The Annals of Statistics
2010, Vol. 38, No. 3, 1320–1340
DOI: 10.1214/09-AOS746

ASYMPTOTIC DISTRIBUTION OF CONICAL-HULL ESTIMATORS OF DIRECTIONAL EDGES

BY BYEONG U. PARK[1,3], SEOK-OH JEONG[2,3] AND LÉOPOLD SIMAR[3]

Seoul National University, Hankuk University of Foreign Studies and Université catholique de Louvain

Nonparametric data envelopment analysis (DEA) estimators have been widely applied in analysis of productive efficiency. Typically they are defined in terms of convex-hulls of the observed combinations of inputs $\times$ outputs in a sample of enterprises. The shape of the convex-hull relies on a hypothesis on the shape of the technology, defined as the boundary of the set of technically attainable points in the inputs $\times$ outputs space. So far, only the statistical properties of the smallest convex polyhedron enveloping the data points has been considered which corresponds to a situation where the technology presents variable returns-to-scale (VRS). This paper analyzes the case where the most common constant returns-to-scale (CRS) hypothesis is assumed. Here the DEA is defined as the smallest conical-hull with vertex at the origin enveloping the cloud of observed points. In this paper we determine the asymptotic properties of this estimator, showing that the rate of convergence is better than for the VRS estimator. We derive also its asymptotic sampling distribution with a practical way to simulate it. This allows to define a bias-corrected estimator and to build confidence intervals for the frontier. We compare in a simulated example the bias-corrected estimator with the original conical-hull estimator and show its superiority in terms of median squared error.

1. Introduction.

Consider a convex set Ψ in $\mathbb{R}_+^{p+1}$ which takes the form

$$\Psi = \{(\mathbf{x}, y) \in \mathbb{R}_+^{p+1} : 0 \le y \le g(\mathbf{x})\},$$

where g is a nonnegative convex function defined on $\mathbb{R}_+^p$ such that $g(a\mathbf{x}) = ag(\mathbf{x})$ for all $a > 0$. Suppose that we have a random sample $(\mathbf{X}_i, Y_i)$ drawn from a distribution which is supported on Ψ. In this paper, we are interested in estimating the "boundary" function g from the random sample. In particular, we study the

Received March 2009; revised July 2009.

[1]Supported by KOSEF funded by the Korea government (MEST 2009-0052815).

[2]Supported by a National Research Foundation of Korea Grant funded by the Korean Government (2009-0067024).

[3]Supported by the "Interuniversity Attraction Pole," Phase VI (No. P6/03) from the Belgian Science Policy.

AMS 2000 subject classifications. Primary 62G05; secondary 62G20.

Key words and phrases. Conical-hull, asymptotic distribution, efficiency, data envelopment analysis, DEA, constant returns-to-scale, CRS.

asymptotic distribution of the estimator

$$\hat{g}(\mathbf{x}) = \max\{y > 0 : (\mathbf{x}, y) \in \widehat{\Psi}\}, \tag{1}$$

where $\widehat{\Psi}$ is the convex-hull of the rays $\mathbf{R}_i \equiv \{(\gamma \mathbf{X}_i, \gamma Y_i) : \gamma \geq 0\}$ for all sample points $(\mathbf{X}_i, Y_i)$.

The problem arises in an area of econometrics where one is interested in evaluating the performance of an enterprise in terms of technical efficiency. In this context, $\mathbf{X}_i$ is the observed input vectors of the ith enterprise, Y_i is its observed productivity and Ψ is the production set of technically feasible pairs of input and output. The property that $g(a\mathbf{x}) = ag(\mathbf{x})$ for all $a > 0$, or, equivalently, $\Psi = a\Psi$ for all $a > 0$, is called "constant returns-to-scale" (CRS), and the commonly used estimator of Ψ in this case is the CRS-version of the data envelopment analysis (DEA) estimator defined by

$$\widehat{\Psi}_0 = \left\{(\mathbf{x}, y) \in \mathbb{R}_+^{p+1} : \mathbf{x} \geq \sum_{i=1}^n \gamma_i \mathbf{X}_i,\ y \leq \sum_{i=1}^n \gamma_i Y_i \text{ for some } \gamma_i \geq 0, i = 1, \dots, n\right\}.$$

In fact, $\widehat{\Psi}_0$ given above is nothing else than the smallest convex set containing all the rays $\mathbf{R}_i$ and the hyperplane $\{(\mathbf{x}, 0) : \mathbf{x} \in \mathbb{R}^p\}$. To see this, suppose that $(\mathbf{x}, y)$ belongs to $\widehat{\Psi}_0$. Then, there exist $\gamma_i \geq 0$ such that $\mathbf{x} \geq \sum_{i=1}^n \gamma_i \mathbf{X}_i$ and $y \leq \sum_{i=1}^n \gamma_i Y_i$. For these constants γ_i, define

$$\gamma_i^* = \gamma_i \left(\frac{y}{\sum_{j=1}^n \gamma_j Y_j}\right) \leq \gamma_i$$

for $1 \leq i \leq n$. Then $\sum_{i=1}^n \gamma_i^* Y_i = y$. Since $\mathbf{x} \geq \sum_{i=1}^n \gamma_i \mathbf{X}_i \geq \sum_{i=1}^n \gamma_i^* \mathbf{X}_i$, we have $\mathbf{x}^* \equiv \mathbf{x} - \sum_{i=1}^n \gamma_i^* \mathbf{X}_i \geq \mathbf{0}$. This shows $(\mathbf{x}, y) = \sum_{i=1}^n (\gamma_i^* \mathbf{X}_i, \gamma_i^* Y_i) + (\mathbf{x}^*, 0)$. The estimator $\hat{g}$ defined in (1) and the one based on $\widehat{\Psi}_0$ are identical with probability tending to one if the density of $(\mathbf{X}_i, Y_i)$ is bounded away from zero in a neighborhood of the boundary point $(\mathbf{x}, g(\mathbf{x}))$.

