Econometric Theory, **24**, 2008, 1663–1697. Printed in the United States of America.
doi:10.1017/S0266466608080651

ASYMPTOTICS AND CONSISTENT BOOTSTRAPS FOR DEA ESTIMATORS IN NONPARAMETRIC FRONTIER MODELS

ALOIS KNEIP
Universität Bonn

LÉOPOLD SIMAR
Université Catholique de Louvain

PAUL W. WILSON
Clemson University

Nonparametric 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 proved consistent, making inference doubtful. This paper derives the asymptotic distribution of DEA estimators under variable returns to scale. This result is used to prove consistency of two different bootstrap procedures (one based on subsampling, the other based on smoothing). The smooth bootstrap requires smoothing the irregularly bounded density of inputs and outputs and smoothing 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 input levels and the mix of outputs in the case of output-oriented measures.

1. INTRODUCTION

Nonparametric data envelopment analysis (DEA) estimators based on the original ideas of Debreu (1951), Farrell (1957), and Shephard (1970) and employing linear programming methods along the lines of Charnes, Cooper, and Rhodes (1978, 1979) and Färe, Grosskopf, and Lovell (1985) have been widely applied. Until recently, however, little was known about their statistical properties. Under certain assumptions, the DEA *frontier* estimator is a consistent, maximum likelihood estimator (Banker, 1993) with a known rate of convergence (Koroste-

Research support from the Texas Advanced Computation Center and from the "Interuniversity Attraction Pole," Phase V (P5/24) and Phase IV (P6/03) from the Belgian government (Belgian Science Policy) is gratefully acknowledged. Address correspondence to Paul W. Wilson, Department of Economics, 222 Sirrine Hall, Clemson University, Clemson, South Carolina 29634, USA; e-mail: pww@clemson.edu.

 0266-4666/08 $15.00

lev, Simar, and Tsybakov, 1995). In addition, consistency and convergence rates of DEA *efficiency* estimators have been established (Kneip, Park, and Simar, 1998; for a survey of recent developments regarding statistical properties of DEA estimators, see Simar and Wilson, 2000b). But until now, the asymptotic distribution of DEA efficiency estimators has remained unknown except for the limited case of one input, one output derived by Gijbels, Mammen, Park, and Simar (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 deriving 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—one based on subsampling, the other on smoothing—yield consistent inference.

It is not surprising that a bootstrap based on subsampling would work in the DEA context; Swanepoel (1986) discussed this approach for inference about the boundary of support for a univariate distribution. Our simulation results presented in Section 4 indicate that the choice of the size of the subsamples is critical; suboptimal choices can be catastrophic for realized coverages of estimated confidence intervals. Unfortunately, there seems to be no reliable way of determining a reasonable value of the subsample size in applied settings. Experimentation with an iterated subsampling bootstrap has proved almost useless; for any realistic original sample size, the inner bootstrap loops contain too few observations to provide useful information on the "optimal" subsample size.

The second bootstrap approach provides a tractable approach to inference with DEA estimators but at a cost of increased complexity over the subsampling 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 that the sample size n is sufficiently large for the given dimensionality of the problem (this caveat should be of no surprise, because it is now well known that the curse of dimensionality affects the quality of the initial DEA point estimates; again, for discussion, see Simar and Wilson, 2000b).

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

$$\Psi = \{(x, y) | x \quad \text{can produce } y\} \tag{1}$$

may be described in terms of its sections

$$\mathcal{Y}(x) \equiv \{y | (x, y) \in \Psi\} \tag{2}$$

and

$$\mathcal{X}(y) \equiv \{x | (x, y) \in \Psi\}, \tag{3}$$

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

$$\mathcal{X}^{\partial}(y) = \{x | (x, y) \in \Psi, \qquad (\delta x, y) \notin \Psi \quad \forall\, \delta < 1\}. \tag{4}$$

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 $\tilde{x} \geq x$, $\tilde{y} \leq y$, if $(x, y) \in \Psi$ then $(\tilde{x}, y) \in \Psi$ and $(x, \tilde{y}) \in \Psi$; 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 $\tilde{x}, x \in \mathbb{R}_+^p$, $\tilde{x} \geq x$ means that some number $\ell \in \{0, 1, \ldots, p\}$ of the corresponding elements of $\tilde{x}$ and x are equal, whereas $(p - \ell)$ of the elements of $\tilde{x}$ 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

$$\theta(x, y) \equiv \inf\{\delta | (\delta x, y) \in \Psi, \qquad \delta > 0\} \tag{5}$$

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

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. To conserve space, we consider only the input orientation; all of our results extend to output-oriented measures after 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 data generating process (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 (i.i.d.) random variables on the convex attainable set Ψ.

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

Clearly, Assumption 5(c) imposes a discontinuity in f at frontier points where $\theta(x, y) = 1$, ensuring a significant, nonnegligible probability of observing production units close to the production frontier. For technically nonattainable points that lie outside Ψ, $f \equiv 0$. In most practical situations, $\mathcal{D} = \Psi$; however, Assumption 5 does not exclude the possibility that $\mathcal{D}$ is a strict subset of Ψ.

Assumption 6. The function $\theta(x, y)$ is twice continuously differentiable for all $(x, y) \in \mathcal{D}$.

By definition of θ, we obtain $\theta(x, y) = \lambda\theta(\lambda x, y)$ for any $\lambda > 0$. Hence, if θ is twice continuously differentiable at some point (x, y) in the interior of $\mathcal{D}$, it is also twice continuously differentiable at any point $(\lambda x, y)$ for arbitrary λ. Essentially, Assumption 6 only requires that the boundary of Ψ is sufficiently smooth.

To illustrate Assumptions 5 and 6 let us consider the case of a single input, $p = 1$. Then $x \in \mathbb{R}_+$, and there exists a well-defined frontier function $g(y) := \inf\{x | (x, y) \in \Psi\}$, i.e., the well-known *production function* that for each output y gives the corresponding efficient input $g(y)$. Consequently, $\theta(x, y) = g(y)/x$, and Assumption 6 is satisfied if the production function $g(y)$ is twice continuously differentiable. Observations (X_i, Y_i) may be rewritten in the form $X_i = g(Y_i) + \epsilon_i$, where $\epsilon_i = X_i - g(Y_i) \geq 0$. Input-oriented parametric approaches to frontier analysis then usually rely on explicit modeling of the structure of g and of the distribution of ϵ_i. For example, it is frequently assumed that ϵ_i possesses a half-normal distribution and that ϵ_i is independent of Y_i. Let ϕ_+ denote the density of the half-normal probability density function, with $\phi_+(v) = 0$ for $v < 0$, $\phi_+(v) > 0$ for $v \geq 0$. If, in addition, Y_i possesses a continuous density $\tilde{f}$ with $\tilde{f}(y) > 0$ for all $y \in \mathbb{R}_+^q$, then x and y have

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

joint density $f(x, y) = \phi_+(x - g(y))\tilde{f}(y)$. Then $f(\theta(x, y)x, y) = f(g(y), y) = \phi_+(0)\tilde{f}(y) > 0$ for all $(x, y) \in \Psi$, and Assumption 5 holds with $\mathcal{D} = \Psi$. Of course, this only constitutes a particular example with $p = 1$. Assumption 5 will be fulfilled in much more complex situations.

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 $\mathcal{S}_n$ and is given by

$$\hat{\Psi} = \{(x, y) | y \leq Ya, \quad x \geq Xa, \quad i'a = 1, \quad a \in \mathbb{R}^n_+\}, \tag{6}$$

where $Y = [y_1 \dots y_n]$, $X = [x_1 \dots x_n]$, i denotes an $(n \times 1)$ vector of ones, and a is an $(n \times 1)$ vector of intensity variables. The corresponding DEA estimator of $\theta(x, y)$ is obtained by replacing Ψ with $\hat{\Psi}$ in (5); i.e.,

$$\hat{\theta}(x, y) = \min\{\delta > 0 | y \leq Ya, \quad \delta x \geq Xa, \quad i'a = 1, \quad a \in \mathbb{R}^n_+\}. \tag{7}$$

Minimization of the linear program in (7) provides a solution for both δ and a. Whereas $\theta(x, y)$ defined in (5) gives a measure of distance from a point $(x, y) \in \mathbb{R}^{p+q}_+$ to the boundary of Ψ, $\hat{\theta}(x, y)$ measures distance from the same point to the boundary of the convex hull of the free-disposal hull of the n sample observations. Note that necessarily $\hat{\Psi} \subset \Psi$ and hence $\hat{\theta}(x, y) \geq \theta(x, y)$ for all (x, y). The statistical performance of the DEA estimator $\hat{\theta}(x, y)$ of $\theta(x, y)$ depends on the smoothness of the frontier. Kneip et al. (1998) derive different rates of convergence depending on the degree of smoothness. Per Assumption 6, we consider only the case where $\theta(x, y)$ is twice-differentiable. For this case, Kneip et al. (1998) prove that $\hat{\theta}(x, y) = \theta(x, y) + O_p(n^{-2/(p+q+1)})$; as with many nonparametric estimators, DEA estimators suffer from the curse of dimensionality.

2. 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 and provides results on uniform convergence. Theorem 2 establishes the asymptotic distribution. It thus provides a basis to prove consistency of the bootstrap methods that are given in Section 3. All proofs are given in the Appendix.

Before actually stating our results some conceptual work has to be done. Recall the definition of the "frontier" $\mathcal{X}^\partial(y)$ defined in (4) that establishes the sets of all technologically feasible, efficient input vectors for a given output vector y. As discussed previously, we have $\mathcal{X}^\partial(y) = \{g(y)\}$ if $p = 1$. A basic problem when dealing with *multiple* inputs and outputs is the nonexistence of a unique, well-defined production function $g(y)$. If $p > 1$, then $\mathcal{X}^\partial(y)$ defined in (4) contains a set of efficient input vectors corresponding to the output level y. Inefficient points with output level y can be projected onto $\mathcal{X}^{\partial}(y)$; e.g., for two linearly independent vectors x^*, x^{**} we have $\theta(x^*, y)x^* \in \mathcal{X}^{\partial}(y)$ and $\theta(x^{**}, y)x^{**} \in \mathcal{X}^{\partial}(y)$, but $\theta(x^*, y)x^* \neq \theta(x^{**}, y)x^{**}$.

