I N S T I T U T D E S T A T I S T I Q U E

B I O S T A T I S T I Q U E E T

S C I E N C E S A C T U A R I E L L E S

(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2013/12

When Bias Kills the Variance: Central Limit Theorems for DEA and FDH Efficiency Scores

KNEIP, A., SIMAR, L. and P. WILSON

When Bias Kills the Variance: Central Limit Theorems for DEA and FDH Efficiency Scores

Alois Kneip   Léopold Simar   Paul W. Wilson*

January 2013

Abstract

Data envelopment analysis (DEA) and free disposal hull (FDH) estimators are widely used to estimate efficiencies of production units. In applications, both efficiency scores for individual units as well as average efficiency scores are typically reported. While several bootstrap methods have been developed for making inference about the efficiencies of individual units, until now no methods have existed for making inference about mean efficiency levels. This paper shows that standard central limit theorems do not apply in the case of means of DEA or FDH efficiency scores due to the bias of the individual scores, which is of larger order than either the variance or covariances among individual scores. The main difficulty comes from the fact that such statistics depend on efficiency estimators evaluated at random points. Here, new central limit theorems are developed for means of DEA and FDH scores, and their efficacy for inference about mean efficiency levels is examined via Monte Carlo experiments.

*Kneip: Institute für Gessellschafts- und Wirtschaftswissennschafen, Statistische Abteilung, Universität Bonn, Bonn, Germany; email: akneip@uni-bonn.de.

Simar: Institut de Statistique, Biostatistique, et Sciences Actuarielles, Université Catholique de Louvain-la-Neuve, Louvain-la-Neuve, Belgium; email: leopold.simar@uclouvain.be.

Wilson: Department of Economics and School of Computing, Clemson University, Clemson, South Carolina, USA; email: pww@clemson.edu.

Research support from the "Interuniversity Attraction Pole", Phase VI (No. P6/03) of the Belgian Science Policy, from the INRA-GREMAQ, Toulouse, France, from U.S. National Science Foundation (grant no. SMA-1243436), and from the Cyber Infrastructure Technology Integration group at Clemson University is gratefully acknowledged. Any remaining errors are solely our responsibility.

1 Introduction

Nonparametric envelopment estimators are widely used to measure producers' performances. These estimators are based on estimators of the attainable set obtained by "enveloping" the observed cloud of points given by a sample of observed input and output levels of firms. Among estimators that have appeared in the literature, those that envelop the observed input-output combinations with a convex set are known as data envelopment analysis (DEA) estimators, while those that do not impose convexity are known as free-disposal hull (FDH) estimators.[1] The statistical properties of these estimators evaluated at a single, fixed point, including their asymptotic distributions and rates of convergence, are well known; see Simar and Wilson (2013) for a recent survey.

Sample means of efficiency estimates are frequently used to (i) summarize results, (ii) compare different groups of producers, (iii) to test hypotheses about returns to scale, convexity of the production set, etc., and in dynamic settings, to (iv) characterize changes in productivity, efficiency, technology, etc. While bootstrap methods are available for making inference about the efficiency of a single fixed point (e.g., see Kneip et al., 2008, 2011; Simar and Wilson, 2011a), to date little is known about the statistical properties of sample means of efficiency estimators. For practitioners, this is a serious problem as sample means are a useful way of summarizing results for large numbers of observations, making "on average" comparisons across groups, or construcitng test statistics to test various model features or restrictions.

This paper derives new results on the properties of sample means of nonparametric efficiency estimators. The problem is complicated because the estimators are biased, and in the case of sample means, the efficiency estimators are computed at random points. Hence the results existing to date are not helpful. The results presented below establish that existing central limit theorems (e.g., the Lindeberg-Feller Theorem) cannot be used for inference about population means except in a few special cases where the number of dimensions is quite small. New theorems are given establishing properties of moments of nonparametric efficiency estimators. The proofs are complicated due to the fact that there is a support boundary, i.e., a frontier, which affects the rate of convergence for points lying near the frontier.

The results in these theorems are then used to establish new central limit theorems that

[1] On December 19, 2012, Google Scholar returned about 38,300 results for "data envelopment analysis" and about 6,140 results for "FDH" and "efficiency."

confirm that whenever the number of dimensions exceeds a small number (depending on the particular estimator that is used), ordinary sample means of efficiency estimators will have limiting distributions involving unknown bias, or will be degenerate in the sense that variance tends to zero as sample size tends to infinity. The results are then used to fix this problem, providing several approaches to inference about (population) mean efficiency. A new central limit theorem is provided, involving an estimate of bias. In addition, a new, re-scaled estimator of mean efficiency is given along with a corresponding central limit theorem. This result allows construction of confidence intervals with asymptotically correct coverage. Finally, a simple trick allows re-centering of these confidence intervals to obtain confidence intervals of the same width, but with improved coverage.

The next section introduces the nonparametric efficiency estimator, establishes some notation, and describes the main problem. Section 3 gives the results on moments of the efficiency estimators when evaluated at random data points. These results are in turn used to establish the results allowing inference about mean efficiency in Section 4. The results are then extended to two-stage problems, where estimated efficiencies are regressed on some covariates; this problem has been examined in Simar and Wilson (2007, 2011b), but here some new results and strategies for inference are provided. Proofs of the theorems in Section 3 are given in Section 6, followed by simulation results in Section 7. Finally, concluding remarks are presented in Section 8.

2 DEA and FDH Estimators of Technical Efficiency

Consider a production process where input quantities $x \in \mathbb{R}_+^p$ are transformed into output quantities $y \in \mathbb{R}_+^q$. The production set

$$\Psi = \{(x, y) \in \mathbb{R}_+^{p+q} \mid x \text{ can produce } y\} \tag{2.1}$$

is the set of feasible combinations of inputs and outputs. The technology, or efficient frontier of Ψ, is defined by

$$\Psi^\partial = \left\{(x, y) \in \Psi \mid \left(\gamma^{-1}x, \gamma y\right) \notin \Psi \text{ for all } \gamma > 1\right\}. \tag{2.2}$$

The Farrell (1957) input-oriented measure of technical efficiency is given by

$$\theta(x, y) = \inf\{\theta > 0 \mid (\theta x, y) \in \Psi\}. \tag{2.3}$$

By construction, $\theta(x,y) \in (0,1]$ for all $(x,y) \in \Psi$. This measure gives the feasible, proportionate reduction in input levels, holding output levels constant, for a firm operating at $(x,y) \in \Psi$. If $\theta(x,y) = 1$, the firm is said to be technically efficient in the input direction, while if $\theta(x,y) < 1$ the firm is said to be technically *in*efficient.

Similar measures can be defined to measure technical efficiency in the output direction, in a hyperbolic direction, or in an arbitrary, linear direction toward the frontier; see Färe et al. (1985), Chambers et al. (1996), Simar and Wilson (2000), Wilson (2011), Simar and Vanhems (2012), and Simar et al. (2012) for details. For simplicity, the analysis below is presented only in terms of the input-oriented measure defined in (2.3); however, all of the results can be extended to the other directions after straightforward changes in notation.

Standard assumptions regarding the production set Ψ (e.g., Shephard, 1970; Färe, 1988; Simar and Wilson (2000); etc.) include the following.

Assumption 2.1. Ψ *is closed, and* Ψ^{∂} *exists.*

Assumption 2.2. *Both inputs and outputs are strongly disposable; i.e., for* $\widetilde{x} \geq x,\ 0 \leq \widetilde{y} \leq y$, *if* $(x,y) \in \Psi$ *then* $(\widetilde{x},y) \in \Psi$ *and* $(x,\widetilde{y}) \in \Psi$.[2]

Assumption 2.2 amounts to an assumption of weak monotonicity for the frontier, and is standard in micro-economic theory of the firm. Of course, the set Ψ is unobserved, and hence must be estimated from a sample $\mathcal{X}_n = \{(X_i, Y_i)\}_{i=1}^n$ of observed input-output pairs $X_i \in \mathbb{R}_+^p$, $Y_i \in \mathbb{R}_+^q$. Additional assumptions (e.g., convexity of Ψ, or assumptions about returns to scale) will be introduced later in Section 3.

Deprins et al. (1984) proposed estimating Ψ by the free-disposal hull of the sample observations in $\mathcal{X}_n$, i.e., by

$$\widehat{\Psi}_{\mathrm{FDH}}(\mathcal{X}_n) = \bigcup_{(X_i,Y_i)\in\mathcal{X}_n} \left\{(x,y) \in \mathbb{R}^{p+q} \mid y \leq Y_i,\ x \geq X_i\right\}. \tag{2.4}$$

Then the FDH estimator $\widehat{\theta}_{\mathrm{FDH}}(x,y \mid \mathcal{X}_n)$ of $\theta(x,y)$ is obtained by replacing Ψ on the right-hand side (RHS) of (2.3) with $\widehat{\Psi}_{\mathrm{FDH}}(\mathcal{X}_n)$.[3]

[2] Note that as usual, inequalities involving vectors are defined on an element-by-element basis.

[3] Afriat (1972, Theorem 1.1) defines a left- (but not right-) continuous function similar to the FDH estimator $\widehat{\Psi}_{\mathrm{FDH}}(\mathcal{X}_n)$ for the case $p \geq 1,\ q = 1$. However, $\widehat{\Psi}_{\mathrm{FDH}}(\mathcal{X}_n)$ is not a function, and is defined for arbitrary $p \geq 1$ as well as $q \geq 1$. Moreover, Afriat's function does not permit measurement of efficiency in the input direction, nor (in general) in hyperbolic or directional orientations.

Alternatively, if Ψ is believed to be convex, then Ψ can be estimated by the convex hull of $\widehat{\Psi}_{\text{FDH}}(\mathcal{X}_n)$ as in Farrell (1957) and Banker et al. (1984), i.e., by

$$\widehat{\Psi}_{\text{VRS}}(\mathcal{X}_n) = \left\{(x, y) \in \mathbb{R}^{p+q} \mid y \leq \boldsymbol{Y}\boldsymbol{\omega},\ x \geq \boldsymbol{X}\boldsymbol{\omega},\ \boldsymbol{i}_n'\boldsymbol{\omega} = 1,\ \boldsymbol{\omega} \in \mathbb{R}_n\right\}, \tag{2.5}$$

where $\boldsymbol{X} = \big(X_1, \ldots, X_n\big)$ and $\boldsymbol{Y} = \big(Y_1, \ldots, Y_n\big)$ are $(p \times n)$ and $(q \times n)$ matrices of input and output vectors, respectively; $\boldsymbol{i}_n$ is an $(n \times 1)$ vector of ones, and $\boldsymbol{\omega}$ is a $(n \times 1)$ vector of weights. This is the VRS-DEA estimator of Ψ, and the VRS-DEA estimator of $\theta(x, y)$ is obtained by replacing Ψ on the RHS of (2.3) with $\widehat{\Psi}_{\text{VRS}}(\mathcal{X}_n)$.

If Ψ^∂ exhibits globally constant returns to scale (CRS), i.e, if $(ax, ay) \in \Psi$ for all $(x, y) \in \Psi$ and $a \in [0, \infty)$, then Ψ can be estimated by the CRS version of the DEA estimator of Ψ obtained by dropping the constraint $\boldsymbol{i}_n\boldsymbol{\omega} = 1$ from the RHS of (2.5). The resulting estimator of Ψ, used by Charnes et al. (1978) and denoted by $\widehat{\Psi}_{\text{CRS}}(\mathcal{X}_n)$, is the conical hull of $\widehat{\Psi}_{\text{FDH}}(\mathcal{X}_n)$. The CRS-DEA estimator $\widehat{\theta}_{\text{VRS}}(x, y \mid \mathcal{X}_n)$ of $\theta(x, y)$ is obtained by replacing Ψ on the RHS of (2.3) with $\widehat{\Psi}_{\text{CRS}}(\mathcal{X}_n)$.

Computation of the FDH and DEA efficiency estimators is straightforward. FDH efficiency estimates can be computed as

$$\widehat{\theta}_{\text{FDH}}(x, y) = \min_{i \in \mathcal{I}(y)} \left(\max_{j=1,\, \ldots,\, p} \left(\frac{X_i^j}{x^j} \right) \right), \tag{2.6}$$

where $\mathcal{I}(y) = \{i \mid Y_i \geq y,\ i = 1, \ldots, n\}$ and X_i^j, x^j are the jth elements of X_i and x, respectively (throughout, subscripts will be used to index different vectors, while superscripts will be used to index elements of vectors). DEA efficiency estimates are typically computed by solving linear programs; for the VRS-DEA estimator, one can compute

$$\widehat{\theta}_{\text{VRS}}(\boldsymbol{x}, \boldsymbol{y}) = \min_{\theta, \boldsymbol{\omega}} \left\{\theta \mid y \leq \boldsymbol{Y}\boldsymbol{\omega},\ \theta x \geq \boldsymbol{X}\boldsymbol{\omega},\ \boldsymbol{i}_n'\boldsymbol{\omega} = 1,\ \boldsymbol{\omega} \in \mathbb{R}_n\right\}. \tag{2.7}$$

The CRS-DEA estimator $\widehat{\theta}_{\text{CRS}}(x, y \mid \mathcal{X}_n)$ can be computed similarly by dropping the constraint $\boldsymbol{i}_n'\boldsymbol{\omega} = 1$ on the RHS of (2.7).

Asymptotic properties of input-oriented VRS-DEA efficiency estimators are investigated in Kneip et al. (1998), Jeong (2004), Jeong and Park (2006), Kneip et al. (2008), Jeong et al. (2010); and for the input-oriented FDH efficiency estimator by Park et al. (2000). These results have been extended to the hyperbolic and directional orientations by Wilson (2011), Simar and Vanhems (2012), and Simar et al. (2012), with similar limiting distributions and rates of convergence. In each case, the estimators are consistent under appropriate assumptions, and converge at rate n^{κ}, where $\kappa = 2/(p+q+1)$, $2/(p+q)$ or $1/(p+q)$ for the VRS-DEA, CRS-DEA, and FDH cases, respectively, and have limiting distributions.

To date, there are no tractable, analytical expressions for the asymptotic distributions of the VRS-DEA and CRS-DEA efficiency estimators. The FDH estimators have been shown to have limiting Weibull distributions, but these involve unknown parameters that are difficult to estimate. Consequently, bootstrap methods appear to be the only practical avenue toward inference on $\theta(x, y)$; see Kneip et al. (2008, 2011) and Simar and Wilson (2011a) for results on consistent inference about $\theta(x, y)$ using bootstrap or subsampling methods.

As noted in Section 1, however, much less is known about how to make inference about the population mean $\mu_\theta = E(\theta(X, Y))$ from a sample $\mathcal{X}_n$ of n identically, independently distributed (iid) observations (X_i, Y_i). One might wish to make inference about μ_θ for several reasons. For example, means are often used to summarize results, to compare efficiency among different groups of producers, to characterize what one might expect "on average," etc. An empirical mean of VRS-DEA, CRS-DEA, or FDH estimators, using the n observations in $\mathcal{X}_n$, might seem to be a natural estimator of μ_θ. For example, one might use

$$\widehat{\mu}_n = n^{-1} \sum_{i=1}^{n} \widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) \tag{2.8}$$

to estimate μ_θ where $\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n)$ denotes either the VRS-DEA, CRS-DEA, or FDH estimator of $\theta(X_i, Y_i)$. The notation on the RHS of (2.8) makes explicit that the estimators are conditional on the sample $\mathcal{X}_n$; this notation will be used as necessary to avoid confusion.

Sample means are arguably the most common statistics in use, and existence of the well-known set of central limit theorem results makes inference straightforward in many contexts. Here, however, a number of problems arise. The sample mean in (2.8) involves a mean of estimators, as opposed to true values $\theta(X_i, Y_i)$. In addition, the $\widehat{\theta}(X_i, Y_i)$ on the RHS of (2.8) are evaluated at random points (X_i, Y_i), instead of at fixed points.[4] This is an important distinction, because all of the available results on VRS-DEA, CRS-DEA, and FDH estimators cited above are for fixed points, and not for random points. Since the $\widehat{\theta}(X_i, Y_i)$ are evaluated

[4] By "fixed point," we mean a point chosen, perhaps arbitrarily, by the researcher. This could correspond to one of the observed input-output pairs, or could represent an hypothetical firm. In either case, the point is non-stochastic and is not a realization of the random variables (X_i, Y_i).

at random points, one must consider covariances among the terms on the RHS of (2.8). Still another complication arises from the fact that only points in a neighborhood of the frontier (as opposed to those in the interior of Ψ lying "far" from the frontier Ψ^{∂}) have potential to affect $\widehat{\theta}(X_i, Y_i)$, and some of the (X_i, Y_i) on the RHS of (2.8) may fall near the frontier. As will be seen below, the bias of the estimators of $\theta(X_i, Y_i)$ turns out to be far more critical than the covariance. In fact, due to the rates of the bias and variance, standard central limit theorem results cannot be used with (2.8) to make inference about μ_θ except in special cases where the number of dimension $(p+q)$ is exceptionally small.

3 Asymptotic moments of efficiency estimators

This section presents new results on the moments of VRS-DEA, CRS-DEA, and FDH efficiency estimators when evaluated at random points; these results will be used later in Section 4 to (i) to demonstrate why standard central limit theorems such as the Lindeberg-Feller Theorem cannot be used in the case of sample means of nonparametric efficiency estimators when there are more than a small number of dimensions; (ii) derive new results which can be used to make inference about μ_θ, and (iii) in Section 5 to show that there are additional problems, beyond those described by Simar and Wilson (2007, 2011b), when efficiency estimates from a first-stage analysis are regressed on some covariates in a second stage.

Some additional assumptions are needed. Proofs of the theorems, which can be skipped by less-technical readers, are given in Section 6. The following assumptions are needed for the case of the VRS-DEA estimator; some of these will be used also for the CRS-DEA and FDH cases.

Assumption 3.1. *(i) The random variables (X, Y) possess a joint density f with support $\mathcal{D} \subset \Psi$; and (ii) f is continuously differentiable on $\mathcal{D}$.*

Assumption 3.2. *(i) $\mathcal{D}^* := \{\theta(x,y)x, y) \mid (x,y) \in \mathcal{D}\} \subset \mathcal{D}$; (ii) $\mathcal{D}^*$ is compact; and (iii) $f(\theta(x,y)x, y) > 0$ for all $(x,y) \in \mathcal{D}$.*

Assumption 3.3. *$\theta(x,y)$ is three times continuously differentiable on $\mathcal{D}$.*

Assumption 3.4. *$\mathcal{D}$ is* almost strictly convex*; i.e., for any $(x,y), (\widetilde{x}, \widetilde{y}) \in \mathcal{D}$ with $(\frac{x}{\|x\|}, y) \neq (\frac{\widetilde{x}}{\|\widetilde{x}\|}, \widetilde{y})$, the set $\{(x^*, y^*) \mid (x^*, y^*) = (x,y) + \alpha((\widetilde{x}, \widetilde{y}) - (x,y))$ for some $0 < \alpha < 1\}$ is a subset of the* interior *of $\mathcal{D}$.*

Assumptions 3.1–3.3 are similar to assumptions needed by Kneip et al. (2008) to establish the limiting distribution of the VRS-DEA estimator, except that there, $\theta(x, y)$ was only required to be twice continuously differentiable. Here, the addition of Assumption 3.4 and the additional smoothness of $\theta(x, y)$ in Assumption 3.3 are needed to establish results beyond those obtained in Kneip et al. (2008).