The problem that we describe in the first paragraph can be generalized to the case of vector-valued $\mathbf{y} \in \mathbb{R}^q$. This is particularly important in the specific problem that we mention in the above paragraph where productivity is typically measured in several variables. For this, we consider a conical-hull of a convex set A in $\mathbb{R}_+^{p+q}$ which is given by

$$\Psi \equiv \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}_+^{p+q} : \text{there exists a constant } a > 0 \text{ such that } (a\mathbf{x}, a\mathbf{y}) \in A\} \cup \{\mathbf{0}\}.$$

The set Ψ is convex and satisfies the CRS condition

$$a\Psi = \Psi \qquad \text{for all } a > 0. \tag{2}$$

We are interested in estimating the "directional edge" of Ψ in the $\mathbf{y}$-space, defined by

$$\lambda(\mathbf{x}, \mathbf{y}) = \sup\{\lambda > 0 : (\mathbf{x}, \lambda \mathbf{y}) \in \Psi\}$$

using a random sample from a density supported on Ψ. In the case where $q = 1$, the directional edge is linked directly to the boundary function g by the identity $g(\mathbf{x}) = y\lambda(\mathbf{x}, y)$. We consider the estimator

$$\hat{\lambda}(\mathbf{x}, \mathbf{y}) = \sup\{\lambda > 0 : (\mathbf{x}, \lambda\mathbf{y}) \in \widehat{\Psi}\}, \tag{3}$$

where $\widehat{\Psi}$ is the convex-hull of the rays $\mathbf{R}_i \equiv \{(\gamma\mathbf{X}_i, \gamma\mathbf{Y}_i) : \gamma \geq 0\}$ for all sample points $(\mathbf{X}_i, \mathbf{Y}_i)$.

To date, nonparametric data envelopment analysis (DEA) estimators have been discussed or applied in more than 1800 articles published in more than 400 journals [see Gattoufi, Oral and Reisman (2004) for a comprehensive bibliography]. DEA estimators are used to estimate various types of productive efficiency of firms in a wide variety of industries as well as governmental agencies, national economies and other decision-making units. The estimators employ linear programming methods, similar to the one appearing in (3), along the lines of Charnes, Cooper and Rhodes (1978) who popularized the basic ideas of Farrell (1957).

Typically these DEA estimators are indeed defined in terms of convex-hulls of the combinations of inputs $\times$ outputs $(\mathbf{X}_i, \mathbf{Y}_i)$ in a sample of firms. The shape of the convex-hull relies on a hypothesis on the shape of the technology defined as the boundary of the set Ψ of technically attainable points in the inputs $\times$ outputs space. So far, only the statistical properties of the smallest convex polyhedron enveloping the data points has been considered which corresponds to a situation where the technology presents variable returns-to-scale (VRS). Convergence results for DEA–VRS have been derived by Korostelev, Simar and Tsybakov (1995) in the case of univariate output and by Kneip, Park and Simar (1998) in the multivariate case. Asymptotic distribution of the DEA–VRS estimators was obtained in the bivariate case ($p = q = 1$) by Gijbels et al. (1999), for univariate output by Jeong and Park (2006) and for the full multivariate case by Jeong (2004) and Kneip, Simar and Wilson (2008).

VRS is a flexible assumption, but in many situations the economist assumes that the technology presents CRS: the first version of the DEA estimator derived by Farrell (1957) was for this situation. Here the DEA estimator $\widehat{\Psi}$ is defined, as above, after (3), as the smallest conical-hull with a vertex at the origin enveloping the cloud of observed points. The properties of this estimator have not been investigated, yet it was conjectured that one would gain some efficiency in the estimation by imposing the appropriate CRS structure to the estimator.

In this paper we determine the asymptotic properties of the DEA–CRS estimator defined in (3), showing that the rate of convergence is better than that of the VRS estimator. We derive also its asymptotic sampling distribution with a practical way to simulate it. This allows us to define a bias-corrected estimator and to build confidence intervals for the frontier. We compare, in a simulated example, the bias-corrected estimator with the original DEA–CRS estimator and show its superiority in terms of median squared error.

2. Rate of convergence. In this section we give the first theoretical result, the convergence rate of the estimator $\hat{\lambda}$, as defined in (3), in the general case of $p, q \geq 1$. Before presenting the result, we first give two lemmas which will be used in the proof of the first theorem.

LEMMA 1. *For any* $\alpha, \beta > 0$, *it holds that* $\lambda(\alpha\mathbf{x}, \beta\mathbf{y}) = \frac{\alpha}{\beta}\lambda(\mathbf{x}, \mathbf{y})$ *whenever* $(\alpha\mathbf{x}, \beta\mathbf{y}) \in \Psi$ *and* $(\mathbf{x}, \mathbf{y}) \in \Psi$. *The same identity holds for* $\hat{\lambda}$.

PROOF. The lemma follows from the CRS property (2) since

$$\sup\{\lambda > 0 \colon (\alpha\mathbf{x}, \lambda\beta\mathbf{y}) \in \Psi\} = \sup\left\{\lambda > 0 \colon \left(\mathbf{x}, \frac{\lambda\beta}{\alpha}\mathbf{y}\right) \in \Psi\right\}. \qquad \square$$

The following lemma is also derived from the convexity of Ψ and $\widehat{\Psi}$.

LEMMA 2. *For all* $r \in [0, 1]$ *and for all* $(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2) \in \Psi$,

$$\lambda[r(\mathbf{x}_1, \mathbf{y}_1) + (1 - r)(\mathbf{x}_2, \mathbf{y}_2)] \geq r\lambda(\mathbf{x}_1, \mathbf{y}_1) + (1 - r)\lambda(\mathbf{x}_2, \mathbf{y}_2).$$

The same inequality holds for $\hat{\lambda}$.

Our first theorem on the rate of convergence relies on the following assumptions. In what follows, we fix the point in Ψ where we want to estimate λ, and denote it by $(\mathbf{x}_0, \mathbf{y}_0)$. Throughout the paper, we assume that $(\mathbf{X}_i, \mathbf{Y}_i)$ are independent and identically distributed with a density f supported on $\Psi \subset \mathbb{R}_+^p \times \mathbb{R}_+^q$ and that $(\mathbf{x}_0, \mathbf{y}_0)$ is in the interior of Ψ.

(A1) $\lambda(\mathbf{x}, \mathbf{y})$ is twice partially continuously differentiable in a neighborhood of $(\mathbf{x}_0, \mathbf{y}_0)$.
(A2) The density f of $(\mathbf{X}, \mathbf{Y})$ on $\{(\mathbf{x}, \mathbf{y}) \in \Psi \colon \|(\mathbf{x}, \mathbf{y}) - (\mathbf{x}_0, \lambda(\mathbf{x}_0, \mathbf{y}_0)\mathbf{y}_0)\| \leq \varepsilon\}$ for some $\varepsilon > 0$ is bounded away from zero.

THEOREM 1. *Under the assumptions* (A1) *and* (A2), *it follows that* $\hat{\lambda}(\mathbf{x}_0, \mathbf{y}_0) - \lambda(\mathbf{x}_0, \mathbf{y}_0) = O_p(n^{-2/(p+q)})$.