It is then only possible to characterize the frontier as a function of y *and* of suitable coordinates of input vectors. There are infinitely many coordinate systems, and we will concentrate on a decomposition of the vectors X_i of inputs that is specific for a particular point of interest $x \in \mathbb{R}_+^p$. 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

$$X_i = \gamma_i \frac{x}{\|x\|} + Z_i \quad \text{for some } Z_i \in \mathcal{V}(x) \quad \text{and} \quad \gamma_i = \frac{x^T X_i}{\|x\|}, \tag{8}$$

where $\|\cdot\|$ denotes the euclidean norm. Let $\Psi^*(x) = \mathcal{V}(x) \times \mathbb{R}_+^q$ and note that the point of interest $(x, y) \in \Psi$ has coordinates $(0, y)$ in $\Psi^*(x)$. We can infer from Assumption 3 that for any $(z, y) \in \Psi^*(x)$ there exists $\gamma > 0$ such that $(\gamma(x/\|x\|) + z, y) \in \Psi$. The boundary of Ψ can thus be described through the following function defined for any $(z, y) \in \Psi^*(x)$:

$$g_x(z, y) = \inf\left\{\gamma \middle| \left(\gamma \frac{x}{\|x\|} + z, y\right) \in \Psi\right\}. \tag{9}$$

This definition implies that

$$\Psi = \left\{\left(\gamma \frac{x}{\|x\|} + z, y\right) \middle| (z, y) \in \mathcal{V}(x), \gamma \geq g_x(z, y)\right\} \tag{10}$$

and

$$\mathcal{X}^{\partial}(y) = \left\{g_x(z, y) \frac{x}{\|x\|} + z \middle| z \in \mathcal{V}(x)\right\}. \tag{11}$$

Thus g_x may be interpreted as a "frontier function" that characterizes the frontier $\mathcal{X}^{\partial}(y)$ in the coordinate system (z, y). For any $v \in \mathbb{R}_+^p$ we have $v = (x^T v/\|x\|^2)x + z$ for $z = v - (x^T v/\|x\|^2)x$. Because $\theta(v, y)v \in \mathcal{X}^{\partial}(y)$ for all $(v, y) \in \Psi$, (10)–(11) yield

$$\theta(v, y) \frac{x^T v}{\|x\|} = g_x(\theta(v, y)z, y) \quad \text{and} \quad \theta(x, y) = \frac{g_x(0, y)}{\|x\|}. \tag{12}$$

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

$$\hat{g}_x(z, y) = \inf\left\{\gamma \middle| \left(\gamma \frac{x}{\|x\|} + z, y\right) \in \hat{\Psi}\right\} \tag{13}$$

and

$$\hat{\theta}(v, y)\, \frac{x^T v}{\|x\|} = \hat{g}_x(\hat{\theta}(v, y) z, y) \quad \text{and} \quad \hat{\theta}(x, y) = \frac{\hat{g}_x(0, y)}{\|x\|}. \tag{14}$$

Finally, with only a small abuse of notation, one may extend the definition of g_x to all (v, y) with $(v - (x^T v/\|x\|^2)x, y) \in \Psi^*(x)$ by taking $g_x(v, y) = g_x(v - (x^T v/\|x\|^2)x, y)$. Note that in this notation, because of (12), $g_x(v, y)$ is continuous in x.

In the case of one input ($p = 1$), the function g_x is simply the production 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 want to emphasize that the preceding relations hold for an *arbitrary* choice of x. When using a different base vector $x^* \neq x$, then for all possible y the alternative "frontier" function g_{x^*} will describe the same frontier $\mathcal{X}^\partial(y)$ in a different coordinate system, and (12) remains true when replacing x by x^*. Also note that $g_x = g_{\lambda x}$ for $\lambda > 0$.

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 effi-

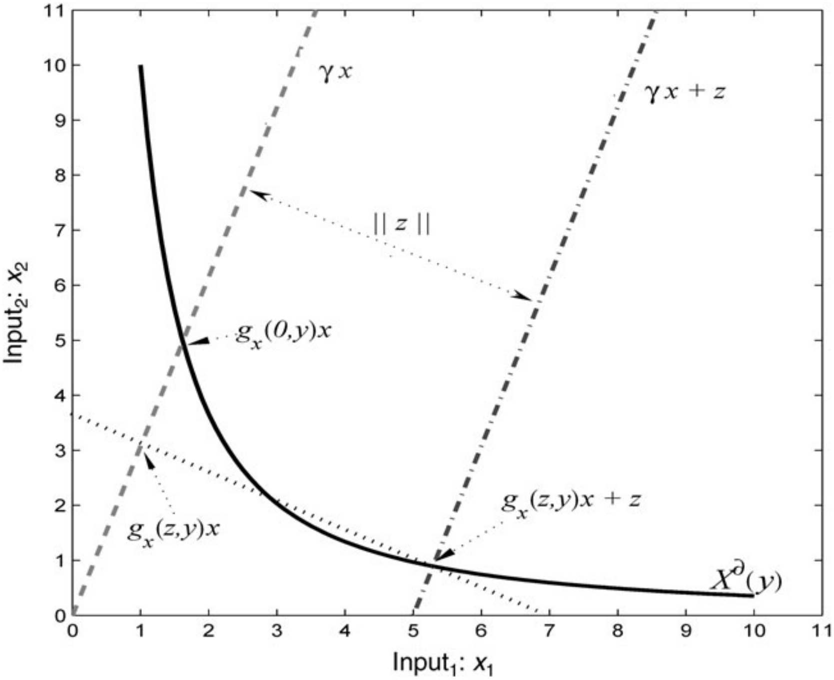

FIGURE 1. Illustration of g_x for the case $p = 2$ for $\|x\| = 1$.

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

ciency boundary $\mathcal{X}^{\partial}(y)$, shown by the solid curve. Now consider an input vector x with $\|x\| = 1$. The ray γx, $\gamma \geq 0$, is represented by the dashed line passing through the origin. For a vector z with $z^T x = 0$, $\gamma x + z$ represented by the dash-dot-dash line is parallel to γx. The intersection between $\gamma x + z$ and $\mathcal{X}^{\partial}(y)$ then determines the point $g_x(z, y)x + z$.

Let $z^{(1)}, \ldots, z^{(p-1)}$ be an orthonormal basis of $\mathcal{V}(x)$. Every vector $Z \in \mathcal{V}(x)$ can be expressed in the form $Z = \sum_{j=1}^{p-1} \zeta_j z^{(j)}$. Let $\zeta_i = (\zeta_{i1}, \ldots, \zeta_{i,p-1})$. Because $\theta_i = \theta(X_i, Y_i)$ and $Z_i = X_i - (x^T X_i / \|x\|^2)x$ 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.

For any point (x, y) of interest the corresponding frontier function g_x and the density $\bar{f}_x$ will provide the basic tools for our theoretical study. Roughly speaking, our approach provides a decomposition into a function g_x characterizing the technological frontier and a density $\bar{f}_x$ describing the distribution of observations *relative* to the frontier. This corresponds to the strategy adopted by all existing parametric frontier models. If $\mathcal{D} = \Psi$ then $\bar{\mathcal{D}} = (0,1] \times \mathbb{R}^{p-1} \times \mathbb{R}_+^q$, and if f is continuous on Ψ then $\bar{f}_x$ is continuous on its whole domain $(0,1] \times \mathbb{R}^{p-1} \times \mathbb{R}_+^q$. Nonparametric estimates of g_x and f_x will play a key role in the derivation of the smoothed bootstrap of Section 3.2.

The following lemma summarizes the most important properties of g_x and f_x.

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

(i) g_x and $\hat{g}_x$ are convex functions;
(ii) the function $g_x(z, y)$ is twice continuously differentiable for all points $(z, y) \in \Psi^*(x)$; 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 $\|w\| = 1$; moreover, $g_x''(0, y)$ changes continuously in x;
(iii) $\bar{f}_x(\cdot, \cdot, \cdot)$ is continuous on $\bar{\mathcal{D}}$, and $\bar{f}_x(1, 0, y) > 0$; furthermore, $\bar{f}_x(1, 0, y)$ changes continuously in 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 $\hat{\theta}(x, y)/\theta(x, y)$ 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 $\delta > 0$*

$$\text{Prob}\left(\frac{\hat{\theta}(x, y)}{\theta(x, y)} - 1 \leq \delta n^{-2/(p+q+1)}\right) = \text{Prob}(A[\delta, n]), \qquad \textbf{(15)}$$

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

$$\sum_{i=1}^{n} \alpha_i Z_i = 0, \qquad \sum_{i=1}^{n} \alpha_i Y_i = y, \quad \text{and}$$

$$\sum_{i=1}^{n} \alpha_i \frac{g_x(\theta_i Z_i, Y_i)}{\theta_i g_x(0, y)} - 1 \leq \delta n^{-2/(p+q+1)}, \tag{16}$$

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

Theorem 1, which follows, provides a basis for our theoretical development and for the construction of bootstrap procedures in Section 3. Although Theorem 1(ii) provides results on uniform convergence on suitable subsets, Theorem 1(i) plays an important role by "localizing" the frontier problem. The value of $\hat{\theta}(x, y)$ is essentially determined by those observations that 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, e.g., the frontier is linear or conical, because in such cases $\hat{\theta}(x, y)$ may be determined by points very far from the point of interest (x, y).

More precisely, it will be shown that under our assumptions $\text{Prob}(A[\delta, n])$ asymptotically coincides with the probabilities of events $A[\delta, n; h]$, which only depend on observations in a suitable neighborhood of the point (x, y) of interest. Note that the sample of observations $\mathcal{S}_n$ can be represented equivalently by the corresponding samples $\tilde{\mathcal{S}}_n = \{(\theta_i, Z_i, Y_i)\}_{i=1}^n$ 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

$$C(x, y; h) = \Big\{ (\theta, \tilde{z}, \tilde{y}) \in (0, 1] \times \Psi^*(x) \mid 1 - \theta \leq h^2,$$
$$z = \sum_j \zeta_j z^{(j)} \quad \text{with } |\zeta_j| \leq h \quad \forall\, j = 1, \dots, p-1,$$
$$|y_r - \tilde{y}_r| \leq h \quad \forall\, r = 1, \dots, q \Big\}. \tag{17}$$

The point $(1, 0, y)$ in the transformed space $\{(\theta(v, \tilde{y}), v - (x^T v / \|x\|^2)x, \tilde{y}) | (v, \tilde{y}) \in \Psi\}$ 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 $k \leq n$ and $i_1, \dots, i_k \in \{1, \dots, n\}$, there exist some $(X_{i_1}, Y_{i_1}), \dots, (X_{i_k}, Y_{i_k})$ with $(\theta_{i_1}, Z_{i_1}, Y_{i_1}), \dots, (\theta_{i_k}, Z_{i_k}, Y_{i_k}) \in \tilde{\mathcal{S}}_n \cap C(x, y; h \cdot n^{-1/(p+q+1)})$ and some $\alpha_1 \geq 0, \dots, \alpha_k \geq 0$ 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

$$\sum_{j=1}^{k} \alpha_j \frac{g_x(\theta_{i_j} Z_{i_j}, Y_{i_j})}{\theta_{i_j} g_x(0, y)} - 1 \leq \delta n^{-2/(p+q+1)}. \tag{18}$$

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

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

(i) for any $\epsilon > 0$ there exists an $h_\epsilon < \infty$ such that for all $h \geq h_\epsilon$, every $\delta > 0$, and all sufficiently large n,

$$|\text{Prob}(A[\delta, n]) - \text{Prob}(A[\delta, n; h])| \leq \epsilon; \tag{19}$$

(ii) for any closed, bounded subset $\mathcal{N}$ of the interior of $\mathcal{D}$ with $\inf_{\tilde{x}, \tilde{y} \in \mathcal{N}} f(\tilde{x}, \tilde{y}) > 0$,

$$\text{Prob}\left(\sup_{(\tilde{x}, \tilde{y}) \in \mathcal{N}} \left| \frac{\hat{\theta}(\tilde{x}, \tilde{y})}{\theta(\tilde{x}, \tilde{y})} - 1 \right| \leq n^{-2/(p+q+1)} (\log n)^{2/(p+q+1)} \right) \to 1 \quad \text{as } n \to \infty \tag{20}$$

and

$$\text{Prob}\left(\sup_{(\tilde{x}, \tilde{y}) \in \mathcal{N}} \left| \frac{\hat{g}_x\left(\theta(\tilde{x}, \tilde{y})\left(\tilde{x} - \frac{x^T \tilde{x}}{\|x\|^2}\right)x, \tilde{y}\right)}{g_x\left(\theta(\tilde{x}, \tilde{y})\left(\tilde{x} - \frac{x^T \tilde{x}}{\|x\|^2}\right)x, \tilde{y}\right)} - 1 \right| \leq n^{-2/(p+q+1)} (\log n)^{2/(p+q+1)} \right) \to 1 \quad \text{as } n \to \infty. \tag{21}$$

To examine the probabilities $P(A[\delta, n; h])$, still more notation is required. Let $(\tilde{\vartheta}_1, \tilde{\zeta}_1, \tilde{y}_1), (\tilde{\vartheta}_2, \tilde{\zeta}_2, \tilde{y}_2), \ldots$ denote a sequence of i.i.d. 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

$$\sum_{j=1}^{k} \alpha_j \tilde{y}_j = 0 \quad \text{and} \quad \sum_{j=1}^{k} \alpha_j \tilde{z}^{(j)} = 0, \tag{22}$$