The first result gives moments of the VRS-DEA estimator of $\theta(X_i, Y_i)$.

Theorem 3.1. *Under Assumptions 2.1–2.2 and 3.1–3.4, there exists a constant* $0 < C_0 < \infty$ *such that for all* $i, j \in \{1, \ldots, n\}$, $i \neq j$,

$$E\left(\widehat{\theta}_{VRS}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right) = C_0 n^{-\frac{2}{p+q+1}} + O\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{p+q+4}{p+q+1}}\right), \tag{3.1}$$

$$VAR\left(\widehat{\theta}_{VRS}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right) = O\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}\right), \tag{3.2}$$

and

$$\left|COV\Big(\widehat{\theta}_{VRS}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i), \widehat{\theta}_{VRS}(X_j, Y_j \mid \mathcal{X}_n) - \theta(X_j, Y_j)\Big)\right| = O\left(n^{-\frac{p+q+2}{p+q+1}} (\log n)^{\frac{p+q+2}{p+q+1}}\right) = o\left(n^{-1}\right). \tag{3.3}$$

The value of the constant C_0 *depends on* f *and on the structure of the set* $\mathcal{D} \subset \Psi$.

As will be seen in the proof of Theorem 3.1 in Section 6, for any (x, y) in the interior of $\mathcal{D}$, the asymptotic variance of the VRS-DEA estimator is of order $n^{-\frac{4}{p+q+1}}$. The slower rate of convergence established in (3.2) is due to (a rough approximation of) boundary effects.

For the case of the CRS-DEA estimator, Assumption 3.4 must be replaced by the following condition.

Assumption 3.5. *(i) For any* $(x, y) \in \Psi$ *and any* $a \in [0, \infty)$, $(ax, ay) \in \Psi$*; (ii) the support* $\mathcal{D} \subset \Psi$ *of* f *is such that for any* (x, y), $(\widetilde{x}, \widetilde{y}) \in \mathcal{D}$ *with* $(\frac{x}{\|x\|}, \frac{y}{\|y\|}) \neq (\frac{\widetilde{x}}{\|\widetilde{x}\|}, \frac{\widetilde{y}}{\|\widetilde{y}\|})$*, the set* $\{(x^*, y^*) \mid (x^*, y^*) = (x, y) + \alpha((\widetilde{x}, \widetilde{y}) - (x, y))$ *for some* $0 < \alpha < 1\}$ *is a subset of the* interior *of* $\mathcal{D}$*; and (iii)* $(x, y) \notin \mathcal{D}$ *for any* $(x, y) \in \mathbb{R}_+^p \times \mathbb{R}^q$ *with* $y^1 = 0$*, where* y^1 *denotes the first element of the vector* y.

The conditions on the structure of Ψ (and $\mathcal{D}$) given in Assumptions 3.4 and 3.5 are incompatible. It is not possible that both assumptions hold simultaneously.

In the following, for any compact, convex set $\mathcal{H} \subset \mathbb{R}_+^p \times \mathbb{R}_+^q$, let $\widehat{\theta}_{\rm CH}(x, y \mid \mathcal{H}) := \min\{\theta > 0 \mid (\theta x, y) \in \mathcal{H}\}$ for all (x, y) with $(ax, y) \in \mathcal{H}$ for some $a > 0$. Furthermore, let $\mathcal{H}_n^0$ denote the convex hull of $\mathcal{X}_n^0 := \mathcal{X}_n \cup \{(0, 0)\}$.

Theorem 3.2. *Under Assumptions 2.1–2.2, 3.1–3.3, and 3.5, the following results hold for any $(x, y) \in \mathcal{H}_n^0$:*

(i) For any $y = (y^1, \ldots, y^q)' \in \mathbb{R}^q$ with $y^1 > 0$, define the $q - 1$ dimensional vector $\widetilde{y} = (y^2/y^1, \ldots, y^q/y^1)'$, and let $\mathcal{X}_n^* = \left\{(X_i/y_i^1, \tilde{Y}_i)\right\}_{i=1}^n$. Then

$$\widehat{\theta}_{CRS}(x, y \mid \mathcal{X}_n) = \widehat{\theta}_{CH}(x/y^1, \widetilde{y} \mid \mathcal{H}_n^*), \tag{3.4}$$

where $\mathcal{H}_n^*$ is the convex hull of $\mathcal{X}_n^*$. Furthermore, let $\mathcal{X}_{n,\geq y^1} := \{(X_i, Y_i) \mid y_i^1 \geq y^1\}$, and let $\mathcal{H}_{n,\geq y^1}^*$ be the convex hull of $\mathcal{X}_{n,\geq y^1}^* = \{(X_i/y_{i1}, \widetilde{Y}_i) \mid y_i^1 \geq y^1\}$. Then

$$\widehat{\theta}_{CRS}(x, y \mid \mathcal{X}_n) \leq \widehat{\theta}_{VRS}(x, y \mid \mathcal{X}_n^0) \leq \widehat{\theta}_{CRS}(x, y \mid \mathcal{X}_{n,\geq y^1}) = \widehat{\theta}_{CH}(x/y_1, \widetilde{y} \mid \mathcal{H}_{n,\geq y^1}^*). \tag{3.5}$$

(ii) There exists a constant $0 < C_1 < \infty$ such that for all $i, j \in \{1, \ldots, n\}$, $i \neq j$,

$$E\left[\widehat{\theta}_{CRS}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right] = C_1 n^{-\frac{2}{p+q}} + O\left(n^{-\frac{3}{p+q}} (\log n)^{\frac{p+q+3}{p+q}}\right), \tag{3.6}$$

$$VAR\left(\widehat{\theta}_{CRS}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right) = O\left(n^{-\frac{3}{p+q}} (\log n)^{\frac{3}{p+q}}\right), \tag{3.7}$$

and

$$\left| COV\Big(\widehat{\theta}_{CRS}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i), \widehat{\theta}_{CRS}(X_j, Y_j \mid \mathcal{X}_n) - \theta(X_j, Y_j)\Big) \right|$$
$$= O\left(n^{-\frac{p+q+1}{p+q}} (\log n)^{\frac{p+q+1}{p+q}}\right) = o\left(n^{-1}\right). \tag{3.8}$$

The value of the constant C_1 depends on f and on the structure of the set $\mathcal{D} \subset \Psi$.

Part (i) of Theorem 3.2 is a key for deriving part (ii), but is otherwise not directly necessary for deriving the central limit theorem results below in Section 4. Since the number of observations with $y_i^1 \geq y$ will be proportional to n, the inequality in (3.5) indicates that under the assumption of CRS, the ordinary VRS-DEA estimator (when adding the point $(0, 0)$ to $\mathcal{X}_n$) also converges at rate $n^{-\frac{2}{p+q}}$. This result is new and unexpected.

Turning now to the FDH estimator, the following assumption is needed.

Assumption 3.6. *(i) $\theta(x,y)$ is twice continuously differentiable on $\mathcal{D}$; and (ii) all the first-order partial derivatives of $\theta(x,y)$ with respect to x and y are nonzero at any point $(x,y) \in \mathcal{D}$.*

Note that the free disposability assumed in Assumption 2.2 implies that $\theta(x,y)$ is monotone, increasing in x and monotone, decreasing in y. Assumption 3.6 additionally requires that the frontier is strictly monotone and does not possess constant segments. Finally, part (i) of Assumption 3.6 is weaker than Assumption 3.3.

Theorem 3.3. *Under Assumptions 2.1–2.2, 3.1, 3.2, and 3.6, there exists a constant $0 < C_2 < \infty$ such that for all $i, j \in \{1, \ldots, n\}$, $i \neq j$,*

$$E\left(\widehat{\theta}_{FDH}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right) = C_2 n^{-\frac{1}{p+q}} + O\left(n^{-\frac{2}{p+q}} (\log n)^{\frac{p+q+2}{p+q}}\right), \tag{3.9}$$

$$VAR\left(\widehat{\theta}_{FDH}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right) = O\left(n^{-\frac{2}{p+q}} (\log n)^{\frac{2}{p+q}}\right), \tag{3.10}$$

and

$$\left| COV\Big(\widehat{\theta}_{FDH}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i), \widehat{\theta}_{FDH}(X_j, Y_j \mid \mathcal{X}_n) - \theta(X_j, Y_j)\Big) \right|$$
$$= O\left(n^{-\frac{p+q+1}{p+q}} (\log n)^{\frac{p+q+1}{p+q}}\right) = o\left(n^{-1}\right). \tag{3.11}$$

The value of the constant C_2 depends on f and on the structure of the set $\mathcal{D} \subset \Psi$.

4 Asymptotic distribution of $\widehat{\mu}_n$ and inference on μ_θ

As noted earlier, the results from the previous section can now be used to explain why existing central limit theorems are inapplicable when using sample means of VRS-DEA, CRS-DEA, or FDH efficiency estimators in more than 2, 3, or 1 dimensions (respectively) to make inference about μ_θ, and to derive new results that permit inference about μ_θ.

In order to simplify notation, results from Section 3 can be summarized by writing

$$E\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right) = Cn^{-\kappa} + R_{n,\kappa}, \tag{4.1}$$

where $R_{n,k} = o\left(n^{-\kappa}\right)$,

$$E\left(\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right)^2\right) = o\left(n^{-\kappa}\right), \tag{4.2}$$

and

$$\left|\text{COV}\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i),\ \widehat{\theta}(X_j, Y_j \mid \mathcal{X}_n) - \theta(X_j, Y_j)\right)\right| = o\left(n^{-1}\right) \tag{4.3}$$

for all $i, j \in \{1, \ldots, n\}$, $i \neq j$. The values of the constant C, the rate κ, and the remainder term $R_{n,\kappa}$ depends on which estimator is used (here, we suppress the labels "VRS," "CRS," or "FDH" on $\widehat{\theta}$ and $\widehat{\mu}_n$. Of course, the results outlined here depend on the corresponding relevant assumptions required by Theorems 3.1–3.3. Under VRS with the VRS-DEA estimator, $\kappa = 2/(p+q+1)$ and $R_{n,\kappa} = O(n^{-3\kappa/2}(\log n)^{\alpha_1})$; under CRS with either the VRS-DEA or VRS-CRS estimator, $\kappa = 2/(p+q)$ and $R_{n,\kappa} = O(n^{-3\kappa/2}(\log n)^{\alpha_2})$; while under only the free disposability assumption (but not necessarily CRS or convexity) with the FDH estimator, we have $\kappa = 1/(p+q)$ and $R_{n,\kappa} = O(n^{-2\kappa}(\log n)^{\alpha_3})$. The values of $\alpha_j > 1$, $j = 1, 2, 3$ are given in the theorems from Section 3. For purposes of the results in this section, the $\log n$ factor appearing in the theorems of Section 3 will not play a role. Most of the results below rely on the fact that in each case, $R_{n,\kappa} = o\left(n^{-\kappa}\right)$; the remainder term will only be considered when it is possible to obtain asymptotic refinements.

Denote $\mu_\theta = E(\theta(X, Y))$ and $\sigma_\theta^2 = \text{VAR}(\theta(X, Y))$, and assume both quantities are finite. In order to make inference about μ_θ, consider the quantities

$$\overline{\theta}_n = n^{-1} \sum_{i=1}^{n} \theta(X_i, Y_i) \tag{4.4}$$

and

$$\widehat{\mu}_n = n^{-1} \sum_{i=1}^{n} \widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n). \tag{4.5}$$

Under mild assumptions, the Lindeberg-Feller Central Limit Theorem establishes the limiting distribution of $\overline{\theta}_n$; i.e.,

$$\sqrt{n}\left(\overline{\theta}_n - \mu_\theta\right) \xrightarrow{\mathcal{L}} N(0, \sigma_\theta^2). \tag{4.6}$$

Of course, $\overline{\theta}_n$ is unobserved, as is μ_θ; as noted in Section 1, $\widehat{\mu}_n$ is typically used to estimate μ_θ. The following basic result will be useful for examining the properties of $\widehat{\mu}_n$.

Lemma 4.1. *Under the appropriate set of Assumptions described in Theorem 3.1, for the VRS-DEA estimator, Theorem 3.2 for the CRS-DEA estimator, or Theorem 3.3 for the FDH estimator, with* $\kappa = 2/(p+q+2)$, $2/(p+q)$, *or* $1/(p+q)$, *respectively, we have*

$$E\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n)\right) = \mu_\theta + Cn^{-\kappa} + o\left(n^{-\kappa}\right) \tag{4.7}$$

and

$$VAR\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n)\right) = \sigma_\theta^2 + o\left(n^{-\kappa/2}\right). \tag{4.8}$$

Proof. Assertion 4.7 follows directly from the theorems in Section 3. To prove 4.8, first note that

$$\begin{aligned}\mathrm{VAR}(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n)) =& E\left(\left[\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right]^2\right) + E\left(\left[\theta(X_i, Y_i) - E\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n)\right)\right]^2\right) \\ &+ 2E\left(\left[\theta(X_i, Y_i) - E\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n)\right)\right]\left[\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right]\right).\end{aligned}$$

Using (4.1),

$$\begin{aligned}E\left(\left[\theta(X_i, Y_i) - E\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n)\right)\right]^2\right) &= \sigma_\theta^2 + \left[E\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right)\right]^2 \\ &= \sigma_\theta^2 + C^2 n^{-2\kappa} + o\left(n^{-2\kappa}\right).\end{aligned} \tag{4.9}$$

Using the Cauchy-Schwartz inequality, (4.2) and (4.9), the last term in $\mathrm{VAR}(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n))$ can be bounded by $o\left(n^{-\kappa/2}\right)$, completing the proof. ■

The following theorem provides a consistent estimator of σ_θ^2, and establishes the basic properties of $\widehat{\mu}_n$.

Theorem 4.1. *Let* $\widetilde{\mu}_n = E(\widehat{\mu}_n)$. *Under the assumptions of Lemma 4.1, the following conditions hold: (i)* $\widetilde{\mu}_n = \mu_\theta + Cn^{-\kappa} + R_{n,\kappa}$; *(ii)* $\widehat{\mu}_n - \widetilde{\mu}_n = \bar{\theta}_n - \mu_\theta + o_p(n^{-1/2})$; *(iii)* $\sqrt{n}\left(\widehat{\mu}_n - \widetilde{\mu}_n\right) \xrightarrow{\mathcal{L}} N(0, \sigma_\theta^2)$; *and (iv)* $\widehat{\sigma}^2_{\theta,n} = n^{-1}\sum_{i=1}^n \left[\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \widehat{\mu}_n\right]^2 \xrightarrow{p} \sigma_\theta^2$.

Proof. Consider the sequence of random variables $\zeta_n = n^{-1}\sum_{i=1}^n \left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right)$. From (4.1) we have $E(\zeta_n) = Cn^{-\kappa} + R_{n,\kappa}$, and using (4.2) and (4.3) we obtain

$$\begin{aligned}\mathrm{VAR}(\zeta_n) &= n^{-2}\sum_{i=1}^n \mathrm{VAR}\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right) + o\left(n^{-2}\right) \\ &= n^{-2}\sum_{i=1}^n \left[E\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right)^2 + \left(E\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right)\right)^2\right] \\ &= n^{-1} o\left(n^{-\kappa}\right).\end{aligned}$$

Since $\widetilde{\mu}_n = \mu_\theta + E(\zeta_n)$, we have (i). For part (ii), we have $\widehat{\mu}_n - \widetilde{\mu}_n = \overline{\theta}_n - \mu_\theta + \eta_n$ where $\eta_n = \zeta_n - E(\zeta_n)$ has mean zero and variance $\mathrm{VAR}(\eta_n) = \mathrm{VAR}(\zeta_n) = n^{-1}o(n^{-\kappa})$. Hence

$\eta_n = o_p\left(n^{-1/2}\right)$. Part (iii) is a direct consequence of (ii). The proof of (iv) is also direct:

$$
\begin{aligned}
\widehat{\sigma}^2_{\theta,n} &= n^{-1}\sum_{i=1}^{n}(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n))^2 - \widehat{\mu}_n^2 \\
&\xrightarrow{p} E[(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n))^2] - \mu_\theta^2 \\
&= \text{VAR}(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n)) + \left[E\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n)\right)\right]^2 - \mu_\theta^2.
\end{aligned}
$$

Using the results of Lemma 4.1 yields the desired result. ■

This theorem shows in particular that in each of the three settings (i.e., VRS-DEA, CRS-DEA, or FDH), and under the appropriate set of assumptions, $\widehat{\mu}_n$ is a consistent estimator of μ_θ, with a bias term of order $Cn^{-\kappa}$. But it also illustrates the fact that the bias will kill the variance if we want to use $\widehat{\mu}_n$ to make inference about μ_θ. This can be seen by writing result (iii) explicitly as

$$
\sqrt{n}\left(\widehat{\mu}_n - \mu_\theta - Cn^{-\kappa} - R_{n,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, \sigma_\theta^2). \tag{4.10}
$$

If $\kappa > 1/2$, the bias term in (4.10) is dominated by the factor $\sqrt{n}$ and thus can be ignored; in this case, standard, conventional methods can be used to obtain confidence intervals for μ_θ. Otherwise, the bias is constant if $\kappa = 1/2$, or explodes if $\kappa < 1/2$. Note that $\kappa > 1/2$ if and only if $p+q \leq 2$ in the VRS case, or if and only if $p+q \leq 3$ in the CRS case. In the FDH case, this occurs only in the univariate case with $p = 1$, $q = 0$. Replacing the scale factor $\sqrt{n}$ in (4.10) with n^γ, with $\gamma < \kappa \leq 1/2$, is not a viable option. Although doing so would make the bias disappear as $n \to \infty$, it would cause the variance to converge to zero whenever $\kappa \leq 1/2$.

In general, whenever $\kappa \leq 1/2$, Theorem 4.1 makes clear that additional work is needed to make inference about the mean μ_θ in general situations. The results so far suggest either (i) using a different estimator for μ_θ, or (ii) incorporating a suitable estimator of the bias.

An easy way to address the issue of controlling both bias and variance, for general number of dimensions $(p+q)$, is to rescale the *estimator* of the population mean μ_θ by an appropriate factor different from $\sqrt{n}$ when $\kappa \leq 1/2$. Consider the factor $n_\kappa = [n^{2\kappa}] \leq n$, where $[a]$ denotes the integer part of a (note that this covers the limiting case of $\kappa = 1/2$). Then assume the observations in the sample $\mathcal{X}_n$ are randomly ordered, and consider the latent estimator

$$
\overline{\theta}_{n_\kappa} = n_\kappa^{-1}\sum_{i=1}^{n_\kappa}\theta(X_i, Y_i). \tag{4.11}
$$

Of course, $\overline{\theta}_{n_\kappa}$ is unobserved, but it can be estimated by

$$\widehat{\mu}_{n_\kappa} = n_\kappa^{-1} \sum_{i=1}^{n_\kappa} \widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n), \tag{4.12}$$

where the notation $\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n)$ serves to remind the reader that the individual efficiency estimates are computed from the full sample of n observations. Here again, one can use either the VRS, CRS or FDH version of the estimator. Under the appropriate set of assumptions, the properties of this estimator are given in the next theorem.

