F_{x,h}(\delta) = \sum_{k=1}^{\infty} \text{Prob}\left(U\left[\frac{\delta}{h^2}, k\right]\right) \frac{\tau(h)^k \bar{f}_x(1,0,y)^k}{k!} e^{-\tau(h)\bar{f}_x(1,0,y)}$$

Clearly, $F_{x,h}$ is a continuous distribution function with $F_{x,h}(0) = 0$ and $F_{x,h}(\infty) = 1$. By definition of the respective events we obtain

$$\text{Prob}(A[\delta, n; h]) \leq \text{Prob}(A[\delta, n; h^*]) \leq \text{Prob}(A[\delta, n] \leq 1$$

for all δ, n and all $h^* > h$. From Proposition 1 $F_{x,h}(\delta) \leq F_{x,h^*}(\delta) \leq 1$ for any $\delta > 0$, implying that $\{F_{x,h}(\delta)\}_{h>0}$ is a bounded sequence of monotonically increasing real numbers and thus necessarily converges to a limit value. Together with Theorem 1(a) we can therefore conclude that there exists a monotone function $F_x(\delta)$ such that

$$F_x(\delta) =: \lim_{h\to\infty} F_{x,h}(\delta) = \lim_{n\to\infty} \text{Prob}(A[\delta, n]).$$

Clearly, F_x is a distribution function with $F_x(0) = 0$ and $F_x(\infty) = 1$.

It only remains to verify relation (3.18) as well as to show that F_x is continuous and that $F_x(\delta) < 1$. This requires a closer analysis of $\text{Prob}(U[\frac{\delta}{h^2}, k])$. There exists a $0 < d_0 < \infty$ such that for all $\gamma > 0$ and all sufficiently large k, $|\text{Prob}(U[\gamma, k]) - \text{Prob}(U[\gamma, k+1]| \leq d_0/k$. Consequently, if $[t]$ is the largest integer which is smaller or equal to t,

$$|\text{Prob}(U[\gamma, k]) - \text{Prob}(U[\gamma, [\lambda k]]| \leq d_0 \cdot \max\{\lambda - 1, \frac{1}{\lambda} - 1\} \tag{A.17}$$

holds for any $\gamma > 0$, $\lambda > 0$ and all sufficiently large k. Otherwise, for large h a Poisson distribution with parameter $\tau(h)\bar{f}_x(1,0,y)$ can be well-approximated by a $N(\tau(h)\bar{f}_x(1,0,y),\ \tau(h)\bar{f}_x(1,0,y))$-distribution. Combining these arguments reveals

$$\begin{aligned}
F_x(\delta) &= \lim_{h\to\infty} F_{x,h}(\delta) \\
&= \lim_{h\to\infty} \int \text{Prob}\left(U\left[\frac{\delta}{h^2}, \left[\sqrt{\tau(h)\bar{f}_x(1,0,y)}z + \tau(h)\bar{f}_x(1,0,y)\right]\right]\right) \phi(z)dz \\
&= \lim_{h\to\infty} \int \text{Prob}\left(U\left[\frac{\delta}{h^2}, \left[\left(1 + \frac{z}{\sqrt{\tau(h)\bar{f}_x(1,0,y)}}\right) \tau(h)\bar{f}_x(1,0,y)\right]\right]\right) \phi(z)dz \\
&= \lim_{h\to\infty} \int \text{Prob}\left(U\left[\frac{\delta}{h^2}, \left[\tau(h)\bar{f}_x(1,0,y)\right]\right]\right) \phi(z)dz \\
&= \lim_{h\to\infty} P\left(U\left[\frac{\delta}{h^2}, \left[\tau(h)\bar{f}_x(1,0,y)\right]\right]\right),
\end{aligned}$$

where ϕ denotes the standard normal density. Relation (3.18) then follows from

$$\lim_{h\to\infty} \text{Prob}\left(U\left[\frac{\delta}{h^2}, \left[\tau(h)\bar{f}_x(1,0,y)\right]\right]\right) = \lim_{k\to\infty} \text{Prob}\left(U\left[\delta\frac{\bar{f}_x(1,0,y)^{2/(p+q+1)}}{k^{2/(p+q+1)}}, k\right]\right),$$

and by using (3.16) the continuity of $F_x(\delta)$ for $\delta > 0$ follows from

$$\begin{aligned} |F_x(\lambda\delta) - F_x(\delta)| &= \lim_{k\to\infty} \Bigg|\text{Prob}\left(U\left[\delta\frac{\bar{f}_x(1,0,y)^{2/(p+q+1)}}{(k/\lambda^{(p+q+1)/2})^{2/(p+q+1)}}, \frac{k\lambda^{(p+q+1)/2}}{\lambda^{(p+q+1)/2}}\right]\right) \\ &\qquad -\text{Prob}\left(U\left[\delta\frac{\bar{f}_x(1,0,y)^{2/(p+q+1)}}{(k/\lambda^{(p+q+1)/2})^{2/(p+q+1)}}, \frac{k}{\lambda^{(p+q+1)/2}}\right]\right)\Bigg| \\ &\leq d_0 \cdot \max\{\lambda^{(p+q+1)/2} - 1, \frac{1}{\lambda^{(p+q+1)/2}} - 1\}. \end{aligned}$$