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

$$\sum_{j=1}^{k} \alpha_j \frac{1}{2 g_x(0, y)} [\tilde{z}_j^T g''_{x;zz}(0, y) \tilde{z}_j + 2 \tilde{z}_j^T g''_{x;zy}(0, y) \tilde{y}_j + \tilde{y}_j^T g''_{x;yy}(0, y) \tilde{y}_j] + \sum_{j=1}^{k} \alpha_j \tilde{\vartheta}_j \leq \gamma. \tag{23}$$

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

Here we use

$$g''(x;0,y) = \begin{bmatrix} g''_{x;zz}(0,y) & g''_{x;zy}(0,y)^T \\ g''_{x;zy}(0,y) & g''_{x;yy}(0,y) \end{bmatrix} \tag{24}$$

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,*

$$\left| \text{Prob}(A[\delta,n;h]) - \sum_{k=1}^{\infty} \text{Prob}\left(U\left[\frac{\delta}{h^2},k\right]\right) \frac{\tau(h)^k \bar{f}_x(1,0,y)^k}{k!} e^{-\tau(h)\bar{f}_x(1,0,y)} \right| \to 0 \tag{25}$$

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

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

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

$$F_x(\delta) = \lim_{k\to\infty} \text{Prob}\left(U\left[\delta\, \frac{\bar{f}_x(1,0,y)^{2/(p+q+1)}}{k^{2/(p+q+1)}}, k\right]\right) \tag{26}$$

for $-\infty < \delta < \infty$. Then F_x is a continuous distribution function with $F_x(0) = 0$, $0 \le F_x(\delta) \le 1 - \exp(-\delta\bar{f}_x(1,0,y)^{2/(p+q+1)}) < 1$, and

$$\begin{aligned} F_x(\delta) &= \lim_{n\to\infty} \text{Prob}\left[n^{2/(p+q+1)}\left(\frac{\hat{\theta}(x,y)}{\theta(x,y)} - 1\right) \le \delta\right] \\ &= \lim_{n\to\infty} \text{Prob}(A[\delta,n]) \\ &= \lim_{h\to\infty} \sum_{k=1}^{\infty} \text{Prob}\left(U\left[\frac{\delta}{h^2},k\right]\right) \frac{\tau(h)^k \bar{f}_x(1,0,y)^k}{k!} e^{-\tau(h)\bar{f}_x(1,0,y)}. \end{aligned} \tag{27}$$

The theorem shows that the asymptotic distribution $F_x(\delta)$ is stochastically dominated by the distribution function $1 - \exp(-\delta\bar{f}_x(1,0,y)^{2/(p+q+1)})$ of an exponential distribution with parameter $\bar{f}_x(1,0,y)^{2/(p+q+1)}$. If μ_x denotes the mean of F_x, we therefore obtain $\mu_x \le 1/\bar{f}_x(1,0,y)^{2/(p+q+1)}$.

Indeed, $F_x(\delta) = 1 - \exp(-\delta\bar{f}_x(1,0,y)^{2/(p+q+1)})$ in the special case $p = 1$ and $q = 0$. This is easily verified: obviously $p = 1$ and $q = 0$ corresponds to the (rather uninteresting) situation of firms that all produce a common, fixed output using some variable input $X_i \in \mathbb{R}_+$. The frontier is then given by the min-

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

imal possible input x_{min}. Assumption 6 is trivially satisfied for $\theta(x) \equiv \theta(x, y) = x_{min}/x$, and Assumption 5 is fulfilled if the density $f(x) \equiv f(x, y)$ of X_i is continuous for $x > x_{min}$ with $f(x_{min}) > 0$. For a sample $X_1, \ldots, X_n$ we obtain the trivial DEA estimators $\hat{x}_{min} = \min_i X_i$ and $\hat{\theta}(x) = \hat{x}_{min}/x$. Because there are no z or y variables, g_x is simply a constant equal to x_{min}, whereas $\bar{f}(\theta) \equiv \bar{f}_x(\theta, z, y)$ is the density of $\theta_i = x_{min}/X_i$. Here $f(x_{min}) > 0$ implies $\bar{f}(1) > 0$. With $\tilde{\vartheta}_1, \ldots, \tilde{\vartheta}_k$ denoting i.i.d. random variables uniformly distributed on $[0, 1]$, the event $U[\gamma, k]$ then corresponds to the event that there exists some $\alpha_j \geq 0$, $\sum_j \alpha_j = 1$, such that $\sum_{j=1}^{k} \alpha_j \tilde{\vartheta}_j \leq \gamma$. The latter is equivalent to requiring that $\tilde{\vartheta}_{min;k} = \min_j \tilde{\vartheta}_j \leq \gamma$. It is well known that for i.i.d. random variables $\tilde{\vartheta}_j \sim \text{Uniform}([0, 1])$ the statistics $k\tilde{\vartheta}_{min;k}$ asymptotically follow an exponential distribution with parameter $\lambda = 1$. Therefore, in this situation

$$\lim_{k\to\infty} \text{Prob}\left(U\left[\delta\, \frac{\bar{f}_x(1, 0, y)^{2/(p+q+1)}}{k^{2/(p+q+1)}}, k\right]\right) = \lim_{k\to\infty} \text{Prob}\left(\tilde{\vartheta}_{min;k} \leq \delta\, \frac{\bar{f}(1)}{k}\right)$$
$$= 1 - \exp(-\delta \bar{f}(1)). \quad \textbf{(28)}$$

In the general case $p > 1$, $q \geq 1$, it seems to be difficult to evaluate analytic expressions for (26). Nevertheless F_x 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 then determined by $(p + q)(p + q + 1)/2 + 2$ parameters that are (a) the value $\bar{f}_x(1, 0, y)$ of the density $\bar{f}_x$, (b) the value $g_x(0, y)$ of the function g_x at the corresponding frontier point, and (c) 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.

We finally note that the condition that (x, y) be in the interior of $\mathcal{D}$ does not impose a substantial restriction. The following corollary shows that Theorem 2 also characterizes the distribution of $\hat{\theta}(\lambda x, y)/\theta(\lambda x, y)$ for any $\lambda > 0$. The boundary point $(\theta(x, y)x, y)$ is a particular case with $\lambda = \theta(x, y)$.

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

$$F_x(\delta) = \lim_{n\to\infty} \text{Prob}\left[n^{2/(p+q+1)}\left(\frac{\hat{\theta}(\lambda x, y)}{\theta(\lambda x, y)} - 1\right) \leq \delta\right] \quad \textit{for all } \lambda > 0. \quad \textbf{(29)}$$

3. BOOTSTRAPPING DEA ESTIMATORS

Two bootstrap methods are presented in this section, and their consistency for inference-making purposes is established in Theorems 3 and 4 using the results

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

from Section 2. 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 2, we consider a fixed point $(x, y) \in \mathcal{D}$ 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 (7) to obtain a bootstrap estimate $\hat{\theta}^*(x, y)$ (note that $\hat{\theta}^*(x, y)$ 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 $n^{2/(p+q+1)}(\hat{\theta}^*/\hat{\theta} - 1)$ 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 $C(x, y; hn^{-1/(p+q+1)})$ which are proportional to $1/n$. 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 (see, e.g., Bickel and Freedman, 1981).

We consider two different bootstrap approaches; the first is based on subsampling, whereas the second is based on smoothing.

3.1. Bootstrap with Subsampling

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 (7) to construct bootstrap estimates $\hat{\theta}^*(x, y)$.
3. Repeat steps (1)–(2) B times; use the resulting bootstrap values to approximate the conditional distribution of $m^{2/(p+q+1)}(\hat{\theta}^*(x, y)/\hat{\theta}(x, y) - 1)$ given $\mathcal{S}_n$ and use this approximation to approximate the unknown distribution of $n^{2/(p+q+1)}(\hat{\theta}(x, y)/\theta(x, y) - 1)$. 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

$$\text{Prob}\left[m^{2/(p+q+1)} \left(\frac{\hat{\theta}^*(x, y)}{\hat{\theta}(x, y)} - 1 \right) \le \delta_{\alpha/2,m} | \mathcal{S}_n \right] = \frac{\alpha}{2}, \tag{30}$$

$$\text{Prob}\left[m^{2/(p+q+1)} \left(\frac{\hat{\theta}^*(x, y)}{\hat{\theta}(x, y)} - 1 \right) \le \delta_{1-\alpha/2,m} | \mathcal{S}_n \right] = 1 - \frac{\alpha}{2}. \tag{31}$$

4. Compute $[\hat{\theta}(x,y)/(1 + n^{-2/(p+q+1)}\delta_{1-\alpha/2,m}),\ \hat{\theta}(x,y)/(1 + n^{-2/(p+q+1)}\delta_{\alpha/2,m})]$, 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*

$$\sup_{\delta>0}\left| F_x(\delta) - \text{Prob}\left[m^{2/(p+q+1)}\left(\frac{\hat{\theta}^*(x,y)}{\hat{\theta}(x,y)} - 1\right) \le \delta | \mathcal{S}_n \right]\right| \xrightarrow{p} 0$$

as $n \to \infty$. **(32)**

3.2. 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 later in this section).

Algorithm 2.

1. Compute a *smooth* analogue $\hat{g}_x^*(z,\tilde{y})$ of the frontier function $\hat{g}_x(z,\tilde{y})$; details are given subsequently.
2. Draw a bootstrap sample $\bar{\mathcal{S}}_n^* = \{(\theta_i^*, \zeta_i^*, Y_i^*)\}_{i=1}^n$ by i.i.d. sampling from a smooth, nonparametric estimate $\hat{\bar{f}}_x$ of the density $\bar{f}_x$. Then determine $\tilde{\mathcal{S}}_n^* = \{(\theta_i^*, Z_i^*, Y_i^*)\}_{i=1}^n$ 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

$$X_i^* = \frac{\hat{g}_x^*(\theta_i^* Z_i^*, Y_i^*)}{\theta_i^*}\,\frac{x}{\|x\|} + Z_i^*. \qquad \textbf{(33)}$$

4. Apply the original DEA estimator in (7) to obtain a bootstrap estimate $\hat{\theta}^*(x,y)$.
5. Repeat steps (2)–(4) B times; use the resulting bootstrap values to approximate the conditional distribution of $(\hat{\theta}^*(x,y)/\hat{\theta}(x,y) - 1)$ given $\mathcal{S}_n$ and use this to approximate the unknown distribution of $(\hat{\theta}(x,y)/\theta(x,y) - 1)$. For a given $\alpha \in (0,1)$, use the bootstrap values to estimate the quantiles $\delta_{\alpha/2}$, $\delta_{1-\alpha/2}$ where

$$\text{Prob}\left[\left(\frac{\hat{\theta}^*(x,y)}{\hat{\theta}(x,y)} - 1\right) \le \delta_{\alpha/2} | \mathcal{S}_n\right] = \frac{\alpha}{2}, \qquad \textbf{(34)}$$

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

$$\text{Prob}\left[\left(\frac{\hat{\theta}^*(x,y)}{\hat{\theta}(x,y)}-1\right)\le\delta_{1-\alpha/2}|\mathcal{S}_n\right]=1-\frac{\alpha}{2}. \tag{35}$$

6. Compute $[\hat{\theta}(x,y)/(1+\delta_{1-\alpha/2}), \hat{\theta}(x,y)/(1+\delta_{\alpha/2})]$, 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 $\hat{f}_x$ and g_x only depend on y. However, for $p>1$ the preceding steps define g_x and $\hat{f}_x$ specifically for the point (x,y) that is of interest. Consequently, if confidence intervals are to be constructed for the efficiency measure defined in (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 $\bar{f}_x(\theta,\zeta,\tilde{y})$ at points $(\theta,\zeta,\tilde{y})$ with $\theta=1$. This problem is handled by reflecting observations $(\hat{\theta}_i,\zeta_i,Y_i)$ to obtain $(2-\hat{\theta}_i,\zeta_i,Y_i)$ (where $\hat{\theta}_i$ denotes the efficiency estimate computed from the smoothed frontier $\hat{g}_x^*$ 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 $\hat{g}_x^*$ 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 $\hat{g}_x$ and g_x is of order $n^{-2/(p+q+1)}$; even more importantly, $\hat{g}_x$ is not differentiable and hence does not possess the same degree of smoothness as g_x. Setting $\hat{g}_x^*=\hat{g}_x$ therefore does not seem to lead to a consistent bootstrap. Even if the distributions of (θ_i,Z_i,Y_i) and (θ_i^*,Z_i^*,Y_i^*) were identical, the asymptotic distributions of $\sum_{j=1}^k\alpha_j(g_x(\theta_jZ_j,Y_j)/\theta_jg_x(0,y))-1$ and $\sum_{j=1}^k\alpha_j(\hat{g}_x(\theta_j^*Z_j^*,Y_j^*)/\theta_j^*\hat{g}_x(0,y))-1$ would not in general coincide.