Theorem 4.2. *Under the assumptions of Lemma 4.1, for cases where $\kappa \leq 1/2$, as $n \to \infty$,*

$$n^\kappa \left(\widehat{\mu}_{n_\kappa} - \mu_\theta - Cn^{-\kappa} - R_{n,\kappa} \right) \xrightarrow{\mathcal{L}} N(0, \sigma_\theta^2). \tag{4.13}$$

Proof. Since $n_\kappa = n^{2\kappa} \to \infty$ as $n \to \infty$, $\sqrt{n_\kappa} \left(\overline{\theta}_{n_\kappa} - \mu_\theta \right) \xrightarrow{\mathcal{L}} N\left(0, \sigma_\theta^2\right)$. The result follows by the same arguments leading to Theorem 4.1. In particular,

$$\left(\widehat{\mu}_{\theta,n_\kappa} - \mu_\theta\right) = \left(\overline{\theta}_{n_\kappa} - \mu_\theta\right) + n_\kappa^{-1} \sum_{i=1}^{n_\kappa} \left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i) \right). \tag{4.14}$$

The right hand term has mean given by (4.1), i.e., $Cn^{-\kappa} + R_{n,\kappa}$, and variance $(1/n_\kappa) o\left(n^{-\kappa}\right)$. Multiplying the two terms of the equation by $\sqrt{n_\kappa} = n^\kappa$ yields the result. ■

Since Theorem 4.2 establishes that $\sqrt{n_\kappa} \left(\widehat{\mu}_{n_\kappa} - \mu_\theta\right)$ has a limiting distribution, with unknown mean, bootstrap approaches could be used to estimate this bias and so to provide confidence intervals for μ_θ (note that the variance could also be estimated by the same bootstrap, or by the consistent estimator $\widehat{\sigma}^2_{\theta,n}$ defined above). In theory, subsampling along the lines of Simar and Wilson (2011a) could also be used to make consistent inference about μ_θ. However, the estimator in (4.12) uses only a subset of the original n observations; unless n is extraordinarily large, taking subsamples among a subset of n_κ observations will leave too little information to provide useful inference.

However, Theorem 4.1 provides another way to correct for the bias in (4.13). Assume that the observations (X_i, Y_i) are randomly ordered, and let $\mathcal{X}^{(1)}_{n/2}$ denote the set of the first $n/2$ observations in $\mathcal{X}_n$; let $\mathcal{X}^{(2)}_{n/2}$ denote the set of remaining observations from $\mathcal{X}_n$ (for simplicity, assume n is even). Let $\widehat{\mu}^{(j)}_{n/2} = 2n^{-1} \sum_{i=1}^{n/2} \widehat{\theta}(X_i, Y_i \mid \mathcal{X}^{(j)}_{n/2})$, where $(X_i, Y_i) \in \mathcal{X}^{(j)}_{n/2}$ for $j \in \{1,\ 2\}$. Then set

$$\widehat{\mu}^*_{n/2} = \left(\widehat{\mu}^{(1)}_{n/2} + \widehat{\mu}^{(2)}_{n/2} \right) /2. \tag{4.15}$$

It follows from Theorem 4.1 (ii) that as $n \to \infty$,

$$\widehat{\mu}_{n/2}^{(j)} - \widetilde{\mu}_{n/2} = 2n^{-1} \sum_{i=1}^{n/2} (\theta(X_i, Y_i) - \mu_\theta) + o_p\left(n^{-1/2}\right) \tag{4.16}$$

for $j \in \{1,\ 2\}$. Consequently,

$$\widehat{\mu}_{n/2}^{*} - \widetilde{\mu}_{n/2} = n^{-1} \sum_{i=1}^{n} (\theta(X_i, Y_i) - \mu_\theta) + o_p\left(n^{-1/2}\right), \tag{4.17}$$

while for the original estimator the result in Theorem 4.1 (ii) holds. Subtracting the result in Theorem 4.1 (ii) from (4.17) and rearranging terms yields

$$\widehat{\mu}_{n/2}^{*} - \widehat{\mu}_n = \widetilde{\mu}_{n/2} - \widetilde{\mu}_n + o_p\left(n^{-1/2}\right), \tag{4.18}$$

which makes clear that the difference $\left(\widehat{\mu}_{n/2}^{*} - \widehat{\mu}_n\right)$ reflects the bias differences. Moreover, the estimation error is of order smaller than $n^{1/2}$. On the other hand, Theorem 4.1 (i) implies that

$$\widetilde{\mu}_{n/2} - \widetilde{\mu}_n = C(2^\kappa - 1)n^{-\kappa} + R_{n,\kappa}, \tag{4.19}$$

where the remainder has the same order as the original $R_{n,\kappa}$. Therefore,

$$(2^\kappa - 1)^{-1}\left(\widehat{\mu}_{n/2}^{*} - \widehat{\mu}_n\right) = Cn^{-\kappa} + R_{n,\kappa} + o_p\left(n^{-1/2}\right) \tag{4.20}$$

provides an estimator of the bias term $Cn^{-\kappa}$. Combining results yields the following:

Theorem 4.3. *Under the Assumptions of Lemma 4.1, for $\kappa \geq 2/5$ for the VRS and CRS cases or $\kappa \geq 1/3$ for the FDH case,*

$$\sqrt{n}\left(\widehat{\mu}_n - (2^\kappa - 1)^{-1}\left(\widehat{\mu}_{n/2}^{*} - \widehat{\mu}_n\right) - \mu_\theta + R_{n,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, \sigma_\theta^2). \tag{4.21}$$

as $n \to \infty$.

It is important to note that Theorem 4.3 is not valid for κ smaller that the bounds given in the theorem. This is due to the fact that for a particular definition of $R_{n,\kappa}$ (i.e., in either the VRS/CRS or FDH cases), values of κ smaller than the boundary value cause the remainder term, multiplied by $\sqrt{n}$, to diverge toward infinity. Interestingly, the normal approximation in Theorem 4.3 can be used with either the VRS-DEA or CRS-DEA estimators under the assumption of CRS if and only if $p + q \leq 5$; with the DEA-VRS estimator under convexity (but not CRS) if and only if $p + q \leq 4$; and with the FDH estimator assuming only free disposability (but not necessarily convexity nor CRS) if and only if $p+q \leq 3$. For these cases, an asymptotically correct $(1 - \alpha)$ confidence interval for μ_θ is given by

$$\left[\widehat{\mu}_n - (2^\kappa - 1)^{-1} \left(\widehat{\mu}^*_{n/2} - \widehat{\mu}_n\right) \pm z_{1-\alpha/2}\widehat{\sigma}_{\theta,n}/\sqrt{n}\right], \tag{4.22}$$

where $z_{1-\alpha/2}$ is the corresponding quantile of the standard normal distribution. The expression $\widehat{\mu}_n - (2^\kappa - 1)^{-1} \left(\widehat{\mu}^*_{n/2} - \widehat{\mu}_n\right)$ appearing in (4.22) can be viewed as a generalized jackknife statistic (e.g., see Gray and Schucany, 1972, Definition 2.1).

In cases where κ is smaller than the bounds given in Theorem 4.3, the idea of estimating μ_θ by a sample mean of n_κ efficiency estimates as above in Theorem 4.2 can be used with the bias correction introduced in this section. This leads to the following result.

Theorem 4.4. *Under the assumptions of Lemma 4.1, as $n \to \infty$,*

$$n^\kappa \left(\widehat{\mu}_{n_\kappa} - (2^\kappa - 1)^{-1} \left(\widehat{\mu}^*_{n/2} - \widehat{\mu}_n\right) - \mu_\theta + R_{n,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2_\theta). \tag{4.23}$$

Proof. Since in all the cases $R_{n,\kappa} = o(n^{-\kappa})$, it is clear that $n^\kappa R_{n,\kappa} = o(1)$. Hence the remainder term can be neglected, yielding the result. ■

Theorem 4.4 allows construction of consistent confidence intervals for μ_θ by replacing the unknown σ^2_θ by its consistent estimator $\widehat{\sigma}^2_{\theta,n}$. An asymptotically correct $1 - \alpha$ confidence interval for μ_θ is given by

$$\left[\widehat{\mu}_{n_\kappa} - (2^\kappa - 1)^{-1} \left(\widehat{\mu}^*_{n/2} - \widehat{\mu}_n\right) \pm z_{1-\alpha/2}\widehat{\sigma}_{\theta,n}/n^\kappa\right], \tag{4.24}$$

where $z_{1-\alpha/2}$ is the corresponding quantile of the standard normal distribution. Here, the normal approximation can be used directly; bootstrap methods are not necessary.

Note that when $\kappa < 1/2$, the center of the confidence interval in (4.24) is determined by a random choice of $n_\kappa = n^{2\kappa} < n$ elements $\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n)$. This may be seen as arbitrary, but any confidence interval for μ_θ may be seen arbitrary in practice since asymmetric confidence intervals can be constructed by using different quantiles to establish the endpoints. The main point, however, is always to achieve a high level of coverage without making the confidence interval too wide to be informative.

Again for $\kappa < 1/2$, the arbitrariness of choosing a particular subsample of size n_κ in (4.24) can be eliminated by averaging the center of the interval in (4.24) over all possible draws (without replacement) of subsamples of size n_κ. Of course, this yields an interval centered on $\widehat{\mu}_n$, i.e.,

$$\left[\widehat{\mu}_n - (2^\kappa - 1)^{-1}\left(\widehat{\mu}^*_{n/2} - \widehat{\mu}_n\right) \pm z_{1-\alpha/2}\widehat{\sigma}_{\theta,n}/n^\kappa\right]. \tag{4.25}$$

The only difference between the intervals (4.24) and (4.25) is the centering value. Both intervals are equally informative, because they possess exactly the same length, $(2z_{1-\alpha/2}\widehat{\sigma}_{\theta,n}/n^\kappa)$. The interval (4.25) should be more accurate (i.e., should have higher coverage) because $\widehat{\mu}_n$ is a better estimator of μ_θ (i.e., has less mean-square error) than $\widehat{\mu}_{n_\kappa}$. If $\kappa < 1/2$, then $n_\kappa < n$, and hence the interval in (4.25) contains the true value μ_θ with probability greater than $1-\alpha$, since by the results above, it is clear that the coverage of the interval in (4.25) converges to 1 as $n \to \infty$. This is confirmed by the Monte Carlo evidence presented below in Section 7.

In cases with sufficiently small dimensions, either Theorem 4.3 or 4.4 can be used to provide different asymptotically valid confidence intervals for μ_θ. For the VRS-DEA and CRS-DEA estimators, this is possible whenever $\kappa = 2/5$ and so $n_\kappa < n$. The interval (4.22) uses the scaling $\sqrt{n}$ and neglects, in Theorem 4.3, a term $\sqrt{n}R_{n,\kappa} = O(n^{-1/10})$, whereas the interval (4.24) uses the scaling n^κ, neglecting in Theorem 4.4 a term $n^\kappa R_{n,\kappa} = O(n^{-1/5})$. We thus may expect a better approximation by using the interval (4.24). The same is true for the FDH case when $\kappa = 1/3$, where the interval (4.22) neglects terms of order $O(n^{-1/6})$ whereas the error when using (4.24) is only of order $O(n^{-1/3})$. These remarks will be confirmed in some of our Monte-Carlo experiments.

5 Extension to Two-Stage Approaches

Two-stage estimation procedures where technical efficiencies are estimated in a first stage and then regressed in a second stage on some environmental variables $Z \in \mathbb{R}^r$ are very popular in the efficiency literature. Simar and Wilson (2007, 2011b) observed that hundreds of papers have used this approach for explaining inefficiency in terms of environmental variables, and propose a well-defined, coherent statistical model in which the second stage regression is meaningful. In particular, this involves a separability condition requiring that the shape of the attainable set Ψ is not affected by the variables in Z. Any effect of these variables on the production process is only through the distribution of the efficiencies inside Ψ, but the Z-variables do not affect the support of Ψ itself. Banker and Natarajan (2008) propose a different statistical model where the second-stage regression is meaningful, but the model is rather restrictive and based on unrealistic assumptions on the production process (see Simar and Wilson, 2011b for details).

Even if a statistical model is defined in which the second stage regression is potentially meaningful, an additional difficulty arises in the second stage regression from the fact that, as in the problem described in Section 4, the true unobserved Farrell measures of efficiencies are replaced by their DEA estimators on the left-hand side of the second-stage regression. The analysis from Section 4 can be easily extended to the case of second-stage regressions. For ease of exposition, the discussion below is presented in terms of a simple linear model where the effects of covariates in Z on firms' efficiencies can be estimated by ordinary least squares (OLS).[5] The main part of the message coming from the following analysis is analogous to the message in Theorem 4.1; i.e., under appropriate, mild regularity conditions, second-stage regressions yield consistent estimators of the given model, and inference by appropriate bootstrap methods is possible, but at a much lower rate than $\sqrt{n}$ as the number of dimensions, $p + q$, increases. This was the message in Simar and Wilson (2007, 2011b), but the results obtained in this paper give additional insight into the problem.

Consider a simple model where, in addition to the assumptions of Lemma 4.1, we assume the following:

Assumption 5.1. *(i) The environmental factors Z do not influence the shape of the attainable set Ψ (this is the "separability" condition described by Simar and Wilson, 2007); and (ii) the variables Z influence on the production process through the following simple mechanism:*

$$\theta(X, Y) = \boldsymbol{Z}\boldsymbol{\beta} + \varepsilon, \tag{5.1}$$

where $\boldsymbol{Z}$ denotes the $n \times (r + 1)$ matrix of observed values of Z, $\boldsymbol{\beta}$ is a vector of parameters, $E(\varepsilon \mid Z) = 0$, $VAR(\varepsilon \mid Z) = \sigma^2$ and $E(\varepsilon^4 \mid Z) \leq D$ for some $\sigma^2, D > 0$ independent of Z.

[5] OLS is used here only as an illustration; in practice, one would use truncated regression or other methods as appropriate. The arguments given below are relevant to other second-stage specifications such as the truncated normal regression described in Simar and Wilson (2007) or the nonparametric truncated regression discussed by Park et al. (2008).

Of course, the dependent variable in (5.1) is bounded, and so a truncated regression specification such as the one in Simar and Wilson (2007) would be more appropriate. Although one would use truncated regression in practice, the simple linear specification in (5.1) serves to illustrate the issues without the additional complication of nonlinear estimation required for a truncated regression specification, and allows the parameter vector $\boldsymbol{\beta}$ (which is the object of interest) to be estimated by OLS. For purposes of the discussion here, assume that the first element of Z is 1 to represent an intercept.

If the true, but unknown $\theta(X_i, Y_i)$ were available, the usual OLS estimator of $\boldsymbol{\beta}$ would be given by

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\theta}, \tag{5.2}$$

where $\boldsymbol{\theta}$ is the $n \times 1$ vector of elements $\theta(X_i, Y_i)$, $i = 1, \ldots, n$. Under mild regularity conditions on $\boldsymbol{Z}$,

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2\boldsymbol{Q}). \tag{5.3}$$

However, the result in (5.3) is not helpful because $\boldsymbol{\theta}$, which is needed for $\widehat{\boldsymbol{\beta}}$, is not observable. The only possibility, and what is done in practice, is to replace the unobserved $\boldsymbol{\theta}$ with the vector $\widehat{\boldsymbol{\theta}}$ of efficiency estimators; each element of $\widehat{\boldsymbol{\theta}}$ is an estimator of the corresponding element of $\boldsymbol{\theta}$. After substitution, (5.3) becomes

$$\widehat{\widehat{\boldsymbol{\beta}}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\widehat{\boldsymbol{\theta}}. \tag{5.4}$$

This can be decomposed by writing

$$\widehat{\widehat{\boldsymbol{\beta}}} - \boldsymbol{\beta} = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} + (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right). \tag{5.5}$$

The asymptotic behavior of the latter term depends on the chosen estimator and is given below in Theorem 5.1.

Before turning to the theorem, an additional assumption is needed.

Assumption 5.2. *(i) (X_i, Y_i, Z_i), $i = 1, \ldots, n$, are iid random variables, and the marginal density f of (X_i, Y_i) has support $\mathcal{D}$ (which is assumed to satisfy the assumptions imposed in Sections 2 and 3).*

(ii) *The variables Z_{is}, $s = 1, \dots, r$, have finite fourth moments, and the conditional distributions of (X_i, Y_i) given $Z_i = Z$ possess densities f_Z with support $\mathcal{D}_Z \subset \mathcal{D}$. Moreover, f_Z changes continuously with Z, and there exists a constant $\delta_0 < \infty$ such that $f_Z(x, y) \leq \delta_0$ for all $(x, y) \in \mathcal{D}$ and all possible values of Z.*

(iii) *There exists a positive definite matrix $\boldsymbol{Q}$ such that $n^{-1}(\boldsymbol{Z}'\boldsymbol{Z}) = \boldsymbol{Q} + O_P(n^{-1/2})$.*

Note that Assumption 5.2 (i) does not exclude that, for example, the first element Z_{i1} of Z_i is an intercept and thus identical to 1. In this case Z_{i1}, $i = 1, \dots, n$, can be interpreted as i.i.d. random variables with $P(Z_{i1} = 1) = 1$.

Theorem 5.1. *Under the conditions of Lemma 4.1 together with Assumptions 5.1 and 5.2 we obtain:*

(i) *There exists a vector $\mathbf{C} \in \mathbb{R}^r$ as well as a sequence of vectors $\mathbf{R}_{n,\kappa} \in \mathbb{R}^r$ satisfying $\mathbf{R}_{n,\kappa} = O(n^{-\kappa^*}(\log n)^{1+\kappa^*})$ such that*

$$n^{-1}\boldsymbol{Z}'\widehat{\boldsymbol{\theta}} = n^{-1}\boldsymbol{Z}'\boldsymbol{\theta} + \mathbf{C}n^{-\kappa} + \mathbf{R}_{n,\kappa} + o_p(n^{-1/2}) \tag{5.6}$$

and

$$\widehat{\widehat{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} + o_p(n^{-1/2}) \quad \text{with } \widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + Q^{-1}\mathbf{C}n^{-\kappa} + Q^{-1}\mathbf{R}_{n,\kappa}; \tag{5.7}$$

(ii) $\sqrt{n}\left(\widehat{\widehat{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2\boldsymbol{Q})$*; and*

(iii) $\widehat{\sigma}_n^2 = n^{-1}\|\boldsymbol{Z}\widehat{\widehat{\boldsymbol{\beta}}} - \widehat{\boldsymbol{\theta}}\|_2^2 \xrightarrow{p} \sigma^2$

where $\kappa = 2/(p+q+2)$ and $\kappa^* = 3\kappa/2$, $\kappa = 2/(p+q)$ and $\kappa^* = 3\kappa/2$, or $\kappa = 1/(p+q)$ and $\kappa^* = 2\kappa$ for the VRS-DEA, CRS-DEA, or FDH estimators, respectively.