Clearly, the event $U\left[\delta\frac{\bar{f}_x(1,0,y)^{2/(p+q+1)}}{k^{2/(p+q+1)}},\ k\right]$ implies that $(\widetilde{\vartheta}_j, \widetilde{\zeta}_j, \widetilde{y}_j) \in I_{k,\delta} :=$ $\left[0, \delta\frac{\bar{f}_x(1,0,y)^{2/(p+q+1)}}{k^{2/(p+q+1)}}\right] \times \left[\frac{-1}{k^{1/(p+q+1)}}, \frac{1}{k^{1/(p+q+1)}}\right]^{p-1} \times \left[\frac{-1}{k^{1/(p+q+1)}}, \frac{1}{k^{1/(p+q+1)}}\right]^q$ for at least one observation $j \in \{1, \ldots, k\}$. Since $\text{Prob}(I_{k,\delta}) = \delta\frac{\bar{f}_x(1,0,y)^{2/(p+q+1)}}{k}$ for all sufficiently large k, standard arguments now lead to

$$\begin{aligned} \text{Prob}\left(U\left[\delta\frac{\bar{f}_x(1,0,y)^{2/(p+q+1)}}{k^{2/(p+q+1)}},\ k\right]\right) &\leq \text{Prob}\Big((\widetilde{\vartheta}_j, \widetilde{\zeta}_j, \widetilde{Y}_j) \in I_{k,\delta} \text{ for some } j \in \{1, \ldots, k\}\Big) \\ &= 1 - \exp(-\delta\bar{f}_x(1,0,y)^{2/(p+q+1)}) \quad \text{as } k \to \infty. \end{aligned}$$

Consequently F_x is continuous at $\delta = 0$, and $F_x(\delta) < 1$ for all $\delta > 0$. ■

Proof of Theorem 4: Recall the definitions of the events $A[\delta, n; h]$ and $A[\delta, n]$. Replace (θ_i, Z_i, Y_i) by $(\theta_i^*, Z_i^*, Y_i^*)$ and g_x by $\widehat{g}_x^*$ to define events $A[\delta, n; h]^*$ and $A[\delta, n]^*$. First, note that for all n,

$$\text{Prob}\left(n^{\frac{2}{p+q+1}}\left(\frac{\widehat{\theta}^*(x,y)}{\widehat{\theta}(x,y)} - 1\right) \leq \delta \mid \mathcal{S}_n\right) = \text{Prob}(A[\delta, n]^* \mid \mathcal{S}_n).$$

Conditional on $\mathcal{S}_n$, the essential parts of the arguments used in the proofs of Lemma A1 and Theorem 1 remain valid when applied to $\widehat{g}_x^*$ and $\widehat{f}_x$ instead of g_x and f_x. This is easily seen when noting that $\widehat{g}_x^*$ is necessarily convex and that with probability converging to 1 as $n \to \infty$ the bounds given in (A.8) and (A.15) also apply to $\widehat{g}_x^*$. Since $n^{-\frac{1}{p+q+1}}/b \to 0$, the latter follows from (4.5) and Taylor expansions of g_x^* similar to (4.6). Furthermore, due to

(4.7) relations (A.10)–(A.12) generalize to $\mathcal{S}_n^*$ and $\widehat{f}_x$. Therefore for any $\epsilon > 0$ there exists a $h_\epsilon > 0$ such that for all $h \geq h_\epsilon$,

$$\text{Prob}\left(\sup_\delta \left[\text{Prob}(A[\delta, n]^* \mid \mathcal{S}_n) - \text{Prob}(A[\delta, n, h]^* \mid \mathcal{S}_n)\right] \leq \epsilon\right) \to 1 \quad \text{as } n \to \infty. \tag{A.18}$$

On the other hand, in view of (4.5)–(4.7), one can invoke arguments similar to those used in the proof of Proposition 1 to obtain

$$\sup_\delta \Bigg| \text{Prob}(A[\delta, n, h]^* \mid \mathcal{S}_n) - \sum_{k=1}^{\infty} \text{Prob}\left(U\left[\tfrac{\delta}{h^2},\ k\right]\right) \frac{\tau(h)^k \bar{f}_x(1,0,y)^k}{k!} e^{-\tau(h)\bar{f}_x(1,0,y)} \Bigg| = o_p(1). \tag{A.19}$$

The theorem now follows from Theorem 2. ■

References

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

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

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

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

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

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," *Journal of the Royal Statistical Society A,* 120, 253–281.