It is important to understand the purpose of smoothing the DEA frontier estimate. We do not require $\hat{g}_x^*$ to be closer to g_x than to $\hat{g}_x$. It suffices if the relative distances $\tilde{g}_x(z,\tilde{y})/g_x(z,\tilde{y})$ do not change very much with $(z,\tilde{y})$. If, for some $\beta>0$, we have $\beta g_x(z,\tilde{y})=\tilde{g}_x(z,\tilde{y})$ for all $(z,\tilde{y})$, then $g_x(\theta_iZ_i,Y_i)/g_x(0,y)=\tilde{g}_x(\theta_iZ_i,Y_i)/\tilde{g}_x(0,y)$, 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 $\tilde{g}_x^*$ 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 $\hat{g}_x$. This approach, however, presents a formidable integration problem in $(p+q-1)$-

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

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* $\hat{g}_x^*$ by

$$\hat{g}_x^*(z,\tilde{y}) = \hat{g}_x(0,y) + b^2\left[\hat{g}_x\left(\frac{z}{b}, y + \frac{\tilde{y}-y}{b}\right) - \hat{g}_x(0,y)\right]. \tag{36}$$

Note that setting $b = 1$ in (36) 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 (36), let $b < 1$ and define

$$g_x^*(z,\tilde{y}) = g_x(0,y) + b^2\left[g_x\left(\frac{z}{b}, y + \frac{\tilde{y}-y}{b}\right) - g_x(0,y)\right]. \tag{37}$$

The following properties are easily verified: (a) $\hat{g}_x^*$ and g_x^* are convex functions; (b) $\hat{g}_x^*(0,y) = \hat{g}_x(0,y) = \hat{\theta}(x,y)\|x\|$ and $g_x^*(0,y) = g_x(0,y) = \theta(x,y)\|x\|$; (c) 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 (d) by Theorem 1(ii), with probability tending to 1 as $n \to \infty$,

$$\left|\frac{\hat{g}_x^*(z,\tilde{y})}{\hat{g}_x^*(0,y)} - \frac{g_x^*(z,\tilde{y})}{g_x^*(0,y)}\right| = \left| b^2\,\frac{\hat{g}_x\left(\frac{z}{b}, y + \frac{\tilde{y}-y}{b}\right)}{\hat{g}_x(0,y)} - b^2\,\frac{g_x\left(\frac{z}{b}, y + \frac{\tilde{y}-y}{b}\right)}{g_x(0,y)}\right|$$
$$\leq b^2 n^{-2/(p+q+1)}\log n \tag{38}$$

holds for all $(1,z,\tilde{y}) \in C(x,y;h\cdot n^{-1/(p+q+1)})$, $h > 0$, if $n^{-1/(p+q+1)}/b \to 0$.

By Theorem 1(i), property (d) implies that if $b^2\log n \to 0$ as $n \to \infty$, the difference between $\hat{g}_x^*$ and g_x^* in the relevant neighborhoods $C(x,y;h\cdot n^{-1/(p+q+1)})$ of (x,y) is of *smaller* order than $n^{-2/(p+q+1)}$. Asymptotically, a bootstrap based on $\hat{g}_x^*$ will thus provide the same results as a bootstrap directly relying on g_x^*. On the other hand, it follows from properties (a)–(c) 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 later in this section). 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 that is the focus of this paper. If g_x is replaced by g_x^*, then (A.30), which appears in the Appendix in the proof of Proposition 1, becomes

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

$$\sum_{j=1}^{k} \alpha_j \frac{g_x^*(\theta_{i_j} Z_{i_j}, Y_{i_j})}{\theta_{i_j} g_x^*(0,y)} - 1 = \sum_{j=1}^{k} \alpha_j \frac{g_x^*(Z_{i_j}, Y_{i_j}) - g_x^*(0,y)}{g_x^*(0,y)} + \sum_{j=1}^{k} \alpha_j (1-\theta_{i_j})$$
$$+ O_p(n^{-3/(p+q+1)})$$
$$= \sum_{j=1}^{k} \alpha_j \frac{1}{2 g_x^*(0,y)} [Z_{i_j}^T g_{x;zz}^{*\prime\prime}(0,y) Z_{i_j}$$
$$+ 2 Z_{i_j}^T g_{x;zy}^{*\prime\prime}(0,y)(Y_{i_j} - y)$$
$$+ (Y_{i_j} - y)^T g_{x;yy}^{*\prime\prime}(0,y)(Y_{i_j} - y)]$$
$$+ \sum_{j=1}^{k} \alpha_j (1-\theta_{i_j}) + O_p(b^{-1} n^{-3/(p+q+1)}). \quad \textbf{(39)}$$

Thus, the bootstrap analogue of the assertion in Proposition 1 holds provided $n^{-1/(p+q+1)}/b \to 0$. The approximation error in (39) becomes smaller as b increases. On the other hand, decreasing b reduces the estimation error (38). The remainder terms in (38) and (39) 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 $n^{-1/(3(p+q+1))}$.

An obvious difficulty of the preceding bootstrap consists of the fact that in most bootstrap samples there will exist points (Z_i^*, Y_i^*) with $(Z_i^*/b,\ y + (Y_i^* - y)/b) \notin \hat{\Psi}^*$, where $\hat{\Psi}^*$ 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 because 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 $(Z_{i_j}^*/b, y + (Y_{i_j}^* - y)/b) \notin \hat{\Psi}^*$, $i_j \in \{1,\ldots,n\}$, $j = 1,\ldots,l$. Eliminate these points from the bootstrap samples and calculate $\hat{\theta}^*(x,y)$ from the remaining $(n-\ell)$ bootstrap observations.

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

In the simulations described in Section 4, we use the elimination option.

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

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

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

$$\sup_{(\theta,z,\tilde{y})\in C(x,y;h)} \left|\hat{f}_x(\theta,z,\tilde{y}) - \bar{f}_x(\theta,z,\tilde{y})\right| = o_p(1) \quad \text{as } n\to\infty \tag{40}$$

if h is sufficiently small. Furthermore, $b \to 0$ and $n^{-1/(p+q+1)}/b \to 0$ as $n \to \infty$.

It follows from well-known results in kernel density estimation that condition (40) is fulfilled for a kernel estimator $\hat{f}_x$ as described earlier, provided that the bandwidths τ_j used with respect to the different directions, $j = 1,\dots,p+q$, satisfy $\tau_j \to 0$ and $n \cdot \prod_{j=1}^{p+q} \tau_j / \log n \to \infty$ as $n \to \infty$.

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

THEOREM 4. *Given Assumptions 1–7,*

$$\sup_{\delta>0} \left| F_x(\delta) - \text{Prob}\left(n^{2/(p+q+1)} \left(\frac{\hat{\theta}^*(x,y)}{\hat{\theta}(x,y)} - 1 \right) \le \delta | \mathcal{S}_n \right) \right| \xrightarrow{p} 0 \quad \text{as } n \to \infty. \tag{41}$$

4. 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 each set of experiments, we examined seven 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 $x = x_e e^{0.2|\epsilon|}$, where $\epsilon \sim N(0,1)$ and is independent. The DGP for this case can therefore be written as

$$y = x^{0.8} e^{-0.2|\epsilon|}. \tag{42}$$

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 $[\frac{1}{9}\frac{\pi}{2}, \frac{8}{9}\frac{\pi}{2}]$ 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 $x_1 = x_{1e} e^{0.2|\epsilon|}$ and $x_2 = x_{2e} e^{0.2|\epsilon|}$ where $\epsilon \sim N(0,1)$ 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-subsampling using Algorithm 1 as described in Section 3.1. 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, because of 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-smooth 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 $\hat{g}_x$ in step 1 of the algorithm, using (36). 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 because numerical problems when $n = 25$ or $n = 50$ (see the discussion preceding Assumption 7).

Recall from the discussion surrounding (39) that our theoretical results imply that the optimal value of b should be proportional to $n^{-1/(3(p+q+1))}$. Because 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 $b = n^{-1/(3(p+q+1))}$ as a rule of thumb implies that $b = n^{-1/9}$ 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

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

TABLE 1. Coverage of confidence intervals estimated by subsampling

		$p = q = 1$ $(1-\alpha)$			$p = q = 2$ $(1-\alpha)$		
n	κ	0.90	0.95	0.99	0.90	0.95	0.99
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
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

(*continued*)

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

TABLE 1. Continued

		$p = q = 1$ $(1-\alpha)$			$p = q = 2$ $(1-\alpha)$		
n	κ	0.90	0.95	0.99	0.90	0.95	0.99
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

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 subsampling 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 coverages for $p = q = 2$ shown in Table 1 for the subsampling bootstrap. However, Table 1 also reveals that suboptimal choices of the tuning parameter κ required for the subsampling method can easily result in coverages worse than

Table 2. Coverage of confidence intervals estimated by double-smooth bootstrap

		$p = q = 1$ $(1 - \alpha)$			$p = q = 2$ $(1 - \alpha)$		
n	b	0.90	0.95	0.99	0.90	0.95	0.99
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

those shown in Table 2 when b is chosen according to the rule of thumb discussed earlier. 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.

5. CONCLUSIONS

The analysis in Section 2 establishes the asymptotic distribution of the DEA efficiency estimator for the variable returns to scale case under rather weak assumptions on the DGP, whereas the analysis in Section 3 establishes consistency of two bootstrap procedures. The bootstrap procedures are necessary for any practical application because the asymptotic distribution in Theorem 2 contains unknown terms and would be difficult either to estimate or to simulate. As noted in Sections 1 and 4, there is at present no reliable way to choose the size of subsamples in Algorithm 1, and hence we do not recommend the subsampling bootstrap. Although 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 or she 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).

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

Korostelev, A., L. Simar, & A.B. Tsybakov (1995) On estimation of monotone and convex boundaries. *Publications de l'Institut de Statistique de l'Université de Paris XXXIX* 1, 3–18.

Sheather, S.J. & M.C. Jones (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 University Press.

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

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

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

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

Simar, L. & P.W. Wilson (2000a) A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics* 27, 779–802.
Simar, L. & P.W. Wilson (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.