Proof. In the proof we will concentrate on the VRS-DEA estimator. Following the theorems of Section 3, the proofs for the CRS-DEA as well as the FDH estimator can be derived analogously.

Based on decomposition (5.5) we have to analyze the stochastic behavior of the term $n^{-1}\boldsymbol{Z}'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n} Z_i(\widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i))$. Along the lines of the proof of Theorem 3.1 let $\mathcal{X}_{n,-i}$ denote the sample of size $n-1$ obtained by eliminating the i-th observation (X_i, Y_i), and set $\mathbf{b}(x, y) = E(\widehat{\theta}_{\text{VRS}}(x, y \mid \mathcal{X}_{n,-i}))$ for $(x, y) \in \mathcal{D}$. By Assumption

5.2 (i) the random variables (X_i, Y_i, Z_i) are independent of $\mathcal{X}_{n,-i}$. By Assumption 5.2 (ii) and (6.18), arguments similar to (6.19)–(6.22) in the proof of Theorem 3.1 can now be applied with respect to the conditional distributions of (X_i, Y_i) given Z_i. In particular, when replacing f by the conditional densities f_Z, then for all Z the probabilities that (X_i, Y_i) fall into a boundary region are of the order of magnitude required in (6.19)–(6.22). Furthermore, it can be shown that $Z_i(\widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i))$ and $Z_j(\widehat{\theta}_{\text{VRS}}(X_j, Y_j \mid \mathcal{X}_n) - \theta(X_j, Y_j))$ are asymptotically uncorrelated if $\mathcal{X}_n(X_i, Y_i; 1, \nu_n^*) \cap \mathcal{X}_n(X_j, Y_j; 1, \nu_n^*) = \emptyset$. When analyzing expectations this leads to

$$E(\frac{1}{n}\sum_{i=1}^{n} Z_i(\widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i))) = E(\frac{1}{n}\sum_{i=1}^{n} Z_i(\widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_{n,-i}) - \theta(X_i, Y_i))) + o(n^{-\kappa^*})$$
$$= E\left(Z_i \; E(\mathbf{b}(X_i, Y_i) - \theta(X_i, Y_i) \mid \; Z_i)\right) + o(n^{-\kappa^*}) = \mathbf{C}n^{-\kappa} + O(n^{-\kappa^*}(\log n)^{1+\kappa^*})$$

for some vector $\mathbf{C} \in \mathbb{R}^r$, while the assertions on variances and covariances in (3.2) and (3.3) remain valid when replacing $(\widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i))$, $(\widehat{\theta}_{\text{VRS}}(X_j, Y_j \mid \mathcal{X}_n) - \theta(X_j, Y_j))$ by $Z_i(\widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i))$, $Z_j(\widehat{\theta}_{\text{VRS}}(X_j, Y_j \mid \mathcal{X}_n) - \theta(X_j, Y_j))$. Therefore,

$$n^{-1}\sum_{i=1}^{n} Z_i(\widehat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i))) - \mathbf{C}n^{-\kappa} = O(n^{-\kappa^*}(\log n)^{1+\kappa^*}) + o_p(n^{-1/2}).$$

Together with (5.5) and Assumption 5.2 (iii), these arguments imply (5.6) and (5.7). Assertion (ii) of Theorem (5.1) is then a straightforward consequence of Lyapunov's central limit theorem, while (iii) follows from $n^{-1}\|\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\|_2^2 = o_p(1)$ as well as $n^{-1}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2 = o_p(1)$.
■

Note that in the regression context, Theorem 5.1 constitutes a direct generalization of Theorem 4.1 in Section 4. Again, result (ii) in Theorem 5.1 demonstrates clearly why the usual central limit theorem results cannot be used for making inference unless $\kappa > 1/2$, due to the inherent bias in $\widetilde{\boldsymbol{\beta}}$ which is of order $O(n^{-\kappa})$. In order to construct confidence intervals or test of significance for $\boldsymbol{\beta}$ we can thus rely on techniques similar to those developed in Section 4 for inference about sample means.

A first step consists in defining an estimator of the leading bias term. In the same way as in Section 4 divide $\mathcal{X}_n$ into the sets $\mathcal{X}_{n/2}^{(1)}$ and $\mathcal{X}_{n/2}^{(2)}$ of the first and second $n/2$ observations in $\mathcal{X}_n$. Let $\widehat{\boldsymbol{\theta}}_{n/2}^{(1)}$ and $\widehat{\boldsymbol{\theta}}_{n/2}^{(2)}$ denote the efficiency estimators obtained using only the observations in

$\mathcal{X}_{n/2}^{(1)}$ and $\mathcal{X}_{n/2}^{(2)}$, respectively, and let $\boldsymbol{Z}_{n/2}^{(j)}$, $j = 1, 2$ be the corresponding matrices of Z-values. Then define

$$\widehat{\widehat{\boldsymbol{\beta}}}^*_{n/2} := (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\left((\boldsymbol{Z}_{n/2}^{(1)})'\widehat{\boldsymbol{\theta}}_{n/2}^{(1)} + (\boldsymbol{Z}_{n/2}^{(2)})'\widehat{\boldsymbol{\theta}}_{n/2}^{(2)}\right).$$

By (5.6) and the definition of $\widehat{\boldsymbol{\beta}}$ we have

$$\widehat{\widehat{\boldsymbol{\beta}}}^*_{n/2} = \widehat{\boldsymbol{\beta}} + 2^{\kappa}Q^{-1}\mathbf{C}n^{-\kappa} + O(\mathbf{R}_{n,\kappa}) + o_p(n^{-1/2}),$$

and therefore

$$(2^{\kappa} - 1)^{-1}(\widehat{\widehat{\boldsymbol{\beta}}}^*_{n/2} - \widehat{\widehat{\boldsymbol{\beta}}}) = Q^{-1}\mathbf{C}n^{-\kappa} + O(\mathbf{R}_{n,\kappa}) + o_p(n^{-1/2}) \tag{5.8}$$

provides an approximation of the first order error term. If $\kappa \geq 2/5$ then $\mathbf{R}_{n,\kappa} = o_p(n^{-1/2})$, and the above arguments lead to the following theorem:

Theorem 5.2. *Under the Assumptions of Theorem 5.1, for $\kappa \geq 2/5$ for the VRS and CRS cases or $\kappa \geq 1/3$ for the FDH case,*

$$\sqrt{n}\left(\widehat{\widehat{\boldsymbol{\beta}}} - (2^{\kappa} - 1)^{-1}\left(\widehat{\widehat{\boldsymbol{\beta}}}^*_{n/2} - \widehat{\widehat{\boldsymbol{\beta}}}\right) - \boldsymbol{\beta}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2\boldsymbol{Q}), \tag{5.9}$$

as $n \to \infty$.

The diagonal elements of $\boldsymbol{Q}$ can be consistently estimated by the diagonal elements $\widehat{q}_{ss} := ((n^{-1}\boldsymbol{Z}'\boldsymbol{Z})^{-1})_{ss}$ of the matrix $(n^{-1}\boldsymbol{Z}'\boldsymbol{Z})^{-1}$, $s = 1, \ldots, r$. Theorem 5.1 (c) defines a consistent estimator $\widehat{\sigma}_n^2$ of the error variance σ^2. Hence, under the conditions of Theorem 5.2 an asymptotically valid $(1 - \alpha)$ confidence interval for the s-th element $\boldsymbol{\beta}_s$ of $\boldsymbol{\beta}$ is given by

$$\left[\widehat{\widehat{\boldsymbol{\beta}}}_s - (2^{\kappa} - 1)^{-1}\left(\widehat{\widehat{\boldsymbol{\beta}}}^*_{n/2;s} - \widehat{\widehat{\boldsymbol{\beta}}}_s\right) \pm z_{1-\alpha/2}\widehat{\sigma}_n\widehat{q}_{ss}/\sqrt{n}\right], \tag{5.10}$$

where $z_{1-\alpha/2}$ is the corresponding quantile of the standard normal distribution.

If $p + q$ is large, then the resulting value of κ will be smaller than required by Theorem 5.2. Similar to the approach of Section 4 one may then rely on a reduction of the sample size.

Define the OLS estimator of $\boldsymbol{\beta}$ based on a random subset of size $n_{\kappa} = [n^{2\kappa}]$ of the data (drawn without replacement). For simplicity of notation, assume that the order of the data is random, so that $\boldsymbol{Z}_{n_{\kappa}}$ and $\widehat{\boldsymbol{\theta}}_{n_{\kappa}}$ are the first n_{κ} row entries of $\boldsymbol{Z}$ and of $\widehat{\boldsymbol{\theta}}$, and

$$\widehat{\widehat{\boldsymbol{\beta}}}_{n_{\kappa}} = (\boldsymbol{Z}'_{n_{\kappa}}\boldsymbol{Z}_{n_{\kappa}})^{-1}\boldsymbol{Z}'_{n_{\kappa}}\widehat{\boldsymbol{\theta}}_{n_{\kappa}}. \tag{5.11}$$

Theorem 5.3. *Under the Assumptions of Theorem 5.1, as $n \to \infty$,*

$$n^{\kappa}\left(\widehat{\widehat{\boldsymbol{\beta}}}_{n_\kappa} - (2^{\kappa}-1)^{-1}\left(\widehat{\widehat{\boldsymbol{\beta}}}^*_{n/2} - \widehat{\widehat{\boldsymbol{\beta}}}\right) - \boldsymbol{\beta}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2 \boldsymbol{Q}). \tag{5.12}$$

Proof. Similar to (5.3) we have $n^{\kappa}\left(\widehat{\boldsymbol{\beta}}_{n_\kappa} - \boldsymbol{\beta}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2 \boldsymbol{Q})$ with $\widehat{\boldsymbol{\beta}}_{n_\kappa} := (\boldsymbol{Z}'_{n_\kappa}\boldsymbol{Z}_{n_\kappa})^{-1}\boldsymbol{Z}'_{n_\kappa}\boldsymbol{\theta}_{n_\kappa}$, and it is easily verified that (5.7) generalizes to $\widehat{\widehat{\boldsymbol{\beta}}}_{n_\kappa} - \boldsymbol{\beta} - Q^{-1}\mathbf{C}n^{-\kappa} - Q^{-1}\mathbf{R}_{n,\kappa} = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} + o_p(n^{-\kappa})$. Since $\mathbf{R}_{n,\kappa} = o_p(n^{-\kappa})$ and by (5.8) $(2^{\kappa}-1)^{-1}(\widehat{\widehat{\boldsymbol{\beta}}}^*_{n/2} - \widehat{\widehat{\boldsymbol{\beta}}}) = Q^{-1}\mathbf{C}n^{-\kappa} + o_p(n^{-\kappa})$, the assertion of the theorem is an immediate consequence. ■

Theorem 5.3 implies that for $\kappa \leq 1/2$, an asymptotically correct $1-\alpha$ confidence interval for $\boldsymbol{\beta}_s$, $s = 1, \ldots, r$ is given by

$$\left[\widehat{\widehat{\boldsymbol{\beta}}}_{n_\kappa;s} - (2^{\kappa}-1)^{-1}\left(\widehat{\widehat{\boldsymbol{\beta}}}^*_{n/2;s} - \widehat{\widehat{\boldsymbol{\beta}}}_s\right) \pm z_{1-\alpha/2}\widehat{\sigma}_n\widehat{q}_{ss}/n^{\kappa}\right]. \tag{5.13}$$

Similar to the arguments in Section 4, for $\kappa < 1/2$, the arbitrariness of choosing a particular subsample of size n_κ in (4.24) can be eliminated by replacing $\widehat{\widehat{\boldsymbol{\beta}}}_{n_\kappa;s}$ by the original estimator $\widehat{\widehat{\boldsymbol{\beta}}}_s$, $s = 1, \ldots, r$. Since $\widehat{\widehat{\boldsymbol{\beta}}}_s$ possesses a faster rate of convergence than $\widehat{\widehat{\boldsymbol{\beta}}}_{n_\kappa;s}$, one can conclude that for large n the interval

$$\left[\widehat{\widehat{\boldsymbol{\beta}}}_s - (2^{\kappa}-1)^{-1}\left(\widehat{\widehat{\boldsymbol{\beta}}}^*_{n/2;s} - \widehat{\widehat{\boldsymbol{\beta}}}_s\right) \pm z_{1-\alpha/2}\widehat{\sigma}_n\widehat{q}_{ss}/n^{\kappa}\right] \tag{5.14}$$

contains the true values $\boldsymbol{\beta}_s$ with probability **greater** than $1-\alpha$.

Note that due to "nonlinearities" introduced by (random) differences of the matrices $(\boldsymbol{Z}'_{n_\kappa}\boldsymbol{Z}_{n_\kappa})^{-1}$, $\widehat{\widehat{\boldsymbol{\beta}}}_s$ will typically not be equal to the average of $\widehat{\widehat{\boldsymbol{\beta}}}_{n_\kappa;s}$ over all possible draws (without replacement) of subsamples of size n_κ. This effect can be eliminated by using a modified estimator $\widehat{\widehat{\boldsymbol{\beta}}}^*_{n_\kappa} := (n^{-1}\boldsymbol{Z}'\boldsymbol{Z})^{-1}n_\kappa^{-1}\boldsymbol{Z}'_{n_\kappa}\boldsymbol{\theta}_{n_\kappa}$. Note also that Theorem 5.3 as well as (5.13) remain true if $\widehat{\widehat{\boldsymbol{\beta}}}_{n_\kappa;s}$ is replaced by $\widehat{\widehat{\boldsymbol{\beta}}}^*_{n_\kappa;s}$. However, in any case the interval given by (5.14) possesses a superior coverage probability.

6 Mathematical Details

For any efficiency estimator $\widehat{\theta}(x, y)$ considered in this section we will set $\widehat{\theta}(x, y) := 1$ whenever the set of all possible values θ satisfying the defining inequalities is the empty set.

6.1 Proof of Theorem 3.1

Let $\mathcal{H}_n$ and $\widetilde{\mathcal{H}}_n$ denote the convex hulls of the sample $\mathcal{X}_n := \{(X_i, Y_i) \mid i = 1, \ldots, n\}$ and of the augmented sample $\widetilde{\mathcal{X}}_n := \{(X_i, Y_i) \mid i = 1, \ldots, n\} \cup \{(X_i, 0) \mid i = 1, \ldots, n\}$, respectively. The structure of the VRS-estimator implies that for all $(x, y) \in \mathcal{D}$ with $(ax, y) \in \widetilde{\mathcal{H}}_n$ for some $a > 0$, it is the case that $\left(\widehat{\theta}_{\mathrm{VRS}}(x, y \mid \mathcal{X}_n)x, y\right) \in \widetilde{\mathcal{H}}_n^{\partial}$, and hence $\widehat{\theta}_{\mathrm{VRS}}(x, y \mid \mathcal{X}_n) = \widehat{\theta}_{\mathrm{CH}}(x, y \mid \widetilde{\mathcal{H}}_n)$. If in addition y is sufficiently larger than zero, then $\left(\widehat{\theta}_{\mathrm{VRS}}(x, y \mid \mathcal{X}_n)x, y\right) \in \widetilde{\mathcal{H}}_n^{\partial}$ as well as $\left(\widehat{\theta}_{\mathrm{VRS}}(x, y \mid \mathcal{X}_n)x, y\right) \in \mathcal{H}_n^{\partial}$, which yields $\widehat{\theta}_{\mathrm{VRS}}(x, y \mid \mathcal{X}_n) = \widehat{\theta}_{\mathrm{CH}}(x, y \mid \widetilde{\mathcal{H}}_n) = \widehat{\theta}_{\mathrm{CH}}(x, y \mid \mathcal{H}_n)$. By the assumptions of the theorem it can be seen that for any (x, y) in the interior of $\mathcal{D}$, the probability $\Pr\left(\widehat{\theta}_{\mathrm{VRS}}(x, y \mid \mathcal{X}_n) = \widehat{\theta}_{\mathrm{CH}}(x, y \mid \widetilde{\mathcal{H}}_n) = \widehat{\theta}_{\mathrm{CH}}(x, y \mid \mathcal{H}_n)\right)$ converges to 1 as $n \to \infty$. Finally, note that $\widehat{\theta}_{\mathrm{CH}}(x, y \mid \mathcal{H}_n)$ can be determined by (2.7) when replacing all inequalities by corresponding equalities.

The arguments developed in Kneip et al. (2008) now provide a basis for the proof of the theorem. To begin, consider an arbitrary point $(x, y) \in \mathcal{D}$. 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 $X_i = \gamma_i \frac{x}{||x||} + Z_i$ for some $Z_i \in \mathcal{V}(x)$ and $\gamma_i = \frac{x^T X_i}{||x||}$, where $|| \cdot ||$ denotes the Euclidean norm. Let $\Psi^*(x)$ denote the set of all $(z, y) \in \mathcal{V}(x) \times \mathbb{R}^q$ with $(\gamma \frac{x}{||x||} + z, y) \in \mathcal{D}$ for some $\gamma > 0$. Note that the point of interest $(x, y) \in \Psi$ has coordinates $(0, y)$ in $\Psi^*(x)$.

The maintained assumptions imply that for any $(z, y) \in \Psi^*(x)$, there exists $\gamma > 0$ such that $(\gamma \frac{x}{||x||} + z, y) \in \Psi$. The efficient boundary of Ψ can therefore be described by the function $g_x(z, y) := \inf \left\{\gamma \mid \left(\gamma \frac{x}{||x||} + z, y\right) \in \Psi\right\}$ defined for any $(z, y) \in \Psi^*(x)$. Furthermore, with only a small abuse of notation, one may extend the definition of g_x to all (v, y) with $\left(v - \frac{x^T v}{||x||^2} x, y\right) \in \Psi^*(x)$ by taking $g_x(v, y) = g_x\left(v - \frac{x^T v}{||x||^2} x, y\right)$.

Properties of g_x are discussed in Kneip et al. (2008). In particular, under the assumptions of the theorem, g_x is a three times continuously differentiable, strictly convex function, and there exists a constant $C_1 > 0$ such that $w^T g_x''(0, y) w \geq C_1$ for all $w \in \mathcal{V}(x) \times \mathbb{R}^q$ with $||w|| = 1$ and all $x \in \mathbb{R}^q$ with $(x, y) \in \mathcal{D}$. Moreover, $g_x''(0, y)$ changes continuously in x.

The decomposition described above establishes a new coordinate system in which each observation (X_i, Y_i) can be equivalently represented by the corresponding vector (θ_i, Z_i, Y_i), where $\theta_i := \theta(X_i, Y_i)$. The point (x, y) of interest has coordinates $(\theta(x, y), 0, y)$ in this new system.