Gattoufi, S., Oral, M., and Reisman, A. (2004), "Data envelopment analysis literature: A bibliography update (1951–2001)," *Socio-Economic Planning Sciences,* 38, 159–229.

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

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

Kneip, A., Park, B., and Simar, L. (1998), "A note on the convergence of nonparametric DEA efficiency measures," *Econometric Theory,* 14, 783–793.

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

Sheather, S.J., and Jones, M.C. (1991), "A reliable data-based bandwidth selection method for kernel density estimation," *Journal of the Royal Statistical Society, Series B,* 53, 683–690.

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

Simar, L. (1996), "Aspects of statistical analysis in DEA-type frontier models," *Journal of Productivity Analysis,* 7, 177–185.

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

—— (1999a), "Of course we can bootstrap DEA scores! but does it mean anything? logic trumps wishful thinking," *Journal of Productivity Analysis,* 11, 93–97.

——— (1999b), “Some problems with the ferrier/hirschberg bootstrap idea,” *Journal of Productivity Analysis,* 11, 67–80.

——— (2000a), “A general methodology for bootstrapping in non-parametric frontier models,” *Journal of Applied Statistics,* 27, 779–802.

——— (2000b), “Statistical inference in nonparametric frontier models: The state of the art,” *Journal of Productivity Analysis,* 13, 49–78.

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

Table 1: Coverage of CIs Estimated by Sub-Sampling

		$p = q = 1$			$p = q = 2$		
		$(1-\alpha)$			$(1-\alpha)$		
n	κ	.90	.95	.99	.90	.95	.99
25	0.50	0.949	0.976	0.986	0.934	0.967	0.993
25	0.55	0.958	0.978	0.993	0.934	0.966	0.991
25	0.60	0.948	0.970	0.993	0.899	0.951	0.990
25	0.65	0.949	0.984	0.999	0.891	0.940	0.988
25	0.70	0.945	0.963	0.989	0.822	0.892	0.975
25	0.75	0.927	0.966	0.988	0.779	0.868	0.964
25	0.80	0.920	0.967	0.990	0.704	0.808	0.935
25	0.85	0.908	0.952	0.991	0.641	0.752	0.909
25	0.90	0.877	0.926	0.972	0.567	0.681	0.853
25	0.95	0.872	0.922	0.972	0.499	0.618	0.821
25	1.00	0.801	0.879	0.956	0.419	0.529	0.737
50	0.50	0.975	0.990	1.000	0.968	0.988	0.998
50	0.55	0.974	0.990	0.998	0.943	0.982	0.998
50	0.60	0.969	0.989	0.994	0.920	0.962	0.996
50	0.65	0.968	0.984	0.997	0.874	0.926	0.983
50	0.70	0.956	0.980	0.995	0.834	0.918	0.979
50	0.75	0.952	0.976	0.994	0.766	0.847	0.942
50	0.80	0.928	0.962	0.990	0.713	0.787	0.904
50	0.85	0.902	0.952	0.988	0.636	0.723	0.864
50	0.90	0.905	0.947	0.988	0.533	0.629	0.798
50	0.95	0.857	0.913	0.971	0.437	0.536	0.738
50	1.00	0.827	0.884	0.964	0.384	0.476	0.665
100	0.50	0.975	0.994	0.999	0.962	0.989	1.000
100	0.55	0.978	0.997	1.000	0.935	0.972	0.998
100	0.60	0.981	0.992	0.999	0.905	0.953	0.986
100	0.65	0.979	0.991	0.998	0.887	0.940	0.981
100	0.70	0.976	0.990	0.999	0.842	0.890	0.961
100	0.75	0.965	0.983	0.998	0.787	0.864	0.948
100	0.80	0.939	0.968	0.994	0.688	0.768	0.894
100	0.85	0.914	0.954	0.985	0.639	0.732	0.854
100	0.90	0.890	0.934	0.985	0.520	0.624	0.775
100	0.95	0.808	0.895	0.962	0.461	0.567	0.720
100	1.00	0.775	0.833	0.938	0.371	0.473	0.645

Table 1: (continued)