APPENDIX

Proof of Lemma 1. 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 $[\alpha g_x(z_1, y_1) + (1-\alpha) g_x(z_2, y_2)](x/\|x\|) + \tilde{z}_\alpha \geq g_x(\tilde{z}_\alpha, \tilde{y}_\alpha)(x/\|x\|) + \tilde{z}_\alpha$ with $(\tilde{z}_\alpha, \tilde{y}_\alpha) = (\alpha z_1 + (1-\alpha) z_2, \alpha y_1 + (1-\alpha) y_2) \in \Psi^*(x)$. Consequently, g_x is a convex function. The same arguments apply to $\hat{g}_x$. In view of Assumption 6 twice-differentiability of g_x and continuity of $g_x''(0, y)$ in x are immediate consequences of (12). Assumption 1 implies that

$$1 = \alpha\theta\left(g_x(z_1, y_1)\frac{x}{\|x\|} + z_1, y_1\right) + (1-\alpha)\theta\left(g_x(z_2, y_2)\frac{x}{\|x\|} + z_2, y_2\right)$$
$$> \theta\left((\alpha g_x(z_1, y_1) + (1-\alpha) g_x(z_2, y_2))\frac{x}{\|x\|}, y\right) \quad \textbf{(A.1)}$$

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 $\alpha y_1 + (1-\alpha) y_2 = y$. Because $\theta(g_x(0, y)(x/\|x\|), y) = 1$, 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''.

By our assumptions and assertion (ii) $T(\theta, \zeta, y) = ((g_x(\theta \sum_{j=1}^{p-1} \zeta_j z^{(j)}, y)/\theta)(x/\|x\|) + \sum_{j=1}^{p-1} \zeta_j z^{(j)}, y)$ is a twice continuously differentiable, bijective mapping from $(0,1] \times \mathbb{R}^{p-1} \times \mathbb{R}_+^q$, into Ψ. Note that $T(\theta, \zeta, y) = T_1(T_2(\theta, \zeta, y))$, where $T_1(\gamma, \zeta, y) = (\gamma(x/\|x\|) + \sum_{j=1}^{p-1} \zeta_j z^{(j)}, y)$ and $T_2(\theta, \zeta, y) = ((g_x(\theta \sum_{j=1}^{p-1} \zeta_j z^{(j)}, y)/\theta_y), \zeta, y)$. Here T_1 corresponds to a simple change of the coordinate system, and $|\det(J_1(\gamma, \zeta, y))| > 0$ for all (γ, ζ, y), where J_1 denotes the Jacobian matrix of partial derivatives of T_1; T_2 is a twice continuously differentiable, injective mapping. All rows of the Jacobian matrix J_2 of partial derivatives of T_2 are linearly independent, and therefore $|\det(J_2(\theta, \zeta, y))| > 0$ for all (θ, ζ, y). Therefore, by well-known results on variable transformation we have

$$\bar{f}_x(\theta, \zeta, y) = \left|\det\left(J_1\left(\frac{g_x\left(\theta \sum_{j=1}^{p-1} \zeta_j z^{(j)}, y\right)}{\theta}, \zeta, y\right)\right)\right|$$
$$\times |\det(J_2(\theta, \zeta, y))| \times f(T_1(T_2(\theta, \zeta, y))), \quad \textbf{(A.2)}$$

and continuity of $\bar{f}_x$ on $\bar{\mathcal{D}} = T^{-1}(\mathcal{D})$ follows from the continuity of f on $\mathcal{D}$. Furthermore, $f(\theta(x, y)x, y) > 0$ implies $f_x(1, 0, y) = |\det(J_1(g_x(0, y), \zeta, y))| \times |\det(J_2(1, 0, y))| \times f(\theta(x, y)x, y) > 0$. The continuity of $\bar{f}_x(1, 0, y)$ in x follows from the preceding structure of $\bar{f}_x$ together with $\theta(x, y) = g_x(0, y)/\|x\|$ and our assumption on $\theta(x, y)$. ■

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

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

$$\sum_{i=1}^{k} \alpha_i Y_i = y \quad \text{and} \quad \sum_{i=1}^{k} \alpha_i X_i = \beta x \tag{A.3}$$

hold for some $\alpha_1 \geq 0, \ldots, \alpha_n \geq 0$ with $\sum_{j=1}^{n} \alpha_j = 1$. The relations in (8) and (12) imply $X_i = (g_x(\theta_i Z_i, Y_i)/\theta_i \|x\|)x + Z_i$. Because all Z_i are orthogonal to x, (A.3) holds if and only if (16) is satisfied and $\sum_{i=1}^{n} \alpha_i (g_x(\theta_i Z_i, Y_i)/\theta_i \|x\|) = \beta$. The lemma now follows from $g_x(0,y) = \|x\|\theta(x,y)$. ■

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*

$$\sum_{r=1}^{k} \alpha_r z_r = 0, \qquad \sum_{r=1}^{k} \alpha_r y_r = y \tag{A.4}$$

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

$$\sum_{r=1}^{k-1} \tilde{\alpha}_r z_r + \tilde{\alpha}_k \tilde{z} = 0, \qquad \sum_{r=1}^{k-1} \tilde{\alpha}_r y_r + \tilde{\alpha}_k \tilde{y} = y \tag{A.5}$$

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

$$\sum_{r=1}^{k} \alpha_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0,y)} \geq \sum_{r=1}^{k-1} \tilde{\alpha}_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0,y)} + \tilde{\alpha}_k \frac{g_x(\tilde{z}, \tilde{y})}{g_x(0,y)} + c_1 \cdot \tilde{\alpha}_k h b n^{-2/(p+q+1)}, \tag{A.6}$$

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

Proof. Assume that (A.4) holds with $(\theta_k, z_k, y_k) \notin C(x,y; hn^{-1/(p+q+1)})$. Then either $\theta_k \leq 1 - h^2 n^{-2/(p+q+1)}$ and $(1, z_k, y_k) \in C(x,y; hn^{-1/(p+q+1)})$ or $(1, z_k, y_k) \notin C(x,y; hn^{-1/(p+q+1)})$.

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

$$\frac{g_x(\theta_k z_k, y_k)}{\theta_k g_x(0,y)} \geq \frac{g_x(z_k, y_k)}{g_x(0,y)} + \frac{1}{2}(1 - \theta_k) \geq \frac{g_x(z_k, y_k)}{g_x(0,y)} + \frac{1}{2} h^2 n^{-2/(p+q+1)}. \tag{A.7}$$

Note that $(1, z_k, y_k) \in C(x,y; hn^{-1/(p+q+1)})$ implies that $(1, (b/h)z_k, y + (b/h)(y_k - y)) \in C(x,y; bn^{-1/(p+q+1)})$. Relation (A.5) thus holds for $(\tilde{z}, \tilde{y}) := ((b/h)z_k, y + (b/h)(y_k - y))$ and $\tilde{\alpha}_r = \alpha_r\big((b/h)/((b/h) + \alpha_k(1 - (b/h)))\big)$ and $\tilde{\alpha}_k = \alpha_k\big(1/((b/h) + \alpha_k(1 - (b/h)))\big)$. Then (A.7) and convexity of g_x lead to

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

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

$$+ \frac{\alpha_k\left(1-\frac{b}{h}\right)}{\frac{b}{h} + \alpha_k\left(1-\frac{b}{h}\right)} \tag{A.8}$$

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

$$+ \tilde{\alpha}_k \frac{b}{h} \frac{1}{2} h^2 n^{-2/(p+q+1)} \tag{A.9}$$

$$\geq \sum_{r=1}^{k-1} \tilde{\alpha}_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0,y)} + \tilde{\alpha}_k \frac{g_x(\tilde{z}, \tilde{y})}{g_x(0,y)} + \tilde{\alpha}_k \frac{1}{2} bhn^{-2/(p+q+1)}. \tag{A.10}$$

It now only remains to prove (A.6) for the case where $(1, z_k, y_k) \notin C(x, y; hn^{-1/(p+q+1)})$. Let $\gamma = \max\{\delta | (1, \delta z_k, y + \delta(y_k - y)) \in C(x, y; hn^{-1/(p+q+1)})\}$ and $\alpha_r^* = \alpha_r\big(\gamma/(\gamma + \alpha_k(1-\gamma))\big)$ and $\alpha_k^* = \alpha_k\big(1/(\gamma + \alpha_k(1-\gamma))\big)$. This yields

$$\sum_{r=1}^{k-1} \alpha_r^* z_r + \alpha_k^* \gamma z_k = 0, \qquad \sum_{r=1}^{k-1} \alpha_r^* y_r + \alpha_k^* (y + \gamma(y_k - y)) = y. \tag{A.11}$$

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.10) then imply

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

$$\geq \sum_{r=1}^{k-1} \alpha_r^* \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0,y)} + \alpha_k^* \frac{g_x(\gamma z_k, y + \gamma(y_k - y))}{g_x(0,y)}. \tag{A.12}$$

Finally, define $(\tilde{z}, \tilde{y}) := ((b/h)\gamma z_k, y + (b/h)\gamma(y_k - y))$, $\tilde{\alpha}_r = \alpha_r^*\big((b/h)/((b/h) + \alpha_k^*(1-(b/h)))\big)$, and $\tilde{\alpha}_k = \alpha_k^*\big(1/((b/h) + \alpha_k^*(1-(b/h)))\big)$. Clearly, then, $(1, \tilde{z}, \tilde{y}) \in C(x, y; bn^{-1/(p+q+1)})$, and relation (A.5) is a direct consequence of (A.11). Moreover, for sufficiently large n,

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

$$
\begin{aligned}
&\sum_{r=1}^{k-1} \alpha_r^* \frac{g_x(\theta_r z_r, Y_r)}{\theta_r g_x(0,y)} + \alpha_k^* \frac{g_x(\gamma z_k, y+\gamma(y_k-y))}{g_x(0,y)} \\
&\quad \geq \sum_{r=1}^{k-1} \tilde{\alpha}_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0,y)} + \tilde{\alpha}_k \left[\frac{\frac{b}{h} g_x(\gamma z_k, y+\gamma(y_k-y))}{g_x(0,y)} + \left(1-\frac{b}{h}\right)\frac{g_x(0,y)}{g_x(0,y)} \right] \\
&\quad \geq \sum_{r=1}^{k-1} \tilde{\alpha}_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0,y)} + \tilde{\alpha}_k \frac{g_x(\tilde{z}_k, \tilde{y}_k)}{g_x(0,y)} + \tilde{\alpha}_k \frac{b}{h} \frac{c_0 h^2 n^{-2/(p+q+1)}}{8 g_x(0,y)}.
\end{aligned}
\tag{A.13}
$$

By using Lemma 1(ii) the second inequality follows from Taylor expansions of $g_x(\gamma z_k, y + \gamma(y_k - y))$ and $g_x(0, y)$ at the point $(\tilde{z}, \tilde{y}) := ((b/h)\gamma z_k, y + (b/h)\gamma(y_k - y))$. Note that the first derivatives cancel out because of $(b/h)/(\gamma z_k - (b/h)\gamma z_k) + (1 - (b/h))\cdot(-(b/h)\gamma z_k) = 0$ and $(b/h)(\gamma(y_k - y) - (b/h)\gamma(y_k - y)) + (1 - (b/h))\cdot(-(b/h)\gamma(y_k - y)) = 0$. The bound given in (A.13) is then obtained by an analysis of the second derivatives while taking into account that $1 - (b/h) \geq \frac{1}{2}$, $\left\| \begin{pmatrix} \gamma z_k \\ \gamma(y_k - y) \end{pmatrix} \right\|^2 \geq h^2$, and that $\inf_{(1,z,w)\in C(x,y;bn^{-1/(p+q+1)})} \inf_{\|v\|=1} v^T g_x''(z,w)v \geq (c_0/2)$ for all sufficiently large n, where c_0 is defined in Lemma 1(ii). Combining (A.12) and (A.13) yields (A.6). ■