PROOF. We apply the technique of Kneip, Park and Simar (1998). Put $B_p(\mathbf{t}, r) = \{\mathbf{x} \in \mathbb{R}_+^p \colon \|\mathbf{x} - \mathbf{t}\| \leq r\}$ and consider the balls near $\mathbf{x}_0 \colon C_r = B_p(\mathbf{x}_0^{(r)}, h/2)$, $r = 1, \ldots, 2p$ where $\mathbf{x}_0^{(2j-1)} = \mathbf{x}_0 - h\mathbf{e}_j$, $\mathbf{x}_0^{(2j)} = \mathbf{x}_0 + h\mathbf{e}_j$, $\mathbf{e}_j$ is the unit p-vector with the jth element equal to 1 for $j = 1, 2, \ldots, p$. Similarly, define $D_s = B_q(\mathbf{y}_0^{(s)}, h/2)$ for $s = 1, \ldots, 2q$. Take h small enough so that $C_r \times D_s \subset \Psi$ for all $r = 1, \ldots, 2p$ and $s = 1, \ldots, 2q$. For $r = 1, \ldots, 2p$, consider the conical hull of C_r,

$$\mathcal{C}_r = \{\mathbf{x} \in \mathbb{R}_+^p \colon \exists a > 0 \text{ such that } a\mathbf{x} \in C_r\}.$$

Similarly, define $\mathcal{D}_s$. Define

$$(\mathbf{U}_r, \mathbf{V}_s) = \operatorname*{arg\,min}_{(\mathbf{X}_i, \mathbf{Y}_i) \in \mathcal{C}_r \times \mathcal{D}_s} \lambda(\mathbf{X}_i, \mathbf{Y}_i).$$

Since the number of points in $\mathcal{X}_n$ falling into $\Psi \cap [\mathcal{C}_r \times \mathcal{D}_s]$ is proportional to nh^{p+q-2}, we have by assumption (A2),

$$\lambda(\mathbf{U}_r, \mathbf{V}_s) = 1 + O_p(n^{-1}h^{-p-q+2}), \qquad r = 1, \dots, 2p, s = 1, \dots, 2q. \tag{4}$$

Let $\mathbf{U}_r^* = \alpha_r \mathbf{U}_r$ and $\mathbf{V}_s^* = \beta_s V_s$ for $r = 1, \dots, 2p$ and $s = 1, \dots, 2q$ where α_r and β_s are positive constants such that $\mathbf{U}_r^* \in C_r$ and $\mathbf{V}_s^* \in D_s$. Then from Lemma 1, (4) and the fact that $\lambda, \hat{\lambda} \geq 1$, it holds that for $r = 1, \dots, 2p$ and $s = 1, \dots, 2q$,

$$\frac{\hat{\lambda}(\mathbf{U}_r^*, \mathbf{V}_s^*)}{\lambda(\mathbf{U}_r^*, \mathbf{V}_s^*)} = \frac{\hat{\lambda}(\mathbf{U}_r, \mathbf{V}_s)}{\lambda(\mathbf{U}_r, \mathbf{V}_s)} \geq \frac{1}{\lambda(\mathbf{U}_r, \mathbf{V}_s)} = 1 + O_p(n^{-1}h^{-p-q+2}),$$

which implies that $\hat{\lambda}(\mathbf{U}_r^*, \mathbf{V}_s^*) \geq \lambda(\mathbf{U}_r^*, \mathbf{V}_s^*) + O_p(n^{-1}h^{-p-q+2})$. Since C_r and D_s are balls surrounding the point $(\mathbf{x}_0, \mathbf{y}_0)$, there exist scalars $w_r \geq 0$ and $\omega_s \geq 0$ such that $\sum_{r=1}^{2p} w_r = 1$, $\sum_{s=1}^{2q} \omega_s = 1$, $\mathbf{x}_0 = \sum_{r=1}^{2p} w_r \mathbf{U}_r^*$ and $\mathbf{y}_0 = \sum_{s=1}^{2q} \omega_s \mathbf{V}_s^*$. Thus, from the assumption (A1) we have

$$\sum_{r=1}^{2p} \sum_{s=1}^{2q} w_r \omega_s \lambda(\mathbf{U}_r^*, \mathbf{V}_s^*) = \lambda(\mathbf{x}_0, \mathbf{y}_0) + O_p(h^2)$$

for all r and s. This, with Lemma 2 and the fact that $\lambda \geq \hat{\lambda}$, shows that

$$\begin{aligned} \lambda(\mathbf{x}_0, \mathbf{y}_0) \geq \hat{\lambda}(\mathbf{x}_0, \mathbf{y}_0) &\geq \sum_{r=1}^{2p} \sum_{s=1}^{2q} w_r \omega_s \hat{\lambda}(\mathbf{U}_r^*, \mathbf{V}_s^*) \\ &\geq \sum_{r=1}^{2p} \sum_{s=1}^{2q} w_r \omega_s \lambda(\mathbf{U}_r^*, \mathbf{V}_s^*) + O_p(n^{-1}h^{-p-q+2}) \\ &= \lambda(\mathbf{x}_0, \mathbf{y}_0) + O_p(h^2) + O_p(n^{-1}h^{-p-q+2}). \end{aligned}$$

Taking $h \sim n^{-1/(p+q)}$ completes the proof of the theorem. $\square$

REMARK 1. In the case where Ψ is a convex set in $\mathbb{R}^{p+q}$ without having the CRS property (2), the DEA (data envelopment analysis) estimator defined as in (3) with $\widehat{\Psi}$ replaced by the convex-hull of $(\mathbf{X}_i, \mathbf{Y}_i)$ is commonly used. In this case, the DEA estimator of $\lambda(\mathbf{x}_0, \mathbf{y}_0)$ is known to have $n^{-2/(p+q+1)}$ rate of convergence which is slightly worse than $n^{-2/(p+q)}$ [see Kneip, Park and Simar (1998)]. The CRS property reduces the "effective" dimension by one.

3. Asymptotic distribution. In this section we derive a representation for the asymptotic distribution of the estimator $\hat{\lambda}$ defined in (3). This representation enables one to simulate the asymptotic distribution so that one can correct the bias of the estimator to get an improved version of $\hat{\lambda}$. We work with the case where $q = 1$ first and then move to the general case where $q > 1$. The result for the case $q = 1$ is essential for the generalization to $q > 1$.