		$p = q = 1$			$p = q = 2$		
		$(1-\alpha)$			$(1-\alpha)$		
n	κ	.90	.95	.99	.90	.95	.99
200	0.50	0.975	0.991	0.999	0.945	0.985	0.999
200	0.55	0.983	0.996	1.000	0.951	0.981	0.996
200	0.60	0.985	0.997	1.000	0.941	0.971	0.998
200	0.65	0.984	0.996	0.999	0.910	0.938	0.985
200	0.70	0.973	0.991	0.999	0.863	0.913	0.973
200	0.75	0.963	0.981	1.000	0.770	0.850	0.936
200	0.80	0.926	0.971	0.995	0.699	0.788	0.904
200	0.85	0.901	0.948	0.993	0.641	0.725	0.871
200	0.90	0.837	0.914	0.976	0.534	0.633	0.791
200	0.95	0.805	0.876	0.965	0.418	0.518	0.693
200	1.00	0.733	0.821	0.945	0.348	0.435	0.645
400	0.50	0.968	0.993	0.999	0.964	0.996	1.000
400	0.55	0.986	0.996	0.999	0.957	0.983	0.996
400	0.60	0.985	0.995	1.000	0.954	0.983	0.999
400	0.65	0.981	0.997	1.000	0.897	0.948	0.987
400	0.70	0.965	0.992	0.999	0.861	0.912	0.971
400	0.75	0.953	0.983	0.994	0.795	0.873	0.955
400	0.80	0.933	0.967	0.998	0.695	0.798	0.915
400	0.85	0.890	0.937	0.985	0.623	0.741	0.876
400	0.90	0.809	0.903	0.971	0.519	0.608	0.785
400	0.95	0.768	0.842	0.948	0.398	0.518	0.706
400	1.00	0.714	0.791	0.902	0.311	0.398	0.573
800	0.50	0.946	0.989	0.995	0.944	0.985	0.998
800	0.55	0.972	0.996	0.998	0.954	0.987	0.998
800	0.60	0.971	0.992	0.998	0.961	0.981	0.995
800	0.65	0.962	0.991	0.999	0.924	0.964	0.988
800	0.70	0.971	0.991	0.998	0.855	0.909	0.975
800	0.75	0.951	0.973	1.000	0.807	0.877	0.961
800	0.80	0.890	0.946	0.992	0.708	0.789	0.922
800	0.85	0.873	0.929	0.978	0.611	0.727	0.863
800	0.90	0.814	0.891	0.968	0.477	0.592	0.773
800	0.95	0.751	0.821	0.927	0.383	0.483	0.653
800	1.00	0.695	0.779	0.902	0.262	0.356	0.548

Table 2: Coverage of CIs Estimated by Double-Smooth Bootstrap

		$p = q = 1$			$p = q = 2$		
		$(1-\alpha)$			$(1-\alpha)$		
n	b	.90	.95	.99	.90	.95	.99
25	0.4	0.793	0.869	0.953	—	—	—
50	0.4	0.831	0.911	0.976	—	—	—
100	0.4	0.870	0.931	0.973	0.672	0.781	0.937
200	0.4	0.907	0.964	0.994	0.678	0.814	0.955
400	0.4	0.910	0.957	0.991	0.762	0.849	0.952
800	0.4	0.937	0.971	0.997	0.763	0.859	0.962
25	0.6	0.810	0.883	0.961	0.456	0.589	0.831
50	0.6	0.861	0.927	0.978	0.643	0.750	0.899
100	0.6	0.888	0.934	0.978	0.722	0.815	0.939
200	0.6	0.916	0.968	0.995	0.746	0.856	0.962
400	0.6	0.913	0.959	0.989	0.808	0.887	0.965
800	0.6	0.916	0.966	0.995	0.821	0.884	0.970
25	0.8	0.833	0.900	0.962	0.641	0.753	0.900
50	0.8	0.868	0.936	0.981	0.665	0.770	0.908
100	0.8	0.881	0.933	0.980	0.744	0.848	0.950
200	0.8	0.907	0.962	0.996	0.794	0.877	0.965
400	0.8	0.892	0.950	0.986	0.808	0.887	0.967
800	0.8	0.882	0.938	0.993	0.813	0.887	0.968
25	1.0	0.844	0.913	0.977	0.667	0.770	0.904
50	1.0	0.871	0.933	0.981	0.684	0.786	0.910
100	1.0	0.878	0.927	0.981	0.760	0.855	0.950
200	1.0	0.891	0.949	0.994	0.793	0.866	0.959
400	1.0	0.866	0.923	0.982	0.792	0.864	0.955
800	1.0	0.855	0.914	0.986	0.773	0.848	0.950

Figure 1: Illustration of g_x for the case $p = 2$