Now consider a point $(x, y) \in \mathcal{D}$ with $\big(\widehat{\theta}_{\rm VRS}(x, y \mid \mathcal{X}_n)x, y\big) \in \mathcal{H}_n^{\partial}$ and thus $\widehat{\theta}_{\rm VRS}(x, y \mid \mathcal{X}_n) = \widehat{\theta}_{\rm CH}(x, y \mid \mathcal{H}_n)$. Following Kneip et al. (2008, proof of Lemma 2, p.1687), the VRS-DEA estimator of $\theta(x, y)$ can be rewritten as

$$\begin{aligned}\widehat{\theta}_{\rm VRS}(x, y \mid \mathcal{X}_n) &= \min\left\{\sum_{i=1}^{n} \omega_i \frac{g_x(\theta_i Z_i, Y_i)}{\|x\|\theta_i} \mid \boldsymbol{i}_n^T \boldsymbol{\omega} = 1,\ \boldsymbol{Z}\boldsymbol{\omega} = 0,\ \boldsymbol{Y}\boldsymbol{\omega} = y, \boldsymbol{\omega} \in \mathbb{R}_+^n\right\} \\ &= \theta(x, y) \times \min\Big\{\sum_{i=1}^{n} \omega_i \frac{g_x(\theta_i Z_i, Y_i)}{g_x(0, y)\theta_i} \mid \boldsymbol{i}_n^T \boldsymbol{\omega} = 1,\ \boldsymbol{Z}\boldsymbol{\omega} = 0, \\ &\qquad\qquad (\boldsymbol{Y} - y\boldsymbol{i}_n^T)\boldsymbol{\omega} = 0, \boldsymbol{\omega} \in \mathbb{R}_+^n\Big\} \end{aligned} \tag{6.1}$$

where $\boldsymbol{i}_n$ and $\boldsymbol{Y}$ are defined as in Section 2, $\theta_i = \theta(X_i, Y_i)$ and $Z_i = X_i - \dfrac{\boldsymbol{x}^T X_i}{||\boldsymbol{x}||^2}\boldsymbol{x}$ is a $(p \times 1)$ vector and $\boldsymbol{Z} = (Z_1, \ \ldots, \ Z_n)$ is a $(p \times n)$ matrix.

Let $z_x^{(1)}, \ \ldots, \ z_x^{(p-1)}$ be an orthonormal basis of $\mathcal{V}(\boldsymbol{x})$. Clearly, the $z_x^{(j)}$ can be chosen as continuous functions of x, and every vector $Z \in \mathcal{V}(\boldsymbol{x})$ can be expressed in the form $Z = \mathbf{Z}_x\zeta$, where $\mathbf{Z}_x$ is the $p \times (p-1)$ matrix with columns $z_x^{(j)}$, $j = 1, \ldots, p-1$, and $\zeta \in \mathbb{R}^{p-1}$. Hence, $Z = \mathbf{Z}_x\zeta_i$. Since $\theta_i = \theta(X_i, Y_i)$ and $Z_i = X_i - \frac{x^T X_i}{||x||^2}x$ are smooth functions of (X_i, Y_i), the joint density f of (X_i, Y_i) translates into a density $\overline{f}_x$ on $(0, 1] \times \mathbb{R}^{p-1} \times \mathbb{R}_+^q$ of (θ_i, ζ_i, Y_i). Let $\overline{\mathcal{D}}$ denote the support of this density. Since f is continuously differentiable, $\overline{f}_x(\theta, \zeta, y)$ is also continuously differentiable on $(\theta, \zeta, y) \in \overline{\mathcal{D}}$. Furthermore, compactness of $\mathcal{D}^*$, as well as $f(\theta(x, y)x, y) > 0$ for all $(x, y) \in \mathcal{D}$, imply that there exists a constant $c_{\inf} > 0$ such that

$$\overline{f}_x(1, \zeta, y) \geq c_{\inf} \tag{6.2}$$

whenever $(\mathbf{Z}_x\zeta, y) \in \Psi^*(x)$ and $x \in \mathbb{R}^q$ with $(x, y) \in \mathcal{D}$.

Next, the localization argument developed in Kneip et al. (2008) is used to show that $\widehat{\theta}_{\rm VRS}(x, y \mid \mathcal{X}_n)$ is asymptotically determined by local information, i.e., only by observations falling into a small neighborhood of (x, y). For any $h_1, h_2 > 0$, define the set

$$\begin{aligned} C(x, y; h_1, h_2) =& \Big\{(\theta, \widetilde{z}, \widetilde{y}) \in (0, 1] \times \Psi^*(x) \mid 1 - \theta \leq h_1, \\ & \widetilde{z} = \sum_j \zeta_j z_x^{(j)}, |\zeta_j| \leq h_2 \ \forall\ j = 1, \ \ldots, \ p-1, \ |y_r - \widetilde{y}_r| \leq h_2 \ \forall\ r = 1, \ \ldots, \ q\Big\}, \end{aligned} \tag{6.3}$$

and let $\mathcal{X}_n(x, y; h_1, h_2) := \{(X_i, Y_i) \in \mathcal{X}_n \mid (\theta_i, Z_i, Y_i) \in C(x, y; h_1, h_2)\}$.

In the following it will be necessary to distinguish between points (x, y) lying in the interior and on the observable boundary of $\mathcal{D}$. For $(x, y) \in \mathcal{D}$, let

$$\Psi^{*\partial}(x) = \left\{ (z, y) \in \mathcal{V}(x) \times \mathbb{R}^q \mid g_x(z, y)\frac{x}{||x||} + z, y) \in \mathcal{D}^* \right\} \tag{6.4}$$

denote the boundary of $\Psi^*(x)$. Define the observable boundary as

$$\mathcal{W}(h) := \left\{ (x, y) \in \mathcal{D} \mid \min \left\{ \min_{j=1,\ldots,p-1} |\zeta_j|, \min_{r=1,\ldots,q} |y_r - \widetilde{y}_r| \right\} \le h \right.$$
$$\left. \text{for some } (\widetilde{z}, \widetilde{y}) \in \Psi^{*\partial}(x) \text{ with } \widetilde{z} = \sum_j \zeta_j z_x^{(j)} \right\}. \tag{6.5}$$

If $p = 1$ and $q = 0$, then $\mathcal{W}(h) = \emptyset$; but for $p + q \ge 2$, compactness of $\mathcal{D}^*$ implies that for any $h > 0$ the observable boundary $\mathcal{W}(h)$ is nonempty. Now choose some constant $b \ge 4(p + q)$ and set $\nu_n^* := b(\frac{\log n}{c_{\inf} n})^{\frac{1}{p+q+1}}$ and $\nu_n := b(\frac{\log n}{n \overline{f}_x(1, 0, y)})^{\frac{1}{p+q+1}}$.

Case (i): We first consider the case where (x, y) is in the interior of $\mathcal{D}$ in the sense that $(x, y) \notin \mathcal{W}(\nu_n^*)$. In this case, it is clear that $C(x, y; \nu_n^2, \nu_n) \subset \mathcal{D}$ for all sufficiently large n.

Following Lemma 1 and the proof of Theorem 1, part (a) in Kneip et al. (2008), one can construct $k = 2(p+q-1)$ sets $B_j \subset C(x, y; \nu_n^2, \nu_n)$, $j = 1, \ldots, k$ centered at values (z_j, y_j) and surrounding $(0, y)$. Each set B_j possesses Lebesgue measure at least $\frac{2 \log n}{n \overline{f}_x(1,0,y)}$, and the probability that there exist at least k observations $(\theta_{i_1}, Z_{i_1}, Y_{i_1}), \ldots, (\theta_{i_1}, Z_{i_1}, Y_{i_1})$ with $(\theta_{i_j}, Z_{i_j}, Y_{i_j}) \in B_j$, $j = 1, \ldots, k$, is of order $1-n^{-2}$ as $n \to \infty$. On the other hand, if such a set of k observations exists, then for any other observation (θ_i, Z_i, Y_i) with $(\theta_i, Z_i, Y_i)) \notin C(x, y; \nu_n^2, \nu_n)$ and any vector $\boldsymbol{\omega} \in \mathbb{R}_+^n$ satisfying the constraints in (6.1), with $\boldsymbol{\omega}_i > 0$, there exists another vector $\boldsymbol{\omega}^* \in \mathbb{R}_+^n$ with $\boldsymbol{\omega}_i^* = 0$ and $\boldsymbol{\omega}_{i_j} \ge 0$, $j = 1, \ldots, k$ such that $\sum_{i=1}^n \boldsymbol{\omega}_i \frac{g_x(\theta_i Z_i, Y_i)}{g_x(0,y)\theta_i} > \sum_{i=1}^n \boldsymbol{\omega}_i^* \frac{g_x(\theta_i Z_i, Y_i)}{g_x(0,y)\theta_i}$. This implies that the minimum in (6.1) is achieved by assigning zero weight $\boldsymbol{\omega}_i = 0$ to each observation $(\theta_i, Z_i, Y_i) \notin C(x, y; \nu_n^2, \nu_n)$. The same type of argument then also allows exclusion of all additional points $(X_i, 0)$, $i = 1, \ldots, n$ of the augmented sample $\widetilde{\mathcal{X}}_n$. This then leads to $\widehat{\theta}_{\text{VRS}}(x, y \mid \mathcal{X}_n) = \widehat{\theta}_{\text{CH}}(x, y \mid \widetilde{\mathcal{X}}_n) = \widehat{\theta}_{\text{CH}}(x, y \mid \mathcal{H}_n(x, y; \nu_n^2, \nu_n))$, where $\mathcal{H}_n(x, y;\ \nu_n^2, \nu_n)$ denotes the convex hull of $\mathcal{X}_n(x, y;\ \nu_n^2, \nu_n)$. Therefore, there exists a $C_2 \in (0, \infty)$ such that for all sufficiently large n,

$$\left| 1 - \Pr\left(\widehat{\theta}_{\text{VRS}}(x, y \mid \mathcal{X}_n) = \widehat{\theta}_{\text{CH}}(x, y \mid \mathcal{H}_n(x, y; \nu_n^2, \nu_n)) \right) \right| \le C_2 n^{-2} \ \forall\ y \notin \mathcal{W}(\nu_n^*). \tag{6.6}$$

Now consider the sums in (6.1) with respect to an arbitrary number $k \leq \#\mathcal{X}_n(x, y; \nu_n^2, \nu_n)$ of observations with coordinates $(\theta_{i_j}, Z_{i_j}, Y_{i_j}) \in C(x, y; \nu_n^2, \nu_n)$. Define

$$\tau((\theta^*_{i_1}, Z_{i_1}, Y_{i_1}), \ \ldots, \ (\theta^*_{i_k}, Z_{i_k}, Y_{i_k}); \boldsymbol{\omega}) := \\ \sum_{j=1}^{k} \boldsymbol{\omega}_j \frac{1}{2g_x(0,y)} \big[Z_{i_j}^T g''_{x;zz}(0,y) Z_{i_j} + 2 Z_{i_j}^T g''_{x;zy}(0,y)(Y_{i_j} - y) + \\ (Y_{i_j} - y)^T g''_{x;yy}(0,y)(Y_{i_j} - y) \big] + \sum_{j=1}^{k} \boldsymbol{\omega}_j \theta^*_{i_j}. \tag{6.7}$$

Using the compactness of $\mathcal{D}^*$, $\sum_{j=1}^k \boldsymbol{\omega}_j = 1$, $\sum_{j=1}^k \boldsymbol{\omega}_j Z_{i_j} = 0$, $\sum_{j=1}^k \boldsymbol{\omega}_j (Y_{i_j} - y) = 0$, as well as the three-times differentiability of g_x, some straightforward Taylor expansions show that there exist constants $C_3 \in (0, \infty)$ and $C_4 \in (0, \infty)$ such that for $\theta^*_{i_j} := 1 - \theta_{i_j}$,

$$\left| \sum_{j=1}^{k} \boldsymbol{\omega}_j \frac{g_x(\theta_{i_j} Z_{i_j}, Y_{i_j})}{g_x(0,y)\theta_{i_j}} - 1 - \tau((\theta^*_{i_1}, Z_{i_1}, Y_{i_1}), \ \ldots, \ (\theta^*_{i_k}, Z_{i_k}, Y_{i_k})) \right| \leq C_3 n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}} \tag{6.8}$$

and

$$\tau((\theta^*_{i_1}, Z_{i_1}, Y_{i_1}), \ \ldots, \ (\theta^*_{i_k}, Z_{i_k}, Y_{i_k}); \boldsymbol{\omega}) \leq C_4 n^{-\frac{2}{p+q+1}} (\log n)^{\frac{2}{p+q+1}}. \tag{6.9}$$

Here, $g''_{x;zz}$, $g''_{x;zy}$ and $g''_{x;yy}$ denote the second partial derivatives of $g_x(z, y)$. From these results it follows that there exists a constant $C_5 \in (0, \infty)$ such that for $\alpha \in \{1, 2\}$,

$$E\left(\left| \widehat{\theta}_{\mathrm{VRS}}(x, y \mid \mathcal{X}_n) - \theta(x,y) \right|^{\alpha} \right) \leq C_5^{\alpha} n^{-\frac{2\alpha}{p+q+1}} (\log n)^{\frac{2\alpha}{p+q+1}}, \ \forall \ (x,y) \notin \mathcal{W}(\nu_n^*). \tag{6.10}$$

Proof of the theorem requires refinements of the inequality in (6.10). Recall that $\theta^*_{i_j} := 1 - \theta_{i_j}$. With $K_n := \#\mathcal{X}_n(x, y; \nu_n^2, \nu_n)$, let $T_{K_n}((\theta^*_{i_1}, Z_{i_1}, Y_{i_1}), \ \ldots, \ (\theta^*_{i_{K_n}}, Z_{i_{K_n}}, Y_{i_{K_n}}))$ denote the the minimum of $\tau((\theta^*_{i_1}, Z_{i_1}, Y_{i_1}), \ \ldots, \ (\theta^*_{i_{K_n}}, Z_{i_{K_n}}, Y_{i_{K_n}}); \boldsymbol{\omega})$ with respect to all possible vectors $\boldsymbol{\omega}$ satisfying the above constraints. Clearly, K_n, as well as the coordinates $(\theta^*_{i_j}, Z_{i_j}, Y_{i_j})$ of the observations in $\mathcal{X}_n(x, y; \nu_n^2, \nu_n)$, are random. By (6.9),

$$\left| E\left(\widehat{\theta}_{\mathrm{VRS}}(x, y \mid \mathcal{X}_n) - \theta(x,y) \right) - \theta(x,y) E\left(T_{K_n}\left((\theta^*_{i_1}, Z_{i_1}, Y_{i_1}), \ \ldots, \ (\theta^*_{i_{K_n}}, Z_{i_{K_n}}, Y_{i_{K_n}}) \right) \right) \right| \\ \leq C_3 n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}.$$

Obviously, $Z_{i_j} = \mathbf{Z}_x \zeta_{i_j}$ and the observations $(\theta_{i_j}, \zeta_{i_j}, Y_{i_j})$ are independent. The conditional distribution of $(\theta^*_{i_j}, \zeta_{i_j}, Y_{i_j})$ given $(X_{i_j}, Y_{i_j}) \in \mathcal{X}_n(x, y; \nu_n^2, \nu_n)$ converges to a uniform distribution.

By assumption densities are continuously differentiable; hence it can be shown that

$$\Big| E\big(T_{K_n}\big((\theta^*_{i_1}, Z_{i_1}, Y_{i_1}),\ \ldots,\ (\theta^*_{i_{K_n}}, Z_{i_{K_n}}, Y_{i_{K_n}})\big)\big) - E\left(T_{K_{n*}}\left((\theta^*_1, Z^*_1, Y^*_1),\ \ldots,\ (\theta^*_{K_n}, Z^*_{K_n}, Y^*_{K_n})\right)\right)\Big| \le C_6 n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}$$

for some $C_6 \in (0, \infty)$, where $Z^*_j = \mathbf{Z}_x \zeta^*_j$ and $(\theta^*_j, \zeta^*_j, Y^*_j)$ are iid random variables with a uniform distribution on $[0, \nu_n^2] \times [\nu_n, \nu_n]^{p-1} \times [y - \nu_n, y + \nu_n]^q$, and where $K^*_n \sim B(b^{p+q+1} n^{-1} \log n, n)$. A generalization of the localization argument leading to (6.6) now yields

$$\left|1 - \Pr\left[T_{K_{n*}}\left((\theta^*_1, Z^*_1, Y^*_1),\ \ldots,\ (\theta^*_{K_n}, Z^*_{K_n}, Y^*_{K_n})\right) = T_n\left((\widetilde{\theta}_1, \widetilde{Z}_1, \widetilde{Y}_1),\ \ldots,\ (\widetilde{\theta}_n, \widetilde{Z}_n, \widetilde{Y}_n)\right)\right]\right| = O(n^{-2}),$$

where $\widetilde{Z}_j = \mathbf{Z}_x \widetilde{\zeta}_j$, and $(\widetilde{\theta}_j, \widetilde{\zeta}_j, \widetilde{Y}_j)$ are iid random variables with a uniform distribution on $\left[0,\ \widetilde{f}^2\right] \times \left[-\widetilde{f},\ \widetilde{f}\right]^{p-1} \times \left[y - \widetilde{f},\ y + \widetilde{f}\right]^q$, where $\widetilde{f}$ is used to denote $\overline{f}_x(1, 0, y)^{-\frac{1}{p+q+1}}$.

It is now useful to consider still another transformation. Let $\widetilde{\nu}_n := \left(\frac{n}{\overline{f}_x(1,0,y)}\right)^{\frac{1}{p+q+1}}$, $\widetilde{Z}^{(n)}_j = \mathbf{Z}_x \widetilde{\zeta}^{(n)}_j$, and let $(\widetilde{\theta}^{(n)}_j, \widetilde{\zeta}^{(n)}_j, \widetilde{Y}^{(n)}_j)$ denote iid random variables with a uniform distribution on $[0, \widetilde{\nu}_n^2] \times [-\widetilde{\nu}_n, \widetilde{\nu}_n]^{p-1} \times [y - \widetilde{\nu}_n, y + \widetilde{\nu}_n]^q$. Then from the geometry of $\tau(\cdot)$, $T(\cdot)$, and from the properties of the uniform distribution, it follows that the distribution of $T_n((\widetilde{\theta}_1, \widetilde{Z}_1, \widetilde{Y}_1), \ldots, (\widetilde{\theta}_n, \widetilde{Z}_n, \widetilde{Y}_n))$ coincides with that of $n^{-\frac{2}{p+q+1}} T_n((\widetilde{\theta}^{(n)}_1, \widetilde{Z}^{(n)}_1, \widetilde{Y}^{(n)}_1),\ \ldots,\ (\widetilde{\theta}^{(n)}_n, \widetilde{Z}^{(n)}_n, \widetilde{Y}^{(n)}_n))$. In particular,

$$E\left(T_n\left((\widetilde{\theta}_1, \widetilde{Z}_1, \widetilde{Y}_1),\ \ldots,\ (\widetilde{\theta}_n, \widetilde{Z}_n, \widetilde{Y}_n)\right)\right) = E\left(n^{-\frac{2}{p+q+1}} T_n\left((\widetilde{\theta}^{(n)}_1, \widetilde{Z}^{(n)}_1, \widetilde{Y}^{(n)}_1),\ \ldots,\ (\widetilde{\theta}^{(n)}_n, \widetilde{Z}^{(n)}_n, \widetilde{Y}^{(n)}_n)\right)\right).$$

As $n \to \infty$, the distributions of $T_n\left((\widetilde{\theta}^{(n)}_1, \widetilde{Z}^{(n)}_1, \widetilde{Y}^{(n)}_1),\ \ldots,\ (\widetilde{\theta}^{(n)}_n, \widetilde{Z}^{(n)}_n, \widetilde{Y}^{(n)}_n)\right)$ converge rapidly to a fixed distribution with finite moments. In order to evaluate convergence of expectations, consider two different sample sizes, e.g., n and n^2. Again using the localization argument of Kneip et al. (2008), note that since the density of $(\widetilde{\theta}^{(n)}_1, \widetilde{\zeta}^{(n)}_1, \widetilde{Y}^{(n)}_1)$ decreases proportional to n, the analogue of the above value ν_n is now $b\left(\frac{\log n}{\overline{f}_x(1,0,y)}\right)^{\frac{1}{p+q+1}}$. For either $m = n$ or $m = n^2$, let $(\widetilde{\theta}^{(m)}_{i_j}, \widetilde{Z}^{(m)}_{i_j}, \widetilde{Y}^{(m)}_{i_j})$, $j = 1, \ldots, \widetilde{K}_m$ denote the coordinates of the K_m out of m random values satisfying $(1 - \widetilde{\theta}^{(m)}_{i_j}, \widetilde{Z}^{(m)}_{i_j}, \widetilde{Y}^{(m)}_{i_j}) \in C\left(x, y; \left(b\left(\frac{\log n^2}{\overline{f}_x(1,0,y)}\right)^{\frac{1}{p+q+1}}\right)^2, b\left(\frac{\log n^2}{\overline{f}_x(1,0,y)}\right)^{\frac{1}{p+q+1}}\right)$.