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 by the corresponding samples $\tilde{\mathcal{S}}_n = \{(\theta_i, Z_i, Y_i)\}_{i=1}^n$ 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 $b_n = b\cdot n^{-1/(p+q+1)}$, $b_n^* = b_n/(2(p-1) + 2q)$. For $i = 1, \ldots, p-1$ and $j = 1, \ldots, q$, define

$$
\bar{B}_{2i-1} = \Big\{ (v,w) \in \mathbb{R}^{p-1} \times \mathbb{R}^q \,|\, \max_{r \neq i} |v_r| \leq b_n^*, |v_i - b_n| \leq b_n^*, \max_{s=1,\ldots,q} |y_s - w_s| \leq b_n^* \Big\}, \tag{A.14}
$$

$$
\bar{B}_{2i} = \Big\{ (v,w) \in \mathbb{R}^{p-1} \times \mathbb{R}^q \,|\, \max_{r \neq i} |v_r| \leq b_n^*, |v_i + b_n| \leq b_n^*, \max_{s=1,\ldots,q} |y_s - w_s| \leq b_n^* \Big\}, \tag{A.15}
$$

$$
\bar{B}_{2j-1+2(p-1)} = \Big\{ (v,w) \in \mathbb{R}^{p-1} \times \mathbb{R}^q \,|\, \max_{r=1,\ldots,p-1} |v_r| \leq b_n^*, \max_{s \neq j} |y_s - w_s| \leq b_n^*, |y_j + b_n - w_j| \leq b_n^* \Big\}, \tag{A.16}
$$

$$
\bar{B}_{2j+2(p-1)} = \Big\{ (v,w) \in \mathbb{R}^{p-1} \times \mathbb{R}^q \,|\, \max_{r=1,\ldots,p-1} |v_r| \leq b_n^*, \max_{s \neq j} |y_s - w_s| \leq b_n^*, |y_j - b_n - w_j| \leq b_n^* \Big\}. \tag{A.17}
$$

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

Finally, for $j = 1,\dots,2(p-1)+2q$ 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 and 5 that if n is sufficiently large,

$$\bar{D}_{j,n} := [1 - b_n^2, 1] \times \bar{B}_j \subset \bar{\mathcal{D}} \tag{A.18}$$

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

For each $j = 1,\dots,2(p-1)+2q$ the set $\bar{D}_{j,n}$ has Lebesgue measure equal to $(2/(2(p-1)+2q))^{p+q-1} b^{p+q+1} \cdot (1/n)$, and our assumptions on the distribution of the random variables (θ_i, ζ_i, Y_i) thus imply $\text{Prob}[(\theta_i, \zeta_i, y_i) \in \bar{D}_{j,n}] = a_0 b^{p+q+1} \cdot (1/n) + o(1/n)$ for $a_0 = \bar{f}_x(1,0,y)(2/(2(p-1)+2q))^{p+q-1}$. Because (θ_i, ζ_i, Y_i), $i = 1,\dots,n$, are independent we have $\lim_{n\to\infty} \text{Prob}(\bar{\mathcal{S}}_n \cap \bar{D}_{j,n} = \emptyset) = \lim_{n\to\infty}(1 - (a_0 b^{p+q+1})/n)^n = \exp(-a_0 b^{p+q+1})$. When additionally using the Bonferroni inequality, we can thus infer that there exist some $0 < d_0$, $d_1 < \infty$ such that for all n sufficiently large,

$$\begin{aligned} 1 - (2(p-1)+2q) \cdot \exp(-d_0 b^{p+q+1}) \\ \leq \text{Prob}(\bar{\mathcal{S}}_n \cap \bar{D}_{j,n} \neq \emptyset \quad \forall j = 1,\dots,2(p-1)+2q) \\ \leq 1 - \exp(-d_1 b^{p+q+1}). \end{aligned} \tag{A.19}$$

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

$$\text{Prob}(\bar{\mathcal{S}}_n \cap \bar{D}_{j,n} \neq \emptyset \quad \forall j = 1,\dots,2(p-1)+2q) \geq 1 - \epsilon. \tag{A.20}$$

By (A.20), assertion (i) of the theorem holds if there is an $h_\epsilon > 0$ such that for all $h > h_\epsilon$ the following conditional probabilities are equivalent for sufficiently large n:

$$\begin{aligned} &\text{Prob}(A[\delta, n] | \bar{\mathcal{S}}_n \cap \bar{D}_{j,n} \neq \emptyset \quad \forall j) \\ &= \text{Prob}(A[\delta, n; h \cdot n^{-1/(p+q+1)}] | \bar{\mathcal{S}}_n \cap \bar{D}_{j,n} \neq \emptyset \quad \forall j). \end{aligned} \tag{A.21}$$

Now we will demonstrate that (A.21) is satisfied for all $h \geq c_3 \cdot b$, where $c_3 < \infty$ denotes a suitable constant that will be specified in what follows.

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

$$\tilde{z} = \sum_{j=1}^{2(p-1)+2q} \gamma_j \tilde{z}_j, \qquad \tilde{y} = \sum_{j=1}^{2(p-1)+2q} \gamma_j \widetilde{w}_j. \tag{A.22}$$

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

$$\sup_{(1,z,w) \in C(x,y;b_n^*)} \left[\sup_{\|v\|=1} v^T g_x''(z,w) v \right] \leq c_0^* \tag{A.23}$$

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

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

$$\frac{g_x(\tilde{z},\tilde{y})}{g_x(0,y)} \le \sum_{j=1}^{2(p-1)+2q} \gamma_j \frac{g_x(\tilde{\theta}_j \tilde{z}_j, \widetilde{w}_j)}{\tilde{\theta}_j g_x(0,y)}$$

$$\le \sum_{j=1}^{2(p-1)+2q} \gamma_j \left(\frac{g_x(\tilde{\theta}_j \tilde{z}_j, \widetilde{w}_j)}{g_x(0,y)} + 1.5\left(\frac{1}{\tilde{\theta}_j} - 1\right)\right) \le \frac{g_x(\tilde{z},\tilde{y})}{g_x(0,y)} + c_2 b^2 n^{-2/(p+q+1)}, \tag{A.24}$$

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

Using the continuity of g_x'', the second inequality can be derived from second-order Taylor expansions of $g_x(\tilde{\theta}_j \tilde{z}_j, \widetilde{w}_j)$ at $(\tilde{z},\tilde{y})$. Note that because of (A.22) 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 $b \ge b_\epsilon$ and $h \ge c_3 b$. Consider an arbitrary $(\theta, z, w) \in \bar{\mathcal{S}}_n$ with $(\theta, z, w) \notin C(x, y; hn^{-1/(p+q+1)})$, and assume that for $k \le n$ there exist some $(\theta_1, z_1, y_1), \ldots, (\theta_{k-1}, z_{k-1}, y_{k-1}) \in \bar{\mathcal{S}}_n$ such that (A.4) holds with $(\theta_k, z_k, y_k) = (\theta, z, w)$. Lemma A1 then implies that there is a $(\tilde{z},\tilde{y})$ with $(1,\tilde{z},\tilde{y}) \in C(x, y; (b/(2(p-1)+2q))n^{-1/(p+q+1)})$ such that relations (A.5) and (A.6) are satisfied when b is replaced by $(b/(2(p-1)+2q))$.

On the other hand, $\bar{\mathcal{S}}_n \cap D_{j,n} \ne \emptyset \ \forall\ j = 1, \ldots, 2(p-1)+2q$ imposes the existence of $2(p-1)+2q$ points $(\tilde{\theta}_1, \tilde{z}_1, \widetilde{w}_1) \in \bar{\mathcal{S}}_n \cap [1-b_n^2, 1] \times B_1, \ldots, (\tilde{\theta}_{2(p-1)+q}, \tilde{z}_{2(p-1)+q}, \widetilde{w}_{2(p-1)+q}) \in \bar{\mathcal{S}}_n \cap [1-b_n^2, 1] \times B_{2(p-1)+q}$. For some suitable $\gamma_1, \ldots, \gamma_{2(p-1)+q} \ge 0$ with $\sum_{j=1}^{2(p-1)+q} \gamma_j = 1$, we then obtain (A.22)–(A.24), and one can conclude from (A.6) that

$$\sum_{r=1}^{k-1} \alpha_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0,y)} + \alpha_k \frac{g_x(\theta z, w)}{\theta g_x(0,y)}$$

$$\ge \sum_{r=1}^{k-1} \tilde{\alpha}_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0,y)} + \tilde{\alpha}_k \frac{g_x(\tilde{z},\tilde{y})}{g_x(0,y)} + \alpha_k \frac{c_1 c_3}{2(p-1)+2q} b^2 n^{-2/(p+q+1)}$$

$$\ge \sum_{r=1}^{k-1} \tilde{\alpha}_r \frac{g_x(\theta_r z_r, y_r)}{\theta_r g_x(0,y)} + \sum_{j=1}^{2(p-1)+2q} \tilde{\alpha}_k \gamma_j \frac{g_x(\tilde{\theta}_j \tilde{z}_j, \widetilde{w}_j)}{\tilde{\theta}_j g_x(0,y)}, \tag{A.25}$$

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

Note that $(\tilde{\theta}_j, \tilde{z}_j, \widetilde{w}_j) \in \bar{\mathcal{S}}_n \cap C(x, y; hn^{-1/(p+q+1)})$ for all j. From (A.25), if $\bar{\mathcal{S}}_n \cap D_{j,n} \ne \emptyset \ \forall\ j$, then the minimal value of $\sum_i \alpha_i (g_x(\theta_i Z_i, Y_i)/\theta_i g_x(0,y))$ over all $\alpha_1, \ldots, \alpha_n \ge 0$ with $\sum \alpha_i = 1$ is achieved by those linear combinations that assign zero weight $\alpha_i = 0$ to all observations with $(\theta, z, w) := (\theta_i, Z_i, Y_i) \notin C(x, y; hn^{-1/(p+q+1)})$. This leads to (A.21) and thus completes the proof of part (i).

Consider part (ii). First note that our assumption on $\mathcal{N}$ implies that $\sup_{(v,w),(\tilde{v},\widetilde{w})\in\mathcal{N}} (\|v-\tilde{v}\| + \|w-\widetilde{w}\|) < \infty$. Therefore, $c_f := \inf_{(\tilde{x},\tilde{y})\in\mathcal{N}} f_{x_r}(1, 0, \tilde{y}) > 0$, and $g_{\tilde{x}}(0,\tilde{y})$ and $\sup_{\|v\|=1} v^T g_{\tilde{x}}''(0,\tilde{y}) v$ are uniformly bounded for $(\tilde{x},\tilde{y}) \in \mathcal{N}$. Let $b = b_0((\log n)/(\log\log n))^{1/(p+q+1)}$, $b_n = b \cdot n^{-1/(p+q+1)}$ and $b_n^* = (b_n/(2(p-1)+2q))$ for some $b_0 > 0$. For sufficiently large n it is then possible to construct a grid $(x_r, y_r) \in \mathcal{N}$, $r = 1, \ldots, n^{(p+q-1)/(p+q)}$, of $n^{(p+q-1)/(p+q)}$ points with the property that $\mathcal{N} \subseteq \{(\gamma(x_r/\|x_r\|) +$