3.1. *The case where $q = 1$.* We consider the set

$$\Psi = \{(\mathbf{x}, y) \in A_c \times \mathbb{R}_+ : 0 \le y \le g(\mathbf{x})\},$$

where g is a nonnegative convex function defined on a conical-hull A_c of a convex set $A \subset \mathbb{R}_+^p$ such that

$$g(a\mathbf{x}) = a g(\mathbf{x}) \qquad \text{for all } a > 0, \tag{5}$$

and that, for all $\mathbf{x}_1, \mathbf{x}_2 \in A_c$ with $\mathbf{x}_1 \ne a\mathbf{x}_2$ for any $a > 0$,

$$g\big(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2\big) > \alpha g(\mathbf{x}_1) + (1-\alpha) g(\mathbf{x}_2) \tag{6}$$

for all $\alpha \in (0, 1)$. In this case, $\lambda(\mathbf{x}_0, y_0) = g(\mathbf{x}_0)/y_0$ so that the problem of estimating $\lambda(\mathbf{x}_0, y_0)$ reduces to that of estimating the function g at $\mathbf{x}_0$. The estimator of $g(\mathbf{x}_0)$ that corresponds to $\hat{\lambda}(\mathbf{x}_0, y_0)$ defined in (3) is given by

$$\hat{g}(\mathbf{x}_0) = y_0 \hat{\lambda}(\mathbf{x}_0, y_0) = \sup\{y : (\mathbf{x}_0, y) \in \widehat{\Psi}\}. \tag{7}$$

We note that the CRS condition (5) is satisfied, not only by linear functions of the form $g(\mathbf{x}) = \mathbf{c}^\top \mathbf{x}$, but also by those functions $g(\mathbf{x}) = c(x_1^r + \cdots + x_p^r)^{1/r}$ for all positive numbers c and positive integers r.

Define $\mathbf{S}_i$ by $\mathbf{S}_i^\top = (\mathbf{X}_i^\top, Y_i)$. Below we describe a canonical transformation T on Ψ such that the transformed data $T(\mathbf{S}_i)$ behave, asymptotically, as an i.i.d. sample from a uniform distribution on a region that can be represented by a simple $(p-1)$-dimensional quadratic function in the transformed space. The reduction of the dimension, by one, for the boundary function is due to the CRS property (5). This is consistent with the dimension reduction as we noted in Remark 1 in the previous section.

The key element in the derivation of the asymptotic distribution of $\hat{g}(\mathbf{x}_0)$ is to project the data $\mathbf{S}_i$ onto a hyperplane which is perpendicular to the vector $\mathbf{x}_0$ and passes through $\mathbf{x}_0$. The projected points lie under the locus of the function g on the hyperplane, and the estimator $\hat{g}(\mathbf{x}_0)$ equals the maximal y such that $(\mathbf{x}_0, y)$ belongs to the convex-hull of the projected points. The asymptotic distribution of the estimator $\hat{g}(\mathbf{x}_0)$ is then obtained by analyzing the statistical properties of the convex-hull of the projected points.

Let Q be a $p \times (p-1)$ matrix whose columns constitute an orthonormal basis for $\mathbf{x}_0^\perp$, the subspace of $\mathbb{R}^p$ that is perpendicular to the vector $\mathbf{x}_0$. Think of the transformation

$$T_1 : \mathbf{x} \mapsto \left(\frac{\mathbf{x}_0^\top \mathbf{x}}{\|\mathbf{x}_0\|}, \mathbf{x}^\top Q \right)^\top.$$

This transformation maps $\mathbf{x}$ to a vector which corresponds to $\mathbf{x}$ in the new coordinate system where the axes are $\mathbf{x}_0$ and the columns of Q. The first component of $T_1(\mathbf{x})$ is nothing other than the projection of $\mathbf{x}$ onto the space spanned by $\mathbf{x}_0$, and the vector of the rest components is its orthogonal complement in $\mathbb{R}^p$. Thus, the inverse transform T_1^{-1} is given by

$$T_1^{-1} : \mathbf{z} \mapsto z_1\left(\frac{\mathbf{x}_0}{\|\mathbf{x}_0\|}\right) + Q\mathbf{z}_2,$$

where $\mathbf{z}^\top = (z_1, \mathbf{z}_2^\top)$.

It would be more convenient to use a transformation that takes $\mathbf{x}_0$ to the origin in the new coordinate system. This can be done by the following transformation:

$$T_2 : \mathbf{x} \mapsto \left[\frac{\mathbf{x}_0^\top(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x}_0\|}, \left(\frac{\|\mathbf{x}_0\|^2}{\mathbf{x}_0^\top \mathbf{x}}\right)\mathbf{x}^\top Q\right]^\top.$$

Scaling by the factor $\|\mathbf{x}_0\|^2/\mathbf{x}_0^\top\mathbf{x}$ is introduced to factor out a common scalar for the inverse map of T_2. In fact, $\|\mathbf{x}_0\|^2/\mathbf{x}_0^\top\mathbf{x}$ equals the scalar c such that the projection of $c\mathbf{x}$ onto the linear span of $\mathbf{x}_0$ equals $\mathbf{x}_0$ itself. Thus

$$\frac{\|\mathbf{x}_0\|^2}{\mathbf{x}_0^\top\mathbf{x}}\mathbf{x} = \mathbf{x}_0 + Q\left(Q^\top \frac{\|\mathbf{x}_0\|^2}{\mathbf{x}_0^\top\mathbf{x}}\mathbf{x}\right)$$

so that the inverse transform of T_2 is given by

$$T_2^{-1} : \mathbf{z} \mapsto \left(\frac{z_1 + \|\mathbf{x}_0\|}{\|\mathbf{x}_0\|}\right)(\mathbf{x}_0 + Q\mathbf{z}_2).$$

Note that $\mathbf{x}_0^\top\mathbf{x} > 0$ if $\mathbf{x} \neq \mathbf{0}$ since then $\mathbf{x}_0, \mathbf{x} > \mathbf{0}$. It is easy to see that $T_2(\mathbf{x}_0) = \mathbf{0}$.

Define a $(p-1)$-dimensional function g^* by $g^*(\mathbf{z}_2) = g(\mathbf{x}_0 + Q\mathbf{z}_2)$. For a function ψ, let $\dot{\psi}$ and $\ddot{\psi}$ denote, respectively, the gradient vector and the Hessian matrix of ψ. Since, for any $\mathbf{u} \in \mathbb{R}^{p-1}$,

$$\mathbf{u}^\top \ddot{g}^*(\mathbf{z}_2)\mathbf{u} = (Q\mathbf{u})^\top \ddot{g}(\mathbf{x}_0 + Q\mathbf{z}_2)(Q\mathbf{u})$$

and also $(Q\mathbf{u})^\top(Q\mathbf{u}) = \mathbf{u}^\top\mathbf{u}$, it can be seen that g^* is convex if g is convex. In particular, (6) implies the strict convexity of g^*. Note that g^* does not have the CRS property (5), however.