The number K_m as well as the corresponding coordinates are, of course, random. Then with probabilities of order $1 - m^{-2}$, $E\left(T_m\left((\widetilde{\theta}^{(m)}_1, \widetilde{Z}^{(m)}_1, \widetilde{Y}^{(m)}_1),\ \ldots,\ (\widetilde{\theta}^{(m)}_m, \widetilde{Z}^{(m)}_m, \widetilde{Y}^{(m)}_m)\right)\right) =$

$E\left(T_{\widetilde{K}_m}\left((\widetilde{\theta}_{i_1}^{(m)},\widetilde{Z}_{i_1}^{(m)},\widetilde{Y}_{i_1}^{(m)}),\ \ldots,\ (\widetilde{\theta}_{i_{\widetilde{K}_m}}^{(m)},\widetilde{Z}_{i_{\widetilde{K}_m}}^{(m)},\widetilde{Y}_{i_{\widetilde{K}_m}}^{(m)})\right)\right)$ for $m=n$ as well as $m=n^2$. Moreover, $\widetilde{K}_m \sim B(b^{p+q+1}m^{-1}\log n^2, m)$, where $B(\cdot,\cdot)$ denotes the Binomial distribution, while the (conditional) distributions of $(\widetilde{\theta}_{i_j}^{(m)},\widetilde{\zeta}_{i_j}^{(m)},\widetilde{Y}_{i_j}^{(m)})$ are uniform and do not depend on m. Hence, for any $k\in\mathbb{N}$, the conditional expectations $E\left(T_k\left((\widetilde{\theta}_{i_1}^{(m)},\widetilde{Z}_{i_1}^{(m)},\widetilde{Y}_{i_1}^{(m)}),\ \ldots,\ (\widetilde{\theta}_{i_k}^{(m)},\widetilde{Z}_{i_k}^{(m)},\widetilde{Y}_{i_k}^{(m)})\right)\mid \widetilde{K}_m=k\right)$ are bounded and independent of m, with

$$
\begin{aligned}
&E\left(T_{\widetilde{K}_m}\left((\widetilde{\theta}_{i_1}^{(m)},\widetilde{Z}_{i_1}^{(m)},\widetilde{Y}_{i_1}^{(m)}),\ \ldots,\ (\widetilde{\theta}_{i_{\widetilde{K}_m}}^{(m)},\widetilde{Z}_{i_{\widetilde{K}_m}}^{(m)},\widetilde{Y}_{i_{\widetilde{K}_m}}^{(m)})\right)\right)\\
&=\sum_{k=1}^{m} E\left(T_k\left((\widetilde{\theta}_{i_1}^{(m)},\widetilde{Z}_{i_1}^{(m)},\widetilde{Y}_{i_1}^{(m)}),\ \ldots,\ (\widetilde{\theta}_{i_k}^{(m)},\widetilde{Z}_{i_k}^{(m)},\widetilde{Y}_{i_k}^{(m)})\right)\mid \widetilde{K}_m=k\right)\times \Pr(K_m=k)\\
&=\sum_{k=1}^{n} E\left(T_k\left((\widetilde{\theta}_{i_1}^{(n)},\widetilde{Z}_{i_1}^{(n)},\widetilde{Y}_{i_1}^{(n)}),\ \ldots,\ (\widetilde{\theta}_{i_k}^{(n)},\widetilde{Z}_{i_k}^{(n)},\widetilde{Y}_{i_k}^{(n)})\right)\mid \widetilde{K}_n=k\right)\frac{s_n^k}{k!}e^{-s_n}+O\left(m^{-1}\log n^2\right),
\end{aligned}
\tag{6.11}
$$

for both $m=n$ and $m=n^2$, where $s_n=b^{p+q+1}\log n^2$, and with the last equality following from well-known results on the approximation of a binomial distribution by the Poisson distribution. The expectations in (6.11) converge rapidly to a finite, fixed value; therefore, for all sufficiently large n,

$$
\left|E\left(\widehat{\theta}_{\mathrm{VRS}}(x,y\mid\mathcal{X}_n)-\theta(x,y)\right)-\theta(x,y)n^{-\frac{2}{p+q+1}}C_{g''_x,\overline{f}_x(1,0,y)}\right|\le C_7 n^{-\frac{3}{p+q+1}}(\log n)^{\frac{3}{p+q+1}} \tag{6.12}
$$

for all $(x,y)\notin\mathcal{W}(\nu_n^*)$ and some $C_7\in(0,\infty)$. Here,

$$
C_{g''_x,\overline{f}_x(1,0,y)} := \lim_{n\to\infty} E(T_n((\widetilde{\theta}_1^{(n)},\widetilde{Z}_1^{(n)},\widetilde{Y}_1^{(n)}),\ \ldots,\ (\widetilde{\theta}_n^{(n)},\widetilde{Z}_n^{(n)},\widetilde{Y}_n^{(n)}))) \tag{6.13}
$$

depends only upon g''_x and $\overline{f}_x(1,0,y)$, and changes continuously in (x,y). Similar arguments establish that there exists some $C_8\in(0,\infty)$ such that

$$
E\left(\left|\widehat{\theta}_{\mathrm{VRS}}(x,y\mid\mathcal{X}_n)-\theta(x,y)\right|^2\right)\le C_8 n^{-\frac{4}{p+q+1}} \tag{6.14}
$$

for all $(x,y)\notin\mathcal{W}(\nu_n^*)$.

Case (ii): Now turn to the alternative case where $(x,y)\in\mathcal{W}(\nu_n^*)$. In this case, in the construction leading to (6.6)–(6.10), the problem that some of the sets $B_j\subset C(x,y;\nu_n^2,\nu_n)$ surrounding $(0,y)$ surpass the boundary and are no longer in $\mathcal{D}$ arises. As a consequence, one cannot exclude that $\widehat{\theta}_{\mathrm{VRS}}(x,y\mid\mathcal{X}_n)$ is influenced by an observation with $\theta_i\le 1-\nu_n^2$. However, by the assumed almost strict convexity of $\mathcal{D}$, it is easily verified that a similar argument allows exclusion of all observations with $(\theta_i, Z_i, Y_i) \notin C(x, y; 1, \nu_n) \cap \mathcal{D}$, and for sufficiently large n, it follows that

$$\left|1 - \Pr\left(\widehat{\theta}_{\mathrm{VRS}}(x, y \mid \mathcal{X}_n) = \widehat{\theta}_{\mathrm{VRS}}(x, y \mid \mathcal{X}_n(x, y; 1, \nu_n))\right)\right| \leq C_9 n^{-2} \tag{6.15}$$

for all $(x, y) \in \mathcal{W}(\nu_n^*)$, for some $C_9 \in (0, \infty)$, and all sufficiently large n. Note that for any subset $\mathcal{X}_n^* \subset \mathcal{X}_n$, $\theta(x, y) \leq \widehat{\theta}_{\mathrm{VRS}}(x, y \mid \mathcal{X}_n) \leq \widehat{\theta}_{\mathrm{CH}}(x, y \mid \mathcal{H}_n^*)$, where $\mathcal{H}_n^*$ denotes the convex hull of $\mathcal{X}_n^*$. For $r = 1, \ldots, p-1$, define

$$v_{r;x,y} := \min_{\substack{(\zeta, y^*) \in \mathbb{R}^{p-1} \times \mathbb{R}^q, \\ (\sum_{j=1}^{p-1} \zeta_j z_x^{(j)}, y^*) \in \Psi^{*\partial}(x)}} \left\{\nu_n,\ |\zeta_r|\right\}. \tag{6.16}$$

Similarly, for $r = 1, \ldots, q$, define $v_{p-1+r;x,y}$ by replacing $|\zeta_r|$ with $|y_r - y_r^*|$ in (6.16). These $v_{r;x,y}$ can be viewed as measuring a "distance" from (x, y) to the boundary, with $v_{r;x,y} \leq \nu_n$.

If $\prod_{r=1}^{p+q-1} v_{r;x,y} \geq \nu_n^{p+q+1}$, i.e. (x, y) is not too near the boundary, an upper bound for $\widehat{\theta}_{\mathrm{VRS}}(x, y \mid \mathcal{X}_n)$ can then be obtained by relying on the observations in $C(x, y; \frac{\nu_n^{p+q+1}}{\prod_{r=1}^{p+q-1} v_{r;x,y}}, \nu_n) \cap \mathcal{D}$. Arguments similar to those used to establish (6.10) then show that for all $(x, y) \in \mathcal{W}(\nu_n^*)$ with $\prod_{r=1}^{p+q-1} v_{r;x,y} \geq \nu_n^{p+q+1}$, we have for $\alpha \in \{1, 2\}$

$$E\left(\left|\widehat{\theta}_{\mathrm{VRS}}(x, y \mid \mathcal{X}_n) - \theta(x, y)\right|^\alpha\right) \leq C_{10}^\alpha \left(\frac{\nu_n^{p+q+1}}{\prod_{r=1}^{p+q-1} v_{r;x,y}}\right)^\alpha, \tag{6.17}$$

for some constant $C_{10} \in (0, \infty)$, and for all sufficiently large n.

Now the moments of $\widehat{\theta}_{\mathrm{VRS}}(X_i, Y_i \mid \mathcal{X}_n)$ can be analyzed. Let $\mathcal{X}_{n,-i}$ denote the sample of size $n-1$ obtained by eliminating the i-th observation (X_i, Y_i). When relying on $\mathcal{X}_{n,-i}$, it is clear that all constants in (6.2)–(6.17) can be chosen independently of x and thus also apply for the (random) coordinate system induced by the specific choice $(x, y) = (X_i, Y_i)$. Obviously,

$$\widehat{\theta}_{\mathrm{VRS}}(X_i, Y_i \mid \mathcal{X}_n) = \min\left\{\widehat{\theta}_{\mathrm{VRS}}(X_i, Y_i \mid \mathcal{X}_{n,-i}),\ 1\right\}. \tag{6.18}$$

Since (X_i, Y_i) is independent of $\mathcal{X}_{n,-i}$, (6.12) and (6.18) imply that

$$E\left(\widehat{\theta}_{\mathrm{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i) \mid (X_i, Y_i) \notin \mathcal{W}(\nu_n^*)\right) = C_0 n^{-\frac{2}{p+q+1}} + O\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}\right) \tag{6.19}$$

for some $C_0 \in (0, \infty)$. If $p = 1$ and $q = 0$, then assertion (3.1) follows directly from (6.19), since in this case there is no boundary problem due to $\mathcal{W}(\nu_n^*) = \emptyset$.

In order to quantify the influence of boundary effects for $p+q \geq 2$, let $\mathcal{W}_{n,1} := \left\{(x,y) \in \mathcal{D} \mid \nu_n^{p+q-1} > \prod_{r=1}^{p+q-1} v_{r;x,y} \geq \nu_n^{p+q+1}\right\}$ contain points in $\mathcal{W}(\nu_n^*)$ but not too near the boundary, and let $\mathcal{W}_{n,2} := \left\{(x,y) \in \mathcal{D} \mid \prod_{r=1}^{p+q-1} v_{r;x,y} < \nu_n^{p+q+1}\right\}$ contain the other points of $\mathcal{W}(\nu_n^*)$ very near the boundary where only the trivial upper bound $|\widehat{\theta}_{\mathrm{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)| \leq 1$ can be used instead of the bound given in (6.17). For points in $\mathcal{W}_{n,1}$, note that for all $r = 1, \ldots, p+q-1$, $\nu_n^3 \leq v_{r;x,y} \leq \nu_n$. Straightforward calculations yield that with $C_{11} = C_{10} \sup_{(x,y)\in\mathcal{W}_{n,1}} f(x,y)$

$$E\left(\widehat{\theta}_{\mathrm{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i) \mid (X_i, Y_i) \in \mathcal{W}_{n,1}\right) \leq C_{10} \int_{\mathcal{W}_{n,1}} \frac{\nu_n^{p+q+1}}{\prod_{r=1}^{p+q-1} v_{r;x,y}} f(x,y) dx dy$$

$$\leq C_{11} \sum_{r=1}^{p+q-1} \int_{\substack{\{(x,y)\in\mathcal{D}| \\ \nu_n > v_{r,x,y} \geq \nu_n^3\}}} \frac{\nu_n^3}{v_{r;x,y}}\, dx\, dy + O_p\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}\right) = O_p\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}} \log n\right), \tag{6.20}$$

as well as $E\left(\widehat{\theta}_{\mathrm{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i) \mid (X_i, Y_i) \in \mathcal{W}_{n,2}\right) \leq \Pr((X_i, Y_i) \in \mathcal{W}_{n,2}) = O\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}\right)$. Together with (6.18) and (6.19), this leads to (3.1).

If $p = 1, q = 0$, no boundary problem exists, and (3.2) is a consequence of (6.18) and (6.14). For $p+q \geq 2$, assertion (3.2) follows from the fact that relations (6.14)–(6.18) imply

$$\begin{aligned}
\mathrm{Var}\left(\widehat{\theta}_{\mathrm{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right) &\leq C_8 n^{-\frac{4}{p+q+1}} \times \Pr\left((X_i, Y_i) \notin \mathcal{W}(\nu_n^*)\right) \\
+ C_{10}^2 \int_{\mathcal{W}_{n,1}} \left(\frac{\nu_n^6}{\prod_{r=1}^{p+q-1} v_{r;x,y}^2}\right) & f(x,y)\, dx\, dy + \Pr\left((X_i, Y_i) \in \mathcal{W}_{n,2}\right) \\
&= O\left(n^{-\frac{4}{p+q+1}} + n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}\right).
\end{aligned} \tag{6.21}$$

It remains to prove (3.3). For all $i, j \in 1, \ldots, n$, $i \neq j$, relations (6.6)–(6.15) together with (6.18) imply that $\left(\widehat{\theta}_{\mathrm{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right)$ and $\left(\widehat{\theta}_{\mathrm{VRS}}(X_j, Y_j \mid \mathcal{X}_n) - \theta(X_j, Y_j)\right)$ are asymptotically uncorrelated if $\mathcal{X}_n(X_i, Y_i; 1, \nu_n^*) \cap \mathcal{X}_n(X_j, Y_j; 1, \nu_n^*) = \emptyset$. Since all observations are iid, the Cauchy-Schwarz inequality yields

$$\begin{aligned}
&\left|\mathrm{COV}\left(\widehat{\theta}_{\mathrm{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i),\ \widehat{\theta}_{\mathrm{VRS}}(X_j, Y_j \mid \mathcal{X}_n) - \theta(X_j, Y_j)\right)\right| \\
&\quad \leq \Pr\left(\mathcal{X}_n(X_i, Y_i; 1, \nu_n^*) \cap \mathcal{X}_n(X_j, Y_j; 1, \nu_n^*) \neq \emptyset\right) \times \mathrm{VAR}\left(\widehat{\theta}_{\mathrm{VRS}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right) + O\left(n^{-2}\right).
\end{aligned} \tag{6.22}$$

Relation (3.2) as well as $\Pr\left(\mathcal{X}_n(X_i, Y_i; 1, \nu_n^*) \cap \mathcal{X}_n(X_j, Y_j; 1, \nu_n^*) \neq \emptyset\right) = O\left(n^{-\frac{p+q-1}{p+q+1}} (\log n)^{\frac{p+q-1}{p+q+1}}\right)$ now lead to assertion (3.3), completing the proof of the theorem. ∎

6.2 Proof of Theorem 3.2

For all $(x,y) \in \mathcal{H}_n^0$,

$$\begin{aligned}\widehat{\theta}_{\text{CRS}}(x,y) &= \min\left\{\theta \mid ay \leq \boldsymbol{Y}\boldsymbol{\omega},\ \theta ax \geq \boldsymbol{X}\boldsymbol{\omega},\ \boldsymbol{i}_n'\boldsymbol{\omega} = 1,\ \text{for some } a \in \mathbb{R}_+\ , \boldsymbol{\omega} \in \mathbb{R}_+^n\right\} \\ &= \min\left\{\theta \mid (\theta ax, ay) \in (\mathcal{H}_n^0)^\partial \text{ for some } a \in \mathbb{R}_+\right\}, \end{aligned} \tag{6.23}$$

since the solution is necessarily attained on the boundary $(\mathcal{H}_n^0)^\partial$ of the convex hull $\mathcal{H}_n^0$. This means that the solution can be determined by replacing inequalities by equalities, and

$$\widehat{\theta}_{\text{CRS}}(x,y) = \min\left\{\theta \mid y = a\boldsymbol{Y}\boldsymbol{\omega},\ \theta x = a\boldsymbol{X}\boldsymbol{\omega},\ \boldsymbol{i}_n'\boldsymbol{\omega} = 1,\ \text{for some } a \in \mathbb{R}_+\ , \boldsymbol{\omega} \in \mathbb{R}_+^n\right\} \tag{6.24}$$

for $(x,y) \in \mathcal{H}_n^0$.