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

$z, \tilde{y}) \in \Psi | (1, z, \tilde{y}) \in C(x_r, y_r, b_n^*), r \in \{1, \ldots, n^{(p+q-1)/(p+q)}\}, \gamma > 0\}$. Replace (x, y) by (x_r, y_r) to define sets $\bar{B}_j^{(r)}$ and $\bar{D}_{j,n}^{(r)}$ analogous to those used in the proof of assertion (i). Because the values $\inf_{x_r, y_r} f_{x_r}(1,0, y_r) \geq c_f > 0$ are uniformly bounded for all r, probability bounds analogous to (A.19) can be established for constants $0 < d_0, d_1 < \infty$ that can be chosen independent of r. The Bonferroni inequality then yields

$$\text{Prob}(\bar{\mathcal{S}}_n \cap \bar{D}_{j,n}^{(r)} \neq \emptyset \quad \text{for all } j = 1, \ldots, 2(p-1) + 2q$$

$$\text{and all } r = 1, \ldots, n^{(2p+2q-1)/(2p+2q+1)})$$

$$\geq 1 - n^{(p+q-1)/(p+q)}(2(p-1) + 2q) \cdot \exp\left(-d_0 b_0^{p+q+1} \frac{\log n}{\log\log n}\right) \to 1 \qquad \textbf{(A.26)}$$

as $n \to \infty$. Now assume that $\bar{\mathcal{S}}_n \cap \bar{D}_{j,n}^{(r)} \neq \emptyset$ for all j, r. There then exist some $(\tilde{\theta}_j^{(r)}, \tilde{z}_j^{(r)}, \tilde{w}_j^{(r)}) \in \bar{\mathcal{S}}_n \cap D_{j,n}^{(r)}$ such that for *all* r and *all* $(1, \tilde{z}, \tilde{y}) \in C(x_r, y_r; b_n^*)$ we obtain $\tilde{z} = \sum_{j=1}^{2(p-1)+2q} \gamma_j \tilde{z}_j^{(r)}$ and $\tilde{y} = \sum_{j=1}^{2(p-1)+2q} \gamma_j \tilde{w}_j^{(r)}$ for some suitable $\gamma_j \geq 0$ with $\sum_j \gamma_j = 1$. Because $((g_{x_r}(\tilde{\theta}_j^{(r)} \tilde{z}_j^{(r)}, \tilde{w}_j^{(r)})/\tilde{\theta}_j^{(r)})(x_r/\|x_r\|) + \tilde{z}_j, \tilde{w}_j) \in \mathcal{S}_n$, definition of $\hat{g}_{x_r}$ and g_{x_r} and arguments similar to (A.24) lead to

$$g_{x_r}(\tilde{z}, \tilde{y}) \leq \hat{g}_{x_r}(\tilde{z}, \tilde{y}) \leq \sum_{j=1}^{2(p-1)+2q} \gamma_j \frac{g_{x_r}(\tilde{\theta}_j^{(r)} \tilde{z}_j^{(r)}, \tilde{w}_j^{(r)})}{\tilde{\theta}_j}$$

$$\leq g_{x_r}(\tilde{z}, \tilde{y}) + \tilde{c}_2 b_0^2 n^{-2/(p+q+1)} \left(\frac{\log n}{\log\log n}\right)^{2/(p+q+1)}, \qquad \textbf{(A.27)}$$

where $\tilde{c}_2$ may be chosen independent of r because $g_{x_r}(0, y_r)$ and $g_{x_r}''(0, y_r)$ are uniformly bounded. Note that also the partial derivatives of $\theta(\tilde{x}, \tilde{y})$ are uniformly bounded for $(\tilde{x}, \tilde{y}) \in \mathcal{N}$. By suitable Taylor expansions of θ it therefore follows from (A.27) that there exists a constant $\tilde{c}_3 < \infty$, independent of r, $(\tilde{z}, \tilde{y})$, so that

$$1 - \theta\left(\hat{g}_{x_r}(\tilde{z}, \tilde{y}) \frac{x_r}{\|x_r\|} + \tilde{z}, \tilde{y}\right)$$

$$= \theta\left(g_{x_r}(\tilde{z}, \tilde{y}) \frac{x_r}{\|x_r\|} + \tilde{z}, \tilde{y}\right) - \theta\left(\hat{g}_{x_r}(\tilde{z}, \tilde{y}) \frac{x_r}{\|x_r\|} + \tilde{z}, \tilde{y}\right)$$

$$\leq \tilde{c}_3 b_0^2 n^{-2/(p+q+1)} \left(\frac{\log n}{\log\log n}\right)^{2/(p+q+1)}. \qquad \textbf{(A.28)}$$

We can then conclude that for some $\tilde{c}_3 < \tilde{c}_4 < \infty$, *all* r and *all* $(1, \tilde{z}, \tilde{y}) \in C(x_r, y_r; b_n^*)$

$$\frac{\hat{\theta}\left(\hat{g}_{x_r}(\tilde{z}, \tilde{y}) \frac{x_r}{\|x_r\|} + \tilde{z}, \tilde{y}\right)}{\theta\left(\hat{g}_{x_r}(\tilde{z}, \tilde{y}) \frac{x_r}{\|x_r\|} + \tilde{z}, \tilde{y}\right)} - 1 \leq \tilde{c}_4 b_0^2 n^{-2/(p+q+1)} \left(\frac{\log n}{\log\log n}\right)^{2/(p+q+1)}. \qquad \textbf{(A.29)}$$

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

Consider an arbitrary $(\tilde{x},\tilde{y}) \in \mathcal{N}$. By construction, $(\tilde{x},\tilde{y}) = ((\hat{g}_{x_r}(\lambda z,\tilde{y})/\lambda)(x_r/\|x_r\|) + z,\tilde{y})$ for some r, $(1,z,\tilde{y}) \in C(x_r,y_r,b_n^*)$, and some $\lambda \geq 1$, which implies $(\hat{\theta}(\tilde{x},\tilde{y})/\theta(\tilde{x},\tilde{y})) = (\hat{\theta}(\hat{g}_{x_r}(\tilde{z},\tilde{y})(x_r/\|x_r\|) + \tilde{z},\tilde{y}))/(\theta(\hat{g}_{x_r}(\tilde{z},\tilde{y})(x_r/\|x_r\|) + \tilde{z},\tilde{y}))$ for $\tilde{z} = \lambda z$. Assertion (20) then is a consequence of (A.26) and (A.29). Furthermore, note that convexity of $\hat{g}_x$ implies that $\hat{g}_x(\lambda_1 z,\tilde{y}) \geq \hat{g}_x(\lambda_2 z,\tilde{y})$ for $\lambda_1 > \lambda_2 > 0$. Therefore, with $z = \tilde{x} - (x^T\tilde{x}/\|x\|^2)x$ (21) follows from (20) and $(\hat{\theta}(\tilde{x},\tilde{y})/\theta(\tilde{x},\tilde{y})) = (\hat{g}_x(\hat{\theta}(\tilde{x},\tilde{y})z,\tilde{y}))/(g_x(\theta(\tilde{x},\tilde{y})z,\tilde{y})) \geq (\hat{g}_x(\theta(\tilde{x},\tilde{y})z,\tilde{y}))/(g_x(\theta(\tilde{x},\tilde{y})z,\tilde{y})) \geq 1$. ■

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

$$\begin{aligned}\sum_{j=1}^{k}\alpha_j\frac{g_x(\theta_{i_j}Z_{i_j},Y_{i_j})}{\theta_{i_j}g_x(0,y)} - 1 &= \sum_{j=1}^{k}\alpha_j\frac{g_x(\theta_{i_j}Z_{i_j},Y_{i_j}) - g_x(0,y)}{g_x(0,y)}\\ &\quad + \sum_{j=1}^{k}\alpha_j(1-\theta_{i_j}) + o_p(n^{-2/(p+q+1)})\\ &= \sum_{j=1}^{k}\alpha_j\frac{1}{2g_x(0,y)}\,[Z_{i_j}^T g''_{x;zz}(0,y)Z_{i_j} + 2Z_{i_j}^T g''_{x;zy}(0,y)(Y_{i_j}-y)\\ &\qquad + (Y_{i_j}-y)^T g''_{x;yy}(0,y)(Y_{i_j}-y)]\\ &\quad + \sum_{j=1}^{k}\alpha_j(1-\theta_{i_j}) + o_p(n^{-2/(p+q+1)}),\end{aligned} \tag{A.30}$$

where the convergence is uniform for all possible $(X_{i_j},Y_{i_j}) \in C(x,y;hn^{-1/(p+q+1)})$. Note that necessarily $\sum_{j=1}^{k}\alpha_j[g_{x;z}(0,y)'\cdot Z_{i_j} + g'_{x;y}(0,y)\cdot(Y_{i_j}-y)] = 0$, 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 $C(x,y;h\cdot n^{-1/(p+q+1)})$ 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 $C(x,y;h\cdot n^{-1/(p+q+1)})$ 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 $(\theta_i,Z_i,Y_i) \in C(x,y;h\cdot n^{-1/(p+q+1)})$, is uniform on $\bar{C}(h\cdot n^{-1/(p+q+1)}) := [1-h^2n^{-2/(p+q+1)},1] \times [-hn^{-1/(p+q+1)},hn^{-1/(p+q+1)}]^{p-1} \times [y_1 - hn^{-1/(p+q+1)},y_1 + hn^{-1/(p+q+1)}] \times \cdots \times [y_q - hn^{-1/(p+q+1)},y_q + hn^{-1/(p+q+1)}]$. Combining these arguments with (A.30) reveals that

$$\left|\mathrm{Prob}(A[\delta,n;h]) - \sum_{k=1}^{\infty}\mathrm{Prob}(\bar{A}[\delta,n;h;k])\,\frac{\tau(h)^k\bar{f}_x(1,0,y)^k}{k!}\,e^{-\tau(h)\bar{f}_x(1,0,y)}\right| \to 0 \tag{A.31}$$

as $n\to\infty$, where for a sequence $(\tilde{\theta}_{1,n},\tilde{\zeta}_{1,n},\widetilde{Y}_{1,n}),\ldots,(\tilde{\theta}_{k,n},\tilde{\zeta}_{k,n},\widetilde{Y}_{k,n})$ of k i.i.d. random variables uniformly distributed on $\bar{C}(h\cdot n^{-1/(p+q+1)})$, 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

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

$$\sum_{j=1}^{k} \alpha_j \tilde{Z}_{j,n} = 0 \quad \text{for } \tilde{Z}_{j,n} = \sum_{r=1}^{p-1} \zeta_{j,n,r} z^{(r)} \tag{A.32}$$

and

$$\sum_{j=1}^{k} \alpha_j \frac{1}{2g_x(0,y)} [\tilde{Z}_{j,n}^T g''_{x;zz}(0,y)\tilde{Z}_{j,n} + 2\tilde{Z}_{j,n}^T g''_{x;zy}(0,y)(\tilde{Y}_{j,n}-y) + (\tilde{Y}_{j,n}-y)^T g''_{x;yy}(0,y)(\tilde{Y}_{j,n}-y)] + \sum_{j=1}^{k} \alpha_j(1-\tilde{\theta}_{j,n}) \leq \delta \cdot n^{-2/(p+q+1)}. \tag{A.33}$$

The assertion of the proposition now follows from the fact that $\bar{A}[\delta,n;h;k]$ is realized iff the event $U[(\delta/h^2),k]$ is realized for $\tilde{\vartheta}_j = (1/h^2 n^{-2/(p+q+1)})(1-\tilde{\theta}_{j,n})$, $\tilde{\zeta}_j = (1/hn^{-1/(p+q+1)})\tilde{\zeta}_{j,n}$ and $\tilde{y}_j = (1/hn^{-1/(p+q+1)})(\tilde{Y}_{j,n}-y)$. It then follows that uniformity of $(\tilde{\theta}_{j,n},\tilde{\zeta}_{j,n},\tilde{Y}_{j,n})$ on $\bar{C}(h\cdot n^{-1/(p+q+1)})$ is equivalent to uniformity of $(\tilde{\vartheta}_j,\tilde{\zeta}_j,\tilde{y}_j)$ on $[0,1]\times[-1,1]^{p-1}\times[-1,1]^q$ and that (A.32) corresponds to (22). Finally, (A.33) implies (23) when γ is replaced by δ/h^2. ∎

Proof of Theorem 2. Let

$$F_{x,h}(\delta) = \sum_{k=1}^{\infty} \text{Prob}\left(U\left[\frac{\delta}{h^2},k\right]\right) \frac{\tau(h)^k \bar{f}_x(1,0,y)^k}{k!} e^{-\tau(h)\bar{f}_x(1,0,y)}. \tag{A.34}$$

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

$$\text{Prob}(A[\delta,n;h]) \leq \text{Prob}(A[\delta,n;h^*]) \leq \text{Prob}(A[\delta,n]) \leq 1 \tag{A.35}$$

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(i) we can therefore conclude that there exists a monotone function $F_x(\delta)$ such that

$$F_x(\delta) =: \lim_{h\to\infty} F_{x,h}(\delta) = \lim_{n\to\infty} \text{Prob}(A[\delta,n]). \tag{A.36}$$