Next, we introduce a further transformation on the new coordinate system $(\mathbf{z}, y)$. This transformation maps the equation $y = g^*(\mathbf{z}_2)$ to a perfect quadratic equation in the further transformed space. Since g^* is strictly convex, $-\ddot{g}^*(\mathbf{0})/2 = Q^\top(-\ddot{g}(\mathbf{x}_0)/2)Q$ is positive definite and symmetric. Thus, there exist an orthogonal matrix P and a diagonal matrix Λ such that $-\ddot{g}^*(\mathbf{0})/2 = P\Lambda P^\top$. The columns of P are the orthonormal eigenvectors, and the diagonal elements of Λ are the eigenvalues of the matrix $-\ddot{g}^*(\mathbf{0})/2$. Let T_3 be a transformation that maps $\mathbb{R}^p$ to $\mathbb{R}^p$ defined by

(8) $$T_3 : \mathbf{z} \mapsto \left(z_1, n^{1/(p+1)}\mathbf{z}_2^\top P \Lambda^{1/2}\right)^\top.$$

Note that this transformation does not change z_1, the first component of $\mathbf{z}$. Also, define a map $T_4 : \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}$ by

$$T_4 : (\mathbf{z}, y) \mapsto n^{2/(p+1)}\left[y\left(\frac{\|\mathbf{x}_0\|}{z_1 + \|\mathbf{x}_0\|}\right) - g^*(\mathbf{0}) - \dot{g}^*(\mathbf{0})^\top \mathbf{z}_2\right]. \tag{9}$$

The transformation we apply to the data $(\mathbf{X}_i, Y_i)$ is now defined by

$$T : (\mathbf{x}, y) \mapsto \big(T_3 \circ T_2(\mathbf{x}), T_4(T_2(\mathbf{x}), y)\big).$$

We explain how the equation $y = g(\mathbf{x})$ can be approximated, locally at $(\mathbf{x}_0, y_0)$, by a $(p-1)$-dimensional quadratic function in the new coordinate system transformed by T. Let $(\mathbf{v}, w) \in \mathbb{R}^p \times \mathbb{R}$ represent the new coordinate system obtained by the transformation T. Write $\mathbf{v}^\top = (v_1, \mathbf{v}_2^\top)$ with $\mathbf{v}_2$ being a $(p-1)$-dimensional vector. Then, the inverse transform of T maps $\mathbf{v}$ and w, respectively, to

$$\mathbf{x} = \left(\frac{v_1 + \|\mathbf{x}_0\|}{\|\mathbf{x}_0\|}\right)\big[\mathbf{x}_0 + n^{-1/(p+1)} Q P \Lambda^{-1/2} \mathbf{v}_2\big],$$
$$y = \left(\frac{v_1 + \|\mathbf{x}_0\|}{\|\mathbf{x}_0\|}\right)\big[g^*(\mathbf{0}) + n^{-1/(p+1)} \dot{g}^*(\mathbf{0})^\top P \Lambda^{-1/2} \mathbf{v}_2 + n^{-2/(p+1)} w\big].$$

Thus, for arbitrary compact sets $C_1 \subset \mathbb{R}^{p-1}$ and $C_2 \subset \mathbb{R}$, we obtain using the CRS property (5) that, uniformly for $v_1 \in \mathbb{R}_+$, $\mathbf{v}_2 \in C_1$ and $w \in C_2$,

$$\begin{aligned} & y = g(\mathbf{x}) \\ & \quad \leftrightarrow \quad g^*(\mathbf{0}) + n^{-1/(p+1)} \dot{g}^*(\mathbf{0})^\top P \Lambda^{-1/2} \mathbf{v}_2 + n^{-2/(p+1)} w \\ & \qquad\qquad = g^*\big(n^{-1/(p+1)} P \Lambda^{-1/2} \mathbf{v}_2\big) \\ & \quad \leftrightarrow \quad w = -\mathbf{v}_2^\top \mathbf{v}_2 + o(1) \end{aligned}$$

as n tends to infinity, provided that $\ddot{g}^*$ is continuous at $\mathbf{0}$.

Now we give a representation of the limit distribution of $\hat{g}$ as given in (7). Define

$$\theta = \|\mathbf{x}_0\| \int_0^\infty u^p f(u\mathbf{x}_0, ug(\mathbf{x}_0))\, du, \tag{10}$$

$$\kappa = \theta \det(\Lambda)^{-1/2}. \tag{11}$$

Define a set $R_n(\kappa) \subset \mathbb{R}^p$ of points $(\mathbf{v}_2, w)$ such that

$$\mathbf{v}_2 \in \big[-\tfrac{1}{2}\kappa^{-1/(p+1)} n^{1/(p+1)}, \tfrac{1}{2}\kappa^{-1/(p+1)} n^{1/(p+1)}\big]^{p-1},$$
$$w \in \big[-\mathbf{v}_2^\top \mathbf{v}_2 - \kappa^{-2/(p+1)} n^{2/(p+1)}, -\mathbf{v}_2^\top \mathbf{v}_2\big].$$

The volume of this set in $\mathbb{R}^p$ equals $n\kappa^{-1}$. Let $(\mathbf{V}_{2i}, W_i)$ be a random sample from the uniform distribution on $R_n(\kappa)$. This random sample can be generated once we know κ. Let $Z_n(\cdot)$ be defined as $\hat{g}$ in (7) with $\widehat{\Psi}$ being replaced by the convex-hull of $(\mathbf{V}_{2i}, W_i)$; that said,

$$Z_n(\mathbf{v}_2) = \sup\left\{\sum_{i=1}^n \gamma_i W_i : \mathbf{v}_2 = \sum_{i=1}^n \gamma_i \mathbf{V}_{2i}, \sum_{i=1}^n \gamma_i = 1, \gamma_i \geq 0, i = 1, \dots, n\right\}. \tag{12}$$

For a small $\varepsilon > 0$, define a set on $\mathbb{R}_+^{p+1}$ by

$$\begin{aligned} H_\varepsilon(\mathbf{x}_0) = \big\{\big(u(\mathbf{x}_0 + Q\mathbf{z}_2), u\big(g(\mathbf{x}_0 + Q\mathbf{z}_2) - y\big)\big) : u \geq 0, \\ \|\mathbf{z}_2\| \leq \varepsilon, 0 \leq y \leq \varepsilon\big\}. \end{aligned} \tag{13}$$

In the theorem below and those that follow, we will measure the distance between two distributions by the following modification of the Mallows distance:

$$d(\mu_1, \mu_2) = \inf_{Z_1, Z_2} \{E(Z_1 - Z_2)^2 \wedge 1 : \mathcal{L}(Z_1) = \mu_1, \mathcal{L}(Z_2) = \mu_2\}.$$

Convergence in this metric is equivalent to weak convergence.