Now recall from the proof of Theorem 3.1 that each observation (X_i, Y_i) can be equivalently represented by the corresponding vector $(\theta(X_i,Y_i), Z_i, Y_i)$. By (6.1), it follows from (6.24) that

$$\begin{aligned}\widehat{\theta}_{\text{CRS}}(x,y \mid \mathcal{X}_n) = \min_{\boldsymbol{\omega}\in\mathbb{R}_+^n} \Bigg\{ \sum_{i=1}^n \boldsymbol{\omega}_i \frac{g_{ax}(\theta(aX_i,aY_i)aZ_i,aY_i)}{\|ax\|\theta(aX_i,aY_i)} \mid \\ \boldsymbol{Z}\boldsymbol{\omega} = 0,\ a\boldsymbol{Y}\boldsymbol{\omega} = y,\ \boldsymbol{i}_n^T\boldsymbol{\omega} = 1,\ a \in \mathbb{R}_+^1,\ \boldsymbol{\omega} \in \mathbb{R}_+^n \Bigg\}. \end{aligned} \tag{6.25}$$

First, note that by definition of g_x, $g_{ax}(\cdot) \equiv g_x(\cdot)$. Kneip et al. (2008) show that necessarily $\theta(v,y)\frac{x^Tv}{||x||} = g_x(\theta(v,y)z,y)$ for all $(v,y),(x,y) \in \Psi$. By the CRS assumption, $\theta(ax,ay) = a\theta(x,y)$ for all $a > 0$, and hence $g_x(\theta(ax,ay)az,ay) = a^2 g_x(\theta(x,y)z,y)$. Consequently,

$$\frac{g_{ax}(\theta(ax,ay)az,ay)}{\|ax\|\theta(ax,ay)} = \frac{a^2 g_x(\theta(x,y)z,y)}{\|ax\|\theta(ax,ay)} = \frac{g_x(\theta(x,y)z,y)}{\|x\|\theta(x,y)}, \tag{6.26}$$

and therefore

$$\begin{aligned}\widehat{\theta}_{\text{CRS}}(x,y \mid \mathcal{X}_n) = \min_{\boldsymbol{\omega}} \Bigg\{ \sum_{i=1}^n \boldsymbol{\omega}_i \frac{g_x(\theta(X_i,Y_i)Z_i,Y_i)}{\|x\|\theta(X_i,Y_i)} \mid \\ \boldsymbol{Z}\boldsymbol{\omega} = 0,\ \boldsymbol{Y}\boldsymbol{\omega} = y,\ \boldsymbol{i}_n^T\boldsymbol{\omega} = a,\ a \in \mathbb{R}_+^1,\ \boldsymbol{\omega} \in \mathbb{R}_+^n \Bigg\}. \end{aligned} \tag{6.27}$$

For $y = (y^1,\ \ldots,\ y^q)'$ with $y^1 > 0$ consider the transformations $x^* = x/y^1$, $z^* = z/y^1$, $y^* = y/y^1 = (1, y^2/y^1, \ldots, y^q/y^1)'$ and $\widetilde{y} = (y^2/y^1, \ldots, y^q/y^1)' \in \mathbb{R}^{q-1}$. With respect to the $q-1$ dimensional output variable $\widetilde{y}$, a production set $\widetilde{\Psi} := \{(x^*, \widetilde{y}) \mid (x^*, (1, \widetilde{y}')') \in \Psi$ can be defined. Efficiencies $\theta^*(x^*, \widetilde{y}) := \theta(x^*, (1, \widetilde{y}')')$ can be similarly defined, as well as a function $g^*_{x^*}(z^*, \widetilde{y}) := g_x(z^*, (1, \widetilde{y}')')$. Furthermore, the density f of (X_i, Y_i) induces a density f^* of

$(X_i^*, \widetilde{Y}_i)$. By Assumption 3.2–3.5, it is easily verified that $\widetilde{\Psi}$, as well as the support $\widetilde{\mathcal{D}}$ of f^*, satisfy conditions Assumptions 3.4–3.2. Moreover, smoothness of θ and f translates into a corresponding smoothness of θ^* and f^*. Therefore, the assumptions of Theorem 3.1 are fulfilled when considering the VRS-DEA estimator for the sample $\mathcal{X}_n^* = \{(X_i^*, \widetilde{Y}_i)\}$. Note that the proof of Theorem 3.1 is heavily based on the asymptotic equivalence of $\widehat{\theta}_{\text{VRS}}(\cdot)$ and a corresponding convex-hull estimator $\widehat{\theta}_{\text{CH}}(\cdot)$. The assertions of the theorem thus remain true when replacing $\widehat{\theta}_{\text{VRS}}(x/y^1, \widetilde{y} \mid \mathcal{X}_n^*)$ with $\widehat{\theta}_{\text{CH}}(x/y^1, \widetilde{y} \mid \mathcal{H}_n^*)$.

For any vector $\boldsymbol{\omega} \in \mathbb{R}_+^n$ with $\sum_{i=1}^n \boldsymbol{\omega}_i Z_i = 0$ and $\sum_{i=1}^n \boldsymbol{\omega}_i Y_i = y$, it follows from (6.26) that the corresponding vector $\boldsymbol{\omega}^* \in \mathbb{R}_+^n$ defined by $\boldsymbol{\omega}_i^* = \boldsymbol{\omega}_i \frac{Y_i^1}{y^1}$, $i = 1, \ldots, n$, satisfies $\boldsymbol{Z}^* \boldsymbol{\omega}^* = 0$, $\widetilde{\boldsymbol{Y}} \boldsymbol{\omega}^* = \widetilde{y}\ \boldsymbol{i}_n' \boldsymbol{\omega}^* = 1$, and

$$\sum_{i=1}^n \boldsymbol{\omega}_i \frac{g_x(\theta(X_i, Y_i) Z_i, Y_i)}{\|x\| \theta(X_i, Y_i)} = \sum_{i=1}^n \boldsymbol{\omega}_i \frac{g_x(\theta(X_i^*, Y_i^*) Z_i^*, Y_i^*)}{\|\frac{1}{Y_i^1} x\| \theta(X_i^*, Y_i^*)} = \sum_{i=1}^n \boldsymbol{\omega}_i^* \frac{g_{x^*}^*(\theta^*(X_i^*, \widetilde{Y}_i) Z_i^*, \widetilde{Y}_i)}{\|x^*\| \theta^*(X_i^*, \widetilde{Y}_i)}.$$

The minimization problem (6.27) allows for arbitrary values of $\boldsymbol{i}_n' \boldsymbol{\omega} = \sum_{i=1}^n \boldsymbol{\omega}_i^* \frac{y^1}{Y_i^1}$, and hence

$$\begin{aligned} \widehat{\theta}_{\text{CRS}}(x, y \mid \mathcal{X}_n) &= \min_{\boldsymbol{\omega}^*} \Big\{ \sum_{i=1}^n \boldsymbol{\omega}_i^* \frac{g_{x^*}^*(\theta^*(X_i^*, \widetilde{Y}_i) Z_i^*, \widetilde{Y}_i)}{\|x^*\| \theta^*(X_i^*, \widetilde{Y}_i)} \boldsymbol{Z}^* \boldsymbol{\omega}^* = 0,\ \widetilde{\boldsymbol{Y}} \boldsymbol{\omega}^* = \widetilde{y},\ \boldsymbol{i}_n' \boldsymbol{\omega}^* = 1, \boldsymbol{\omega}^* \in \mathbb{R}_+^n \Big\} \\ &= \widehat{\theta}_{\text{CH}}(x^*, \widetilde{y} \mid \mathcal{H}_n^*) \end{aligned} \tag{6.28}$$

This proves (3.4), and (3.6)–(3.8) are immediate consequences of Theorem 3.1. Finally, the asserted inequality (3.5) for the VRS-DEA estimator of (x, y) follows from

$$\begin{aligned} &\widehat{\theta}_{\text{VRS}}(x, y \mid \mathcal{X}_n^0) = \widehat{\theta}_{\text{CH}}(x, y \mid \mathcal{H}_n^0) \\ &\quad = \min_{\boldsymbol{\omega}} \Big\{ \sum_{i=1}^n \boldsymbol{\omega}_i \frac{g_x(\theta(X_i, Y_i) Z_i, Y_i)}{\|x\| \theta(X_i, Y_i)} \mid \boldsymbol{Z} \boldsymbol{\omega} = 1,\ \boldsymbol{Y} \boldsymbol{\omega} = y,\ \boldsymbol{i}_n' \boldsymbol{\omega} \le 1, \boldsymbol{\omega} \in \mathbb{R}_+^n \Big\} \\ &\quad = \min_{\boldsymbol{\omega}^*} \Big\{ \sum_{i=1}^n \boldsymbol{\omega}_i^* \frac{g_{x^*}^*(\theta^*(X_i^*, \widetilde{Y}_i) Z_i^*, \widetilde{Y}_i)}{\|x^*\| \theta^*(X_i^*, \widetilde{Y}_i)} \mid \boldsymbol{Z}^* \boldsymbol{\omega}^* = 1,\ \widetilde{\boldsymbol{Y}} \boldsymbol{\omega}^* = \widetilde{y},\ \boldsymbol{i}_n' \boldsymbol{\omega}^* \le 1, \boldsymbol{\omega}^* \in \mathbb{R}_+^n \Big\} \\ &\quad \le \min_{\boldsymbol{\omega}^*} \Big\{ \sum_{i: y_{i1} \ge y_1} \boldsymbol{\omega}_i^* \frac{g_{x^*}^*(\theta^*(X_i^*, \widetilde{Y}_i) Z_i^*, \widetilde{Y}_i)}{\|x^*\| \theta^*(X_i^*, \widetilde{Y}_i)} \mid \\ &\qquad\qquad \sum_{i: y_{i1} \ge y_1} \boldsymbol{\omega}_i^* = 1, \sum_{i: y_{i1} \ge y_1} \boldsymbol{\omega}_i^* Z_i^* = 0, \sum_{i: y_{i1} \ge y_1} \boldsymbol{\omega}_i^* \widetilde{Y}_i = \widetilde{y},\ \boldsymbol{\omega}^* \in \mathbb{R}_+^n \Big\}. \quad \blacksquare \end{aligned}$$

6.3 Proof of Theorem 3.3

The proof relies on the arguments in Park et al. (2000). For any $(x, y) \in \mathcal{D}$ we may define the function $g^p(x_1, \ldots, x_{p-1}, y) := \theta(x, y) x_p$. Note that g^p does not depend on x_p, and that by our assumptions the absolute values of all partial derivatives of g^p are (uniformly) bounded away from 0. Furthermore, $x_p = \frac{g^1(x_1,\ldots,x_{p-1})}{\theta(x,y)}$, which means that each observation (X_i, Y_i) can be equivalently represented in the coordinate system given by $(\theta_i, X_{i1}, \ldots, X_{i,p-1}, Y_i)$ with $\theta_i := \theta(X_i, Y_i)$. For $h_1, h_2 > 0$, let

$$C^{FDH}(x,y;h_1,h_2) = \Big\{(\widetilde{x},\widetilde{y}) \in \mathcal{D} \mid 1-\theta \leq h_1;\ x_j - h_2 \leq \widetilde{x}_j \leq x_j;\ y_r \leq \widetilde{y}_r \leq y_r + h_2; \\ j = 1,\ \ldots,\ p-1,\ r = 1,\ \ldots,\ q\Big\},$$

and let $\mathcal{X}_n^{FDH}(x,y;h_1,h_2) := \{(X_i,Y_i) \in \mathcal{X}_n \mid (X_i,Y_i) \in C^{FDH}(x,y;h_1,h_2)\}$, as well as

$$\mathcal{W}(h)^{FDH} := \Big\{(x,y) \in \mathcal{D} \mid \min\{\min_{j=1,\ \ldots,\ p-1} |x_j - \widetilde{x}_j|, \min_{r=1,\ \ldots,\ q} |y_r - \widetilde{y}_r|\} \leq h \\ \text{for some } (\widetilde{x},\widetilde{y}) \in \mathcal{D}^* \text{ with } \widetilde{x} \leq x,\ \widetilde{y} \geq y\Big\}$$

Moreover, set $d_{inf} = \inf_{(x,y)\in\mathcal{D}^*} f(x,y)$. By a straightforward generalization of the localization arguments of Park et al. (2000, proof of Theorem 3.1), one can show that there is a constant $M_1 \in (0,\infty)$ such that with $\gamma_n^* := 2(\frac{\log n}{d_{inf} n})^{\frac{1}{p+q}}$ as well as $\gamma_n := 2(\frac{\log n}{nf(\theta(x,y)x,y)})^{\frac{1}{p+q}}$,

$$\Big|1 - \Pr\left(\widehat{\theta}_{\mathrm{FDH}}(x,y \mid \mathcal{X}_n) = \widehat{\theta}_{\mathrm{FDH}}(x,y|\mathcal{X}_n^{FDH}(x,y;\gamma_n,\gamma_n))\right)\Big| \leq M_1 n^{-2} \tag{6.29}$$

for $(x,y) \notin \mathcal{W}(\gamma_n^*)^{FDH}$. Asymptotic moments have already been derived in Park et al. (2000). In particular, it is shown that for some $C_{x,y}^{FDH} \in (0,\infty)$, $E(\widehat{\theta}_{\mathrm{FDH}}(x,y \mid \mathcal{X}_n)) - \theta(x,y) = C_{x,y}^{FDH} n^{-\frac{1}{p+q}} + o(n^{-\frac{1}{p+q}})$. The constants $C_{x,y}^{FDH}$ depend on the partial derivatives of $g^p(x_1,\ldots,x_{p-1},y)$ as well as on $f(\theta(x,y)x,y)$, and their analytical structure is derived in Park et al. (2000). Moreover, the arguments used to prove Theorem 3.4 of Park et al. (2000) even allow a more precise approximation of the remainder term in the bias approximation. Under our conditions we can conclude that there are constants $0 < M_3, M_4 < \infty$ such that

$$\left|E\left(\widehat{\theta}_{\mathrm{FDH}}(x,y \mid \mathcal{X}_n) - \theta(x,y)\right) - \theta(x,y) - n^{-\frac{1}{p+q}} C_{x,y}^{FDH}\right| \leq M_3 n^{-\frac{2}{p+q}} (\log n)^{\frac{2}{p+q}}$$

and $\mathrm{VAR}(\widehat{\theta}_{\mathrm{FDH}}(x,y \mid \mathcal{X}_n)) \leq M_4 n^{-\frac{2}{p+q}}$ for all $(x,y) \notin \mathcal{W}(\gamma_n^*)^{FDH}$. Now consider points $(x,y) \in \mathcal{W}(\gamma_n^*)^{FDH}$, and define $v_{r;x,y}^* := \min_{(\widetilde{x},\widetilde{y})\in\mathcal{D}^*}\{|x_r - \widetilde{x}_r|\}$, $r = 1,\ldots,p-1$, as well as $v_{s;x,y}^* := \min_{(\widetilde{x},\widetilde{y})\in\mathcal{D}^*}\{|y_{s-p+1} - \widetilde{y}_{s-p+1}|\}$, $s = p,\ldots,p+q-1$. Arguments similar to those leading to (6.17) may be used to show that for some $M_5 \in (0,\infty)$ and $\alpha \in \{1,2\}$,

$$E\left(\left|\widehat{\theta}_{\mathrm{FDH}}(x,y \mid \mathcal{X}_n) - \theta(x,y)\right|^{\alpha}\right) \leq M_5^{\alpha} \left(\frac{\gamma_n^{p+q}}{\Pi_{r=1}^{p+q-1} v_{r;x,y}^*}\right)^{\alpha} \tag{6.30}$$

for all $(x, y) \in \mathcal{W}(\gamma_n^*)$ with $\gamma_n^{p+q-1} > \prod_{r=1}^{p+q} \upsilon_{r;x,y}^* \geq \gamma_n^{p+q}$. Applying arguments similar to (6.18)–(6.22), the assertions of the theorem now follow from straightforward calculations. ■

7 Monte Carlo Evidence

For purposes of Monte Carlo experiments to analyze the coverages of estimated confidence intervals, two technologies are considered. The first is characterized by variable, non-constant returns to scale, and consists of the part of a (hyper)-sphere centered at $(\mathbf{1}_p, \mathbf{0}_q)$ lying in the space $[0,1]^p \times [0,1]^q$, where $\mathbf{1}_p$ denotes a p-vector of ones and $\mathbf{0}_q$ denotes a q-vector of zeros. The second technology is characterized by globally constant returns to scale, where

$$y = \prod_{j=1}^{p} x_j^{1/p} \tag{7.1}$$

with $x_j \in (0,1)\ \forall\ j \in \{1,\ \ldots,\ p\}$. In addition, let the true (marginal) density of the input inefficiency be given by

$$f_\theta(t) = \begin{cases} 3t^2 & \forall\ t \in [0,1], \\ 0 & \text{otherwise} \end{cases} \tag{7.2}$$

so that $E(\theta) = \mu_\theta = 0.75$. In each Monte Carlo experiment, n points (X_i^{eff}, Y_i) are generated uniformly along the simulated technology, and then projected away from the frontier using draws from f_θ to compute $X_i = \theta_i^{-1} X_i^{\text{eff}}$ to create a simulated sample $\mathcal{X}_n = \{(X_i, Y_i)\}_{i=1}^n$.[6]

Each experiment consists of 1,000 Monte Carlo trials. On each trial, a sample of size n is generated, and efficiency is estimated for each simulated observation (X_i, Y_i), using the entire simulated sample as the reference set. Let $\widehat{\mu}_{n_\kappa}$ denote the rescaled sample mean defined by (4.12), computed from either VRS-DEA, CRS-DEA, or FDH efficiency estimates. Confidence intervals are then estimated on each Monte Carlo trial for μ_θ using each of three methods:

(i) normal approximation using (4.22), based on Theorem 4.3, using $\widehat{\mu}_n$ to estimate μ_θ, but incorporating (4.20) to correct for bias;

(ii) normal approximation using (4.24), based on Theorem 4.4, using $\widehat{\mu}_{n_\kappa}$ to estimate μ_θ and incorporating (4.20) to correct for bias; and

[6] Points uniformly distributed on the surface of a hyper-sphere are generated using the method of Muller (1959) and Marsaglia (1972).

(iii) normal approximation using (4.25), based on averaging the intervals obtained by method (iii) over the possible disjoint subsets of size n_κ; and

For each of these three methods, the proportion of Monte Carlo trials where the estimated confidence intervals cover the true value $\mu_\theta = 0.75$ are reported as estimated coverages. Experiments are conducted with two sets of four different dimensionalities. For the VRS technology, these are $p = q = 1$, $p = 2,\ q = 1$, $p = q = 2$, and $p = q = 3$. For the CRS technology, the dimensions are $p = q = 1$; $p = 2,\ q = 1$; $p = 3,\ q = 1$; and $p = 5,\ q = 1$.

Table 1 shows estimated coverages using methods (i)–(iii) listed above for the case of the VRS-DEA estimator and the VRS technology. Since these methods use normal approximations, no bootstrapping is required; this avoids some computational burden, and so experiments were performed in each case for sample sizes $n \in \{100,\ 200,\ 500,\ 1,000,\ 5,000,\ 10,000\}$ and with two, three, four, and six dimensions as described earlier. Table 1 gives three sets of results corresponding to methods (i)–(iii); in each set, estimated coverages of 90-, 95-, and 99-percent confidence intervals are reported. Tables 2 and 3 are organized similarly, with Tables 2 giving results for the CRS-DEA estimator and the CRS technology, and Table 3 showing results for the FDH estimator and the VRS technology.

Turning to the first column of results in Table 1 corresponding to method (i), recall that Theorem 4.3 holds for $\kappa \geq 2/5$. For the VRS-DEA estimator, under VRS, this holds if and only if $p+q \leq 4$. The results shown in column (i) of Table 4.3 confirm that method (i) "works" in the sense that for a given number of dimensions, coverages increase with sample size and approach the nominal coverages (the results in the table show small decreases in coverages in some cases when sample size is increased from 5,000 to 10,000, but the decreases are not statistically significant). For given sample sizes, coverages worsen slightly as the number of dimensions increases, as expected. However, with six dimensions ($p = q = 3$), coverages obtained with method (i) are poor, and begin to decline significantly when the sample size is increased from 5,000 to 10,000. This is to be expected, since Theorem 4.3 holds in this case only if the number of dimensions is no more than four.