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

It only remains to verify relation (26) and to show that F_x is continuous and that $F_x(\delta) \leq 1-\exp(-\delta \bar{f}_x(1,0,y)^{2/(p+q+1)}) < 1$. This requires a closer analysis of $\text{Prob}(U[(\delta/h^2),k])$. 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 that is smaller than or equal to t,

$$|\text{Prob}(U[\gamma,k]) - \text{Prob}(U[\gamma,[\lambda k]])| \leq d_0 \cdot \max\left\{\lambda-1, \frac{1}{\lambda}-1\right\} \tag{A.37}$$

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

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 an $N(\tau(h)\bar{f}_x(1,0,y),\ \tau(h)\bar{f}_x(1,0,y))$-distribution. In particular, with $s_{h,-} := \tau(h)\bar{f}_x(1,0,y) - \log h \cdot \sqrt{\tau(h)\bar{f}_x(1,0,y)}, s_{h,+} := \tau(h)\bar{f}_x(1,0,y) + \log h \cdot \sqrt{\tau(h)\bar{f}_x(1,0,y)}$ we obtain $\sum_{k=1}^{[s_{h,-}]}(\tau(h)^k\bar{f}_x(1,0,y)^k/k!)e^{-\tau(h)\bar{f}_x(1,0,y)} \to 0$ and $\sum_{k=[s_{h,+}]}^{\infty}(\tau(h)^k\bar{f}_x(1,0,y)^k/k!)e^{-\tau(h)\bar{f}_x(1,0,y)} \to 0$ as $h \to \infty$. Combining these arguments reveals

$$\begin{aligned} F_x(\delta) &= \lim_{h\to\infty} F_{x,h}(\delta) \\ &= \lim_{h\to\infty}\left\{\text{Prob}\left(U\left[\frac{\delta}{h^2},[\tau(h)\bar{f}_x(1,0,y)]\right]\right)\right. \\ &\qquad + \sum_{k=[s_{h,-}]}^{[s_{h,+}]}\left[\text{Prob}\left(U\left[\frac{\delta}{h^2},k\right]\right)\right. \\ &\qquad\qquad \left. - \text{Prob}\left(U\left[\frac{\delta}{h^2},[\tau(h)\bar{f}_x(1,0,y)]\right]\right)\right] \\ &\qquad \left.\times \frac{\tau(h)^k\bar{f}_x(1,0,y)^k}{k!}e^{-\tau(h)\bar{f}_x(1,0,y)}\right\} \\ &= \lim_{h\to\infty}\text{Prob}\left(U\left[\frac{\delta}{h^2},[\tau(h)\bar{f}_x(1,0,y)]\right]\right), \end{aligned} \tag{A.38}$$

when noting that for $k \in [[s_{h,-}],[s_{h,+}]]$ relation (A.17) implies that $|\text{Prob}(U[(\delta/h^2), k]) - \text{Prob}(U[(\delta/h^2),[\tau(h)\bar{f}_x(1,0,y)]])| \le d_0 \max\{([s_{h,+}]/[\tau(h)\bar{f}_x(1,0,y)]) - 1, ([\tau(h)\bar{f}_x(1,0,y)]/[s_{h,-}]) - 1\} \to 0$ as $h \to \infty$. Relation (26) then follows from

$$\lim_{h\to\infty}\text{Prob}\left(U\left[\frac{\delta}{h^2},[\tau(h)\bar{f}_x(1,0,y)]\right]\right) = \lim_{k\to\infty}\text{Prob}\left(U\left[\delta\,\frac{\bar{f}_x(1,0,y)^{2/(p+q+1)}}{k^{2/(p+q+1)}},k\right]\right), \tag{A.39}$$

and by using (27) the continuity of $F_x(\delta)$ for $\delta > 0$ is a consequence of

$$\begin{aligned} |F_x(\lambda\delta) - F_x(\delta)| &= \lim_{k\to\infty}\left|\text{Prob}\left(U\left[\delta\,\frac{\bar{f}_x(1,0,y)^{2/(p+q+1)}}{(k/\lambda^{(p+q+1)/2})^{2/(p+q+1)}},\frac{k\lambda^{(p+q+1)/2}}{\lambda^{(p+q+1)/2}}\right]\right)\right. \\ &\qquad \left. - \text{Prob}\left(U\left[\delta\,\frac{\bar{f}_x(1,0,y)^{2/(p+q+1)}}{(k/\lambda^{(p+q+1)/2})^{2/(p+q+1)}},\frac{k}{\lambda^{(p+q+1)/2}}\right]\right)\right| \\ &\le d_0 \cdot \max\left\{\lambda^{(p+q+1)/2} - 1, \frac{1}{\lambda^{(p+q+1)/2}} - 1\right\}. \end{aligned} \tag{A.40}$$

Clearly, the event $U[\delta(\bar{f}_x(1,0,y)^{2/(p+q+1)}/k^{2/(p+q+1)}), k]$ implies that $(\tilde{\vartheta}_j, \tilde{\zeta}_j, \tilde{y}_j) \in I_{k,\delta} := [0, \delta(\bar{f}_x(1,0,y)^{2/(p+q+1)}/k^{2/(p+q+1)})] \times [(-1/k^{1/(p+q+1)}), (1/k^{1/(p+q+1)})]^{p-1} \times [(-1/k^{1/(p+q+1)}), (1/k^{1/(p+q+1)})]^q$ for at least one observation

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press

$j \in \{1,\ldots,k\}$. Because $\text{Prob}(I_{k,\delta}) = \delta(\bar{f}_x(1,0,y)^{2/(p+q+1)}/k)$ for all sufficiently large k, standard arguments now lead to

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

$$\le \text{Prob}((\tilde{\vartheta}_j,\tilde{\zeta}_j,\tilde{Y}_j) \in I_{k,\delta} \quad \text{for some } j \in \{1,\ldots,k\})$$

$$= 1-\exp(-\delta\bar{f}_x(1,0,y)^{2/(p+q+1)}) \quad \text{as } k \to \infty. \qquad \textbf{(A.41)}$$

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

Proof of Corollary 1. By definition of θ, $\hat{\theta}$ we have $\theta(x,y)/\lambda = \theta(\lambda x,y)$ and $\hat{\theta}(x,y)/\lambda = \hat{\theta}(\lambda x,y)$. Consequently, $(\hat{\theta}(\lambda x,y)/\theta(\lambda x,y)) = (\hat{\theta}(x,y)/\theta(x,y))$ for all $\lambda > 0$. Furthermore, $F_{\lambda x} = F_x$ because $x/\|x\| = \lambda x/\|\lambda x\|$ implies that $g_x = g_{\lambda x}$ and also $f_x = f_{\lambda x}$. ■

Proof of Theorem 3. The bootstrap samples $\mathcal{S}_m^*$ can be represented equivalently by the samples $\tilde{\mathcal{S}}_m^* = \{(\theta_i^*,Z_i^*,Y_i^*)\}_{i=1}^m$ 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 (θ_i^*,Z_i^*,Y_i^*) to define events $A[\delta,m;h]^*$ and $A[\delta,m]^*$ and note that by a straightforward generalization of Lemma 2 $\text{Prob}[m^{2/(p+q+1)}((\hat{\theta}^*(x,y)/\theta(x,y))-1) \le \delta|\mathcal{S}_n] = \text{Prob}(A[\delta,m]^*|\mathcal{S}_n)$ holds for all m, δ. Theorem 2 implies $|m^{2/(p+q+1)}((\hat{\theta}(x,y)/\theta(x,y))-1)| \xrightarrow{p} 0$ as $n\to\infty$, and hence

$$\sup_\delta\left|\text{Prob}\left[m^{2/(p+q+1)}\left(\frac{\hat{\theta}^*(x,y)}{\hat{\theta}(x,y)}-1\right)\le\delta|\mathcal{S}_n\right]-\text{Prob}(A[\delta,m]^*|\mathcal{S}_n)\right| = o_p(1). \qquad \textbf{(A.42)}$$

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

$$\left|\frac{\text{Prob}((\theta_i^*,Z_i^*,Y_i^*)\in C(x,y;hm^{-1/(p+q+1)})|\mathcal{S}_n)}{\text{Prob}((\theta_i,Z_i,Y_i)\in C(x,y;hm^{-1/(p+q+1)}))}-1\right| = O_p(n^{(\kappa-1)/2}). \qquad \textbf{(A.43)}$$

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 $C(x,y;hn^{-1/(p+q+1)})$ asymptotically coincide:

$$\sup_{\mathcal{C}}\left|\frac{\text{Prob}[(\theta_i^*,Z_i^*,Y_i^*)\in\mathcal{C}|\mathcal{S}_n]}{\text{Prob}[(\theta_i^*,Z_i^*,Y_i^*)\in C(x,y;hm^{-1/(p+q+1)})|\mathcal{S}_n]} - \frac{\text{Prob}[(\theta_i,Z_i,Y_i)\in\mathcal{C}]}{\text{Prob}[(\theta_i,Z_i,Y_i)\in C(x,y;hm^{-1/(p+q+1)})]}\right| = o_p(1), \qquad \textbf{(A.44)}$$

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

This leads to $\sup_\delta |\text{Prob}(A[\delta, m; h]^* | \mathcal{S}_n) - \text{Prob}(A[\delta, m; h])| \xrightarrow{p} 0$ as $n \to \infty$. By arguments similar to those used to prove Theorem 1, it follows that for all $\epsilon > 0$ there exists an h_ϵ such that for every $h \geq h_\epsilon$, $\text{Prob}(\sup_\delta |\text{Prob}(A[\delta, m; h]^* | \mathcal{S}_n) - \text{Prob}(A[\delta, m])| \geq \epsilon) \xrightarrow{p} 0$ and $\text{Prob}(\sup_\delta |P(A[\delta, m; h]^* | \mathcal{S}_n) - P(A[\delta, m]^* | \mathcal{S}_n)| \geq \epsilon) \xrightarrow{p} 0$ as $n \to \infty$. The assertion of the theorem now follows from (A.42) and Theorems 1 and 2. ■

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 $\hat{g}_x^*$ to define events $A[\delta, n; h]^*$ and $A[\delta, n]^*$. First, note that for all n,

$$\text{Prob}\left(n^{2/(p+q+1)} \left(\frac{\hat{\theta}^*(x, y)}{\hat{\theta}(x, y)} - 1 \right) \leq \delta | \mathcal{S}_n \right) = \text{Prob}(A[\delta, n]^* | \mathcal{S}_n). \qquad \textbf{(A.45)}$$

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 $\hat{g}_x^*$ and $\hat{f}_x$ instead of g_x and f_x. This is easily seen when noting that $\hat{g}_x^*$ is necessarily convex and that with probability converging to 1 as $n \to \infty$ the bounds given in (A.13) and (A.25) also apply to $\hat{g}_x^*$. Because $n^{-1/(p+q+1)}/b \to 0$, the latter follows from (38) and Taylor expansions of g_x^* similar to (39). Furthermore, because of (40) relations (A.19)–(A.21) generalize to $\mathcal{S}_n^*$ and $\hat{f}_x$. Therefore for any $\epsilon > 0$ there exists an $h_\epsilon > 0$ such that for all $h \geq h_\epsilon$,

$$\text{Prob}\left(\sup_\delta [\text{Prob}(A[\delta, n]^* | \mathcal{S}_n) - \text{Prob}(A[\delta, n, h]^* | \mathcal{S}_n)] \leq \epsilon \right) \to 1 \quad \text{as } n \to \infty. \qquad \textbf{(A.46)}$$

On the other hand, in view of (38)–(40), one can invoke arguments similar to those used in the proof of Proposition 1 to obtain

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

The theorem now follows from Theorem 2. ■

https://doi.org/10.1017/S0266466608080651 Published online by Cambridge University Press