THEOREM 2. *Assume* (A1) *and* (A2). *In addition, assume that* $-\ddot{g}^*$ *is positive definite and continuous at* $\mathbf{0}$ *and that the density* f *of* $(\mathbf{X}, Y)$ *is uniformly continuous on* $H_\varepsilon(\mathbf{x}_0)$ *for an arbitrarily small* $\varepsilon > 0$. *Let* L_{n1} *and* L_{n2} *denote the distributions of* $n^{2/(p+1)}[\hat{g}(\mathbf{x}_0) - g(\mathbf{x}_0)]$ *and* $Z_n(\mathbf{0})$, *respectively. Then*, $d(L_{n1}, L_{n2}) \to 0$ *as* n *tend to infinity.*

Computation of the distribution of Z_n solely depends on knowledge of κ. Thus one can approximate the distribution of $\hat{g}(\mathbf{x}_0)$ by estimating κ and then simulating Z_n with the estimated κ. The approximation enables one to correct the downward bias of $\hat{g}(\mathbf{x}_0)$ and get an improved estimator of $g(\mathbf{x}_0)$. Estimation of κ and bias-correction for $\hat{g}(\mathbf{x}_0)$ will be discussed in Section 4.

PROOF OF THEOREM 2. We first give a geometric description of the estimator $\hat{g}$. Consider a hyperplane in $\mathbb{R}^p$ defined by

$$\mathcal{P}(\mathbf{x}_0) = \{\mathbf{x} \in \mathbb{R}_+^p : \mathbf{x}_0^\top(\mathbf{x} - \mathbf{x}_0) = 0\}. \tag{14}$$

This hyperplane is perpendicular to the vector $\mathbf{x}_0$ and passes through $\mathbf{x}_0$. Let $\mathbf{P}_i$ be the point where the ray $\mathbf{R}_i$ meets the hyperplane $\mathcal{P}^\dagger(\mathbf{x}_0) \equiv \mathcal{P}(\mathbf{x}_0) \times \mathbb{R}_+$ in $\mathbb{R}^{p+1}$. It follows that

$$\mathbf{P}_i = \frac{\|\mathbf{x}_0\|^2}{\mathbf{x}_0^\top \mathbf{X}_i}(\mathbf{X}_i, Y_i). \tag{15}$$

Define $\widehat{\Psi}(\mathbf{x}_0)$ to be the convex-hull of the points $\mathbf{P}_i$. We claim that

$$\widehat{\Psi}(\mathbf{x}_0) = \mathcal{P}^\dagger(\mathbf{x}_0) \cap \widehat{\Psi}. \tag{16}$$

This means that $\widehat{\Psi}(\mathbf{x}_0)$ is a section of $\widehat{\Psi}$ obtained by cutting $\widehat{\Psi}$ by the hyperplane $\mathcal{P}^\dagger(\mathbf{x}_0)$. The fact that $\widehat{\Psi}(\mathbf{x}_0) \subset \mathcal{P}^\dagger(\mathbf{x}_0) \cap \widehat{\Psi}$ follows from convexity of $\mathcal{P}^\dagger(\mathbf{x}_0)$ and $\widehat{\Psi}$. The reverse inclusion also holds. To see this, let $(\mathbf{x}, y) \in \mathcal{P}^\dagger(\mathbf{x}_0) \cap \widehat{\Psi}$. Since $\widehat{\Psi}$ is the convex-hull of the rays $\mathbf{R}_i$, it follows that there exist $\gamma_i^* \ge 0$ such that $\mathbf{x} = \sum_{i=1}^n \gamma_i^* \mathbf{X}_i$ and $y = \sum_{i=1}^n \gamma_i^* Y_i$. Since $(\mathbf{x}, y) \in \mathcal{P}^\dagger(\mathbf{x}_0)$, we have

$$\sum_{i=1}^n \gamma_i^* \mathbf{x}_0^\top \mathbf{X}_i = \|\mathbf{x}_0\|^2. \tag{17}$$

Let $\xi_i = (\mathbf{x}_0^\top \mathbf{X}_i / \|\mathbf{x}_0\|^2)\gamma_i^* \ge 0$ for $1 \le i \le n$. By (17), $\sum_{i=1}^{n+1} \xi_i = 1$. By (15), we get $(\mathbf{x}, y) = \sum_{i=1}^n \xi_i \mathbf{P}_i$ which shows $(\mathbf{x}, y) \in \widehat{\Psi}(\mathbf{x}_0)$.

Since $\bigcup_{a \ge 0} a\mathcal{P}^\dagger(\mathbf{x}_0) = \mathbb{R}_+^{p+1}$, the CRS property of $\widehat{\Psi}$ and (16) thus yield

$$\widehat{\Psi} = \bigcup_{a \ge 0} a\widehat{\Psi}(\mathbf{x}_0) = \{(a\mathbf{x}, ay) : (\mathbf{x}, y) \in \widehat{\Psi}(\mathbf{x}_0), a \ge 0\}. \tag{18}$$

Recall the definition of $\hat{g}$ in (7). Also, note that, for $\mathbf{x} \in \mathcal{P}(\mathbf{x}_0)$, we have $(\mathbf{x}, y) \in \widehat{\Psi}$ if and only if $(\mathbf{x}, y) \in \widehat{\Psi}(\mathbf{x}_0)$. This follows from (18) and the fact that $a = 1$ is the only constant $a \ge 0$ such that $(\mathbf{x}, y) \in a\widehat{\Psi}(\mathbf{x}_0)$ if $\mathbf{x} \in \mathcal{P}(\mathbf{x}_0)$. This gives

$$\hat{g}(\mathbf{x}) = \sup\{y : (\mathbf{x}, y) \in \widehat{\Psi}(\mathbf{x}_0)\} \qquad \text{if } \mathbf{x} \in \mathcal{P}(\mathbf{x}_0). \tag{19}$$

See Figure 1 for an illustration in the case of $p = 2$ and $q = 1$.

Let Q be the matrix defined in the paragraph that contains the definition of the transformation T_1 early in this section. Since $\mathcal{P}(\mathbf{x}_0) = \{\mathbf{x}_0 + Q\mathbf{z}_2 \in \mathbb{R}_+^p : \mathbf{z}_2 \in$

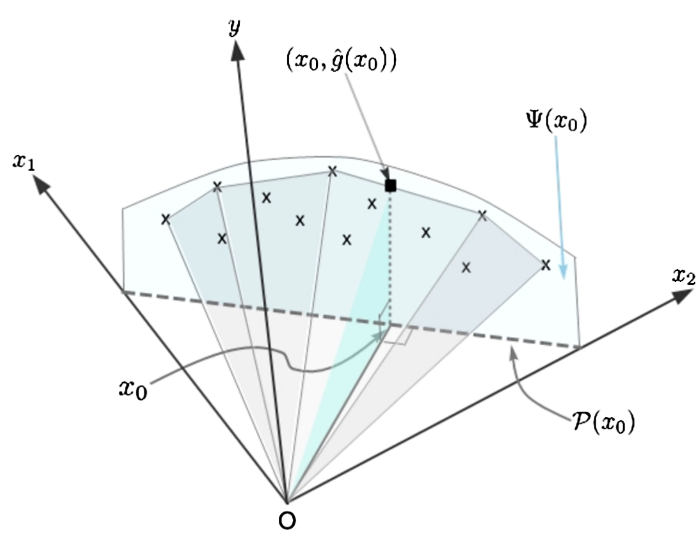

FIG. 1. *An illustration of $\mathcal{P}(\mathbf{x}_0)$, $\mathbf{P}_i$, $\hat{\Psi}$ and $\hat{g}$ in the case of $p = 2$ and $q = 1$. The crosses are the points $\mathbf{P}_i$, and the gray surface is the roof of the conical-hull estimator $\hat{\Psi}$.*