Results shown for method (ii) in Table 1 are identical to those shown for method (i) for $p = q = 1$ and $p = 2,\ q = 1$. This is due to the fact that, for the VRS-DEA estimator, $n_\kappa = n$. With $p = q = 2$ and $p = q = 3$, coverages attained by method (ii) are greater, and closer to nominal values, than those achieved by method (i). In particular, method (ii) does not break down for the six-dimensional case, unlike method (i). With method (ii), achieved coverages are close to their nominal values with $n = 500$, even with six dimensions. For the case where $(p + q) = 4$, coverages by method (ii) are better than those obtained with method (i) (i.e., closer to nominal coverages) in every case, confirming the remarks in the last paragraph of Section 4.

Coverages attained by method (iii) for $p = q = 1$ and $p = 2,\ q = 1$ are also identical to those achieved by methods (i)–(ii), again due to the fact that $n_\kappa = n$ in these cases. For $p = q = 2$, however, the coverages achieved by method (iii) exceed nominal coverages for $n \geq 500$; with $p = q = 3$, achieved coverages are greater than nominal coverages for $n \geq 200$. Recall from the discussion in Section 4 that the intervals in (4.24) and (4.25) are of the same width, and differ only in where they are centered. The width of the intervals in method (ii) reflect the greater uncertainty in $\widehat{\mu}_{n_\kappa}$, as opposed to $\widehat{\mu}_n$ in method (iii). Since method (iii) centers on an estimator of μ_θ with less mean-square error than does method (ii), coverages of intervals constructed using method (iii) are larger than those constructed using method (ii). In addition, since the intervals obtained by method (ii) have asymptotically correct coverage, the intervals from method (iii) must necessarily have coverages larger than nominal values. The results in Table 1 show that the intervals from method (iii) eventually yield coverages of 100-percent as the sample size is increased.

Ordinarily, one might reject intervals that cover in every case; typically, this would happen when intervals are too wide to be informative. Here, however, the situation is different—the intervals obtained from method (iii) are, by construction, of exactly the same width as the intervals from method (ii) which have asymptotically correct coverages. Although the interpretation might differ, the intervals from method (iii) are more informative about μ_θ than those from method (ii).

Turn now to the results in Table 2 obtained with the CRS-DEA estimator and the CRS technology. Results obtained with methods (i)–(iii) are identical for the cases with two, three, or four dimensions. This is due to the fact that for the CRS-DEA estimator under CRS, $n_\kappa = n$ for $p + q \leq 4$. Method (i) is seen in Table 2 to "work" when the number of dimensions is two, three, or four, but not when $p + q = 6$. Again, this is due to the fact that Theorem 4.3 is valid only for $\kappa \geq 2/5$, or $p + q \leq 5$. Results for methods (i)-(iii) with the CRS-DEA estimator are qualitatively similar to those for the VRS-DEA estimator. For smaller numbers of dimensions, the coverages in Table 2 are in many cases closer to the corresponding nominal coverages than are the coverages in Table 1 when the sample size is 100 or 200. This may be due to the faster convergence rate of the CRS-DEA estimator.

In the case of the FDH estimator, the condition $\kappa \geq 1/3$ in Theorem 4.3 means that $p + q$ must be no more than three in order for method (i) to provide. The first column of results in Table 3 indicate that with the FDH estimator, method (i) yields good coverage when $p = q = 1$. The results also show that coverage increases with sample size when $p = 2$, $q = 1$, but in fact coverages are well short of their nominal levels even with $n = 10,000$ with three dimensions. With four or six dimensions, coverages appear to tend toward zero as sample size increases. By contrast, methods (ii)–(iii) are seen to yield coverages that are qualitatively similar to those achieved with the VRS-DEA and CRS-DEA estimators. For the case where $(p + q) = 3$, coverages by method (ii) are better than those obtained with method (i) (i.e., closer to nominal coverages) in every case, again confirming the remarks in the last paragraph of Section 4. With the FDH estimator, coverages when $n = 100$ or 200 are smaller than with the DEA estimators, but the coverages improve as sample size increases. One might reasonably expect coverages in Table 3 to be smaller than corresponding coverages in Tables 1–2 due to the slower convergence rate of the FDH estimator.

8 Conclusions

Nonparametric estimators (DEA or FDH) of efficiency are widely used in production analysis. The statistical properties of estimators of individual efficiencies are well known and bootstrap techniques have been developed for making inference. This paper establishes asymptotic properties of statistics that are functions of these estimators. The main difficulty is that in such statistics, the efficiency estimators are evaluated at random data points, where some of them may fall near the boundary of the attainable set.

We first establish new results for the asymptotic moments (mean, variance and covariances) of the efficiency estimators evaluated at random data points. We then analyze a simple and useful statistic: the mean of the efficiency estimates over the sample points. We consider the FDH, the VRS-DEA, and the CRS-DEA cases.

Our results show that the usual central limit theorems are not applicable unless the dimension of the problem (i.e., the number of inputs and outputs) is exceptionally small. The problem comes mainly from the bias of the individual efficiency estimates, and the fact that this bias does not vanish at an appropriate rate, except in cases involving small dimensions. For the general case, we overcome this problem by using a mean computed over a subsample of data points; the subsample size is chosen to tune the bias and variance in order to obtain a stable, nondegenerate limiting distribution. We then propose a more general central limit theorem for DEA or FDH efficiency estimators.

In all cases, it is still necessary to remove the bias. This is accomplished using an estimator of the bias that allows construction of confidence intervals using normal quantiles, thereby avoiding computationally-burdensome bootstrap techniques. Monte-Carlo experiments confirm our theoretical results.

We show that these results can be extended to more sophisticated statistics, e.g., to OLS estimators that are sometimes used in the literature as part of a second-stage analysis to explain the variation in estimated efficiencies in terms of environmental factors, for cases where such two-stage analysis is appropriate. Future developments include extensions to testing problem, nonparametric second stage regressions, etc.

References

Afriat, S. (1972), Efficiency estimation of production functions, *International Economic Review* 13, 568–598.

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

Banker, R. D. and R. Natarajan (2008), Evaluating contextual variables affecting productivity using data envelopment analysis, *Operations Research* 56, 48–58.

Chambers, R. G., Y. Chung, and R. Färe (1996), Benefit and distance functions, *Journal of Economic Theory* 70, 407–419.

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

Deprins, D., L. Simar, and H. Tulkens (1984), Measuring labor inefficiency in post offices, in M. M. P. Pestieau and H. Tulkens, eds., *The Performance of Public Enterprises: Concepts and Measurements*, Amsterdam: North-Holland, pp. 243–267.

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

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

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

Gray, H. L. and R. R. Schucany (1972), *The Generalized Jackknife Statistic*, New York: Marcel Decker, Inc.

Jeong, S. O. (2004), Asymptotic distribution of DEA efficiency scores, *Journal of the Korean Statistical Society* 33, 449–458.

Jeong, S. O. and B. U. Park (2006), Large sample approximation of the distribution for convex-hull estimators of boundaries, *Scandinavian Journal of Statistics* 33, 139–151.

Jeong, S. O., B. U. Park, and L. Simar (2010), Nonparametric conditional efficiency measures: asymptotic properties, *Annals of Operational Research* 173, 105–122.

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

Kneip, A., L. Simar, and P. W. Wilson (2008), Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models, *Econometric Theory* 24, 1663–1697.

— (2011), A computationally efficient, consistent bootstrap for inference with non-parametric DEA estimators, *Computational Economics* 38, 483–515.

Marsaglia, G. (1972), Choosing a point from the surface of a sphere, *Annals of Mathematical Statistics* 43, 645–646.

Muller, M. E. (1959), A note on a method for generating points uniformly on n-dimensional spheres, *Communications of the Association for Computing Machinery* 2, 19–20.

Park, B. U., L. Simar, and C. Weiner (2000), FDH efficiency scores from a stochastic point of view, *Econometric Theory* 16, 855–877.

Park, B. U., L. Simar, and V. Zelenyuk (2008), Local likelihood estimation of truncated regression and its partial derivative: Theory and application, *Journal of Econometrics* 146, 185–2008.

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

Simar, L. and A. Vanhems (2012), Probabilistic characterization of directional distances and their robust versions, *Journal of Econometrics* 166, 342–354.

Simar, L., A. Vanhems, and P. W. Wilson (2012), Statistical inference for dea estimators of directional distances, *European Journal of Operational Research* 220, 853–864.

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

— (2007), Estimation and inference in two-stage, semi-parametric models of productive efficiency, *Journal of Econometrics* 136, 31–64.

— (2011a), Inference by the m out of n bootstrap in nonparametric frontier models, *Journal of Productivity Analysis* 36, 33–53.

— (2011b), Two-Stage DEA: Caveat emptor, *Journal of Productivity Analysis* 36, 205–218.

— (2013), Estimation and inference in nonparametric frontier models: Recent developments and perspectives, *Foundations and Trends in Econometrics* Forthcoming.

Wilson, P. W. (2011), Asymptotic properties of some non-parametric hyperbolic efficiency estimators, in I. Van Keilegom and P. W. Wilson, eds., *Exploring Research Frontiers in Contemporary Statistics and Econometrics*, Berlin: Springer-Verlag, pp. 115–150.

Table 1: Coverages of Estimated Confidence Intervals using VRS-DEA Efficiency Estimator

			(i)			(ii)			(iii)		
p	q	n	.90	.95	.99	.90	.95	.99	.90	.95	.99
1	1	100	0.8120	0.8740	0.9530	0.8120	0.8740	0.9530	0.8120	0.8740	0.9530
1	1	200	0.8360	0.9050	0.9750	0.8360	0.9050	0.9750	0.8360	0.9050	0.9750
1	1	500	0.8660	0.9230	0.9830	0.8660	0.9230	0.9830	0.8660	0.9230	0.9830
1	1	1000	0.8710	0.9280	0.9830	0.8710	0.9280	0.9830	0.8710	0.9280	0.9830
1	1	5000	0.8870	0.9470	0.9940	0.8870	0.9470	0.9940	0.8870	0.9470	0.9940
1	1	10000	0.9120	0.9570	0.9920	0.9120	0.9570	0.9920	0.9120	0.9570	0.9920
2	1	100	0.7300	0.8220	0.9130	0.7300	0.8220	0.9130	0.7300	0.8220	0.9130
2	1	200	0.7960	0.8570	0.9490	0.7960	0.8570	0.9490	0.7960	0.8570	0.9490
2	1	500	0.8510	0.9030	0.9650	0.8510	0.9030	0.9650	0.8510	0.9030	0.9650
2	1	1000	0.8620	0.9310	0.9790	0.8620	0.9310	0.9790	0.8620	0.9310	0.9790
2	1	5000	0.8930	0.9500	0.9890	0.8930	0.9500	0.9890	0.8930	0.9500	0.9890
2	1	10000	0.8940	0.9450	0.9910	0.8940	0.9450	0.9910	0.8940	0.9450	0.9910
2	2	100	0.4270	0.5040	0.6450	0.6120	0.7040	0.8490	0.6560	0.7660	0.9000
2	2	200	0.5500	0.6350	0.7650	0.7290	0.8190	0.9350	0.8110	0.8880	0.9710
2	2	500	0.6920	0.7820	0.9070	0.8350	0.9020	0.9640	0.9490	0.9800	0.9980
2	2	1000	0.7930	0.8680	0.9440	0.8680	0.9080	0.9800	0.9880	0.9970	1.0000
2	2	5000	0.8710	0.9330	0.9820	0.8940	0.9530	0.9930	1.0000	1.0000	1.0000
2	2	10000	0.8660	0.9270	0.9770	0.8860	0.9460	0.9810	1.0000	1.0000	1.0000
3	3	100	0.1990	0.2470	0.3500	0.5860	0.6870	0.8390	0.6490	0.7690	0.9130
3	3	200	0.3300	0.3960	0.5280	0.7340	0.8140	0.9360	0.9020	0.9580	0.9970
3	3	500	0.4960	0.5680	0.7000	0.8580	0.9090	0.9750	0.9970	1.0000	1.0000
3	3	1000	0.6460	0.7380	0.8530	0.8740	0.9380	0.9860	1.0000	1.0000	1.0000
3	3	5000	0.7270	0.8100	0.9170	0.8840	0.9480	0.9860	1.0000	1.0000	1.0000
3	3	10000	0.6640	0.7410	0.8690	0.8880	0.9440	0.9890	1.0000	1.0000	1.0000

Table 2: Coverages of Estimated Confidence Intervals using CRS-DEA Efficiency Estimator

			(i)			(ii)			(iii)		
p	q	n	.90	.95	.99	.90	.95	.99	.90	.95	.99
1	1	100	0.8980	0.9430	0.9860	0.8980	0.9430	0.9860	0.8980	0.9430	0.9860
1	1	200	0.8840	0.9340	0.9790	0.8840	0.9340	0.9790	0.8840	0.9340	0.9790
1	1	500	0.8980	0.9470	0.9880	0.8980	0.9470	0.9880	0.8980	0.9470	0.9880
1	1	1000	0.9010	0.9510	0.9900	0.9010	0.9510	0.9900	0.9010	0.9510	0.9900
1	1	5000	0.9050	0.9570	0.9930	0.9050	0.9570	0.9930	0.9050	0.9570	0.9930
1	1	10000	0.9070	0.9570	0.9920	0.9070	0.9570	0.9920	0.9070	0.9570	0.9920
2	1	100	0.8570	0.9080	0.9750	0.8570	0.9080	0.9750	0.8570	0.9080	0.9750
2	1	200	0.8650	0.9220	0.9780	0.8650	0.9220	0.9780	0.8650	0.9220	0.9780
2	1	500	0.8840	0.9380	0.9850	0.8840	0.9380	0.9850	0.8840	0.9380	0.9850
2	1	1000	0.8730	0.9280	0.9860	0.8730	0.9280	0.9860	0.8730	0.9280	0.9860
2	1	5000	0.9110	0.9500	0.9880	0.9110	0.9500	0.9880	0.9110	0.9500	0.9880
2	1	10000	0.9050	0.9570	0.9940	0.9050	0.9570	0.9940	0.9050	0.9570	0.9940
3	1	100	0.6590	0.7370	0.8590	0.6590	0.7370	0.8590	0.6590	0.7370	0.8590
3	1	200	0.7400	0.8160	0.9160	0.7400	0.8160	0.9160	0.7400	0.8160	0.9160
3	1	500	0.8160	0.8740	0.9440	0.8160	0.8740	0.9440	0.8160	0.8740	0.9440
3	1	1000	0.8280	0.8980	0.9690	0.8280	0.8980	0.9690	0.8280	0.8980	0.9690
3	1	5000	0.8690	0.9380	0.9900	0.8690	0.9380	0.9900	0.8690	0.9380	0.9900
3	1	10000	0.8930	0.9370	0.9900	0.8930	0.9370	0.9900	0.8930	0.9370	0.9900
5	1	100	0.1890	0.2350	0.3330	0.5110	0.6220	0.8090	0.5990	0.7070	0.8960
5	1	200	0.2120	0.2810	0.4140	0.5980	0.7100	0.8610	0.7200	0.8430	0.9650
5	1	500	0.2420	0.3170	0.4750	0.7140	0.8200	0.9320	0.9150	0.9680	1.0000
5	1	1000	0.2820	0.3640	0.5690	0.7990	0.8900	0.9650	0.9770	0.9980	1.0000
5	1	5000	0.3660	0.4750	0.6770	0.8460	0.9130	0.9760	0.9990	1.0000	1.0000
5	1	10000	0.5010	0.6150	0.7950	0.8760	0.9300	0.9800	1.0000	1.0000	1.0000

Table 3: Coverages of Estimated Confidence Intervals using FDH Efficiency Estimator

			(i)			(ii)			(iii)		
p	q	n	.90	.95	.99	.90	.95	.99	.90	.95	.99
1	1	100	0.3810	0.4710	0.6390	0.3810	0.4710	0.6390	0.3810	0.4710	0.6390
1	1	200	0.4740	0.5640	0.7390	0.4740	0.5640	0.7390	0.4740	0.5640	0.7390
1	1	500	0.5950	0.6950	0.8320	0.5950	0.6950	0.8320	0.5950	0.6950	0.8320
1	1	1000	0.6570	0.7550	0.8930	0.6570	0.7550	0.8930	0.6570	0.7550	0.8930
1	1	5000	0.8130	0.8930	0.9710	0.8130	0.8930	0.9710	0.8130	0.8930	0.9710
1	1	10000	0.8620	0.9200	0.9800	0.8620	0.9200	0.9800	0.8620	0.9200	0.9800
2	1	100	0.2200	0.2780	0.3710	0.5260	0.6290	0.7720	0.5540	0.6710	0.8580
2	1	200	0.2570	0.3090	0.4440	0.6450	0.7470	0.8690	0.7360	0.8490	0.9620
2	1	500	0.3190	0.3990	0.5510	0.7540	0.8460	0.9500	0.9340	0.9760	1.0000
2	1	1000	0.4260	0.5160	0.6710	0.8220	0.8980	0.9690	0.9940	0.9980	1.0000
2	1	5000	0.6320	0.7410	0.8720	0.8740	0.9380	0.9890	1.0000	1.0000	1.0000
2	1	10000	0.7200	0.8170	0.9220	0.8920	0.9420	0.9870	1.0000	1.0000	1.0000
2	2	100	0.0130	0.0160	0.0310	0.2020	0.2680	0.4270	0.1560	0.2260	0.4330
2	2	200	0.0120	0.0130	0.0250	0.2870	0.3810	0.5600	0.2090	0.3410	0.6490
2	2	500	0.0030	0.0040	0.0110	0.4540	0.5620	0.7690	0.4560	0.6990	0.9630
2	2	1000	0.0030	0.0040	0.0110	0.5760	0.7050	0.8970	0.8020	0.9640	1.0000
2	2	5000	0.0030	0.0030	0.0080	0.8000	0.8760	0.9610	1.0000	1.0000	1.0000
2	2	10000	0.0070	0.0100	0.0230	0.8520	0.9130	0.9750	1.0000	1.0000	1.0000
3	3	100	0.0050	0.0070	0.0100	0.1440	0.1840	0.2840	0.1060	0.1630	0.3020
3	3	200	0.0000	0.0010	0.0020	0.1970	0.2670	0.4310	0.1360	0.2360	0.4880
3	3	500	0.0010	0.0010	0.0020	0.3040	0.3990	0.6300	0.2090	0.4380	0.8480
3	3	1000	0.0000	0.0000	0.0000	0.4100	0.5210	0.7570	0.3850	0.7040	0.9890
3	3	5000	0.0000	0.0000	0.0000	0.5700	0.6840	0.8710	0.8600	0.9980	1.0000
3	3	10000	0.0000	0.0000	0.0000	0.7740	0.8780	0.9710	1.0000	1.0000	1.0000