$\mathbb{R}^{p-1}\}$, the set,

$$\Psi(\mathbf{x}_0) \equiv \{(\mathbf{x}_0 + Q\mathbf{z}_2, y) \in A_c \times \mathbb{R}_+ : \mathbf{z}_2 \in \mathbb{R}^{p-1}, 0 \le y \le g(\mathbf{x}_0 + Q\mathbf{z}_2)\}, \tag{20}$$

equals the section of Ψ obtained by cutting Ψ by the hyperplane $\mathcal{P}^\dagger(\mathbf{x}_0)$; that is, $\Psi(\mathbf{x}_0) = \mathcal{P}^\dagger(\mathbf{x}_0) \cap \Psi$. In the new coordinate system

$$(\mathbf{z}, y') \equiv \big(T_2(\mathbf{x}), y\|\mathbf{x}_0\|^2/(\mathbf{x}_0^\top \mathbf{x})\big),$$

the set $\Psi(\mathbf{x}_0)$ in (20) can be represented by $\{0\} \times \Psi^*(\mathbf{x}_0)$ where

$$\Psi^*(\mathbf{x}_0) = \{(\mathbf{z}_2, y') : \mathbf{z}_2 \in \mathbb{R}^{p-1}(\mathbf{x}_0), 0 \le y' \le g^*(\mathbf{z}_2)\} \tag{21}$$

and $\mathbb{R}^{p-1}(\mathbf{x}_0)$ denote the set of $\mathbf{z}_2$ such that $\mathbf{x}_0 + Q\mathbf{z}_2 \in A_c$. Also, in that new coordinate system the points $\mathbf{P}_i$ defined in (15) correspond to $(0, \mathbf{P}_i^*)$ where $\mathbf{P}_i^* = (\mathbf{Z}_{2i}, Y_i')$, $\mathbf{Z}_{2i} = (\|\mathbf{x}_0\|^2/\mathbf{x}_0^\top \mathbf{X}_i) Q^\top \mathbf{X}_i$ and $Y_i' = (\|\mathbf{x}_0\|^2/\mathbf{x}_0^\top \mathbf{X}_i) Y_i$. Since convex-hulls are equivariant under linear transformations, this means that in the new coordinate system, $\widehat{\Psi}(\mathbf{x}_0)$ corresponds to $\{0\} \times \widehat{\Psi}^*(\mathbf{x}_0)$ where $\widehat{\Psi}^*(\mathbf{x}_0)$ is the convex-hull of the points $\mathbf{P}_i^*$. Now define

$$\hat{g}^*(\mathbf{z}_2) = \hat{g}(\mathbf{x}_0 + Q\mathbf{z}_2)$$

on $\mathbb{R}^{p-1}(\mathbf{x}_0)$. Since $(\mathbf{x}_0 + Q\mathbf{z}_2, y) \in \widehat{\Psi}(\mathbf{x}_0)$ is equivalent to $(\mathbf{z}_2, y) \in \widehat{\Psi}^*(\mathbf{x}_0)$, it follows from (19) that

$$\hat{g}^*(\mathbf{z}_2) = \sup\{y : (\mathbf{z}_2, y) \in \widehat{\Psi}^*(\mathbf{x}_0), \mathbf{z}_2 \in \mathbb{R}^{p-1}\}. \tag{22}$$

Let f denote the density of the original random vector $(\mathbf{X}, Y)$ and f^* denote the density of the transformed vector $(\mathbf{Z}_2, Y')$. The arguments in the preceding paragraph imply that the distribution of $\hat{g}(\mathbf{x}_0) - g(\mathbf{x}_0)$ equals that of $\hat{g}^*(\mathbf{0}) - g^*(\mathbf{0})$ where $\hat{g}^*$ is the convex-hull estimator of g^* constructed from a random sample of size n generated from the density f^*. Let $\kappa^* = \det(\Lambda)^{-1/2} f^*(\mathbf{0}, g^*(\mathbf{0}))$ where Λ is the diagonal matrix with its entries being the eigenvalues of $-\ddot{g}^*(\mathbf{0})/2$. Define Z_n^* as a version of $\hat{g}^*$ constructed from a random sample from the uniform distribution on $R_n(\kappa^*) \subset \mathbb{R}^p$ where R_n is defined immediately after (11). Then one can proceed as in the proof of Theorem 1 of Jeong and Park (2006) to show that the asymptotic distribution of $n^{2/(p+1)}(\hat{g}^*(\mathbf{0}) - g^*(\mathbf{0}))$ is identical to that of $Z_n^*(\mathbf{0})$ where one uses the transformations T_3^* and T_4^* defined by

$$T_3^* : \mathbf{z}_2 \mapsto n^{1/(p+1)} \Lambda^{1/2} P^\top \mathbf{z}_2,$$
$$T_4^* : (\mathbf{z}_2, y') \mapsto n^{2/(p+1)}\big(y' - g^*(\mathbf{0}) - \dot{g}^*(\mathbf{0})^\top \mathbf{z}_2\big).$$

Recalling the definitions of the transformations T_3 and T_4 in (8) and (9), respectively, $T_3^*(\mathbf{z}_2)$ equals $T_3(\mathbf{z})$ without the first component, where $\mathbf{z}^\top = (z_1, \mathbf{z}_2^\top)$, and $T_4^*(\mathbf{z}_2, y\|\mathbf{x}_0\|/(z_1 + \|\mathbf{x}_0\|)) = T_4(\mathbf{z}, y)$. Below, we prove that κ^* equals κ defined in (11) so that $Z_n^* = Z_n$ in distribution which concludes the proof of the theorem.

Let T^* denote the transformation that maps $(\mathbf{x}, y)$ to

$$(\mathbf{z}, y') = \big(T_2(\mathbf{x}), y\|\mathbf{x}_0\|^2/(\mathbf{x}_0^\top \mathbf{x})\big).$$

Let $c(z_1) = (z_1 + \|\mathbf{x}_0\|)/\|\mathbf{x}_0\|$. The Jacobian of the inverse transform of T^* equals

$$\begin{aligned} J(\mathbf{z}) &\equiv c(z_1)\det[\|\mathbf{x}_0\|^{-1}(\mathbf{x}_0 + Q\mathbf{z}_2), c(z_1)Q] \\ &= c(z_1)\det{}^{1/2}\begin{bmatrix} 1 + (\|\mathbf{z}_2\|/\|\mathbf{x}_0\|)^2 & (c(z_1)/\|\mathbf{x}_0\|)\mathbf{z}_2^\top \\ (c(z_1)/\|\mathbf{x}_0\|)\mathbf{z}_2 & c(z_1)^2 I_{p-1} \end{bmatrix}, \end{aligned}$$

where I_{p-1} denotes the identity matrix of dimension $(p-1)$. The second equality in the above calculation follows from the fact that the columns of Q are perpendicular to $\mathbf{x}_0$. Thus the joint density of $T^*(\mathbf{X}, Y)$ at the point $(\mathbf{z}, y')$ is given by $J(\mathbf{z}) f(c(z_1)(\mathbf{x}_0 + Q\mathbf{z}_2), c(z_1)y')$. The density $f^*(\mathbf{z}_2, y')$ is simply the marginalization of this joint density with respect to z_1 so that

$$f^*(\mathbf{z}_2, y') = \int_{-\|\mathbf{x}_0\|}^{\infty} J(\mathbf{z}) f\bigl(c(z_1)(\mathbf{x}_0 + Q\mathbf{z}_2), c(z_1)y'\bigr)\, dz_1.$$

Now, since $J(z_1, \mathbf{0}) = c(z_1)^p$, we obtain

$$\begin{aligned} f^*(\mathbf{0}, g^*(\mathbf{0})) &= \int_{-\|\mathbf{x}_0\|}^{\infty} c(z_1)^p f(c(z_1)\mathbf{x}_0, c(z_1)g^*(\mathbf{0}))\, dz_1 \\ &= \theta, \end{aligned}$$

where θ is defined in (10). $\square$

To see how well the distribution of $n^{2/(p+1)}\{\hat{g}(\mathbf{x}_0) - g(\mathbf{x}_0)\}$ is approximated by that of $Z_n(\mathbf{0})$, we took a Cobb–Douglas CRS production function $g(\mathbf{x}) = x_1^{0.4} \times x_2^{0.6}$ $(p = 2)$. We generated 5000 random samples of size $n = 100$ and 400 from $f(x_1, x_2, y) = \lambda x_1^{-0.4\lambda} x_2^{-0.6\lambda} y^{\lambda-1}$ supported on $\Psi = \{(x_1, x_2, y) : 0 \le x_1, x_2 \le 1, 0 \le y \le g(x_1, x_2)\}$. This yielded i.i.d. copies of (X_1, X_2, Y) with $X_1 \sim$ Uniform[0, 1], $X_2 \sim$ Uniform[0, 1] and $Y = g(X_1, X_2)e^{-V/\lambda}$ where $V \sim \mathrm{Exp}(1)$. Figures 2 and 3 depict the empirical distributions of $n^{2/(p+1)}\{\hat{g}(\mathbf{x}_0) - g(\mathbf{x}_0)\}$ and $Z_n(\mathbf{0})$ based on these samples in the case where $\lambda = 3$. The figures suggest that the approximation is fairly good for moderate sample sizes and get better as the sample size increases.

Theorem 2 excludes the case where g is linear; that is, $g(\mathbf{x}) = \mathbf{c}^\top\mathbf{x}$ for some vector $\mathbf{c}$. The latter case needs a different treatment. In the following theorem, we give the limit distribution in this case. To state the theorem, let $(\mathbf{V}_{2i}^L, W_i^L)$ be a random sample from the uniform distribution on the p-dimensional rectangle,

$$\begin{aligned} R_n^L(\theta) = {} & \bigl[-\tfrac{1}{2}\theta^{-1/(p+1)}n^{1/(p+1)}, \tfrac{1}{2}\theta^{-1/(p+1)}n^{1/(p+1)}\bigr]^{p-1} \\ & \times \bigl[-\theta^{-2/(p+1)}n^{2/(p+q)}, 0\bigr], \end{aligned} \tag{23}$$

where θ is defined in (10). The volume of this set in $\mathbb{R}^p$ equals $n\theta^{-1}$. Let $Z_n^L(\cdot)$ be a version of $Z_n(\cdot)$ constructed from $(\mathbf{V}_{2i}^L, W_i^L)$ replacing $(\mathbf{V}_{2i}, W_i)$.

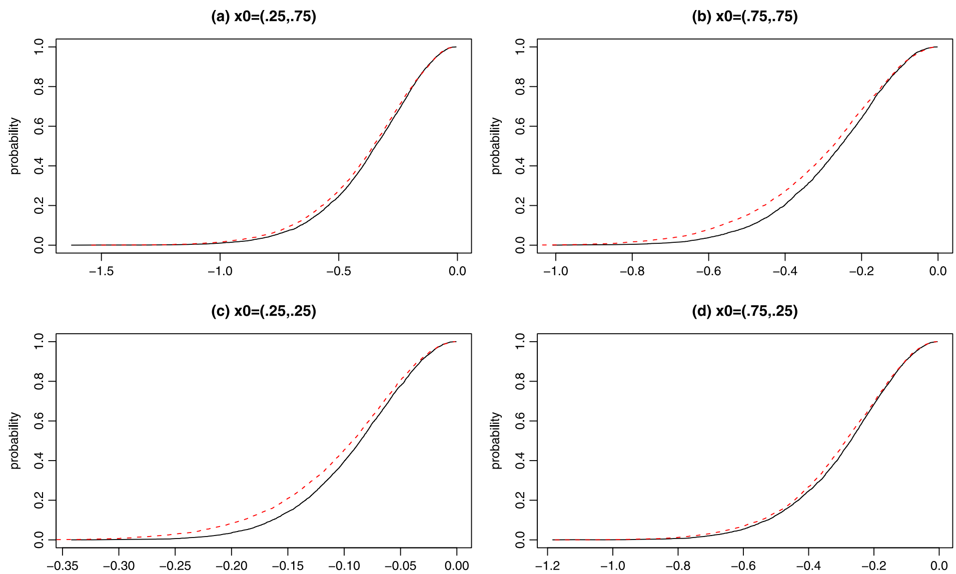

FIG. 2. *Solid curves are the empirical distribution functions of* $Z_n(0)$, *and the dotted curves are those of* $n^{2/(p+1)}\{\hat{g}(\mathbf{x}_0) - g(\mathbf{x}_0)\}$ *in the case where* $n = 100$ *and* $\lambda = 3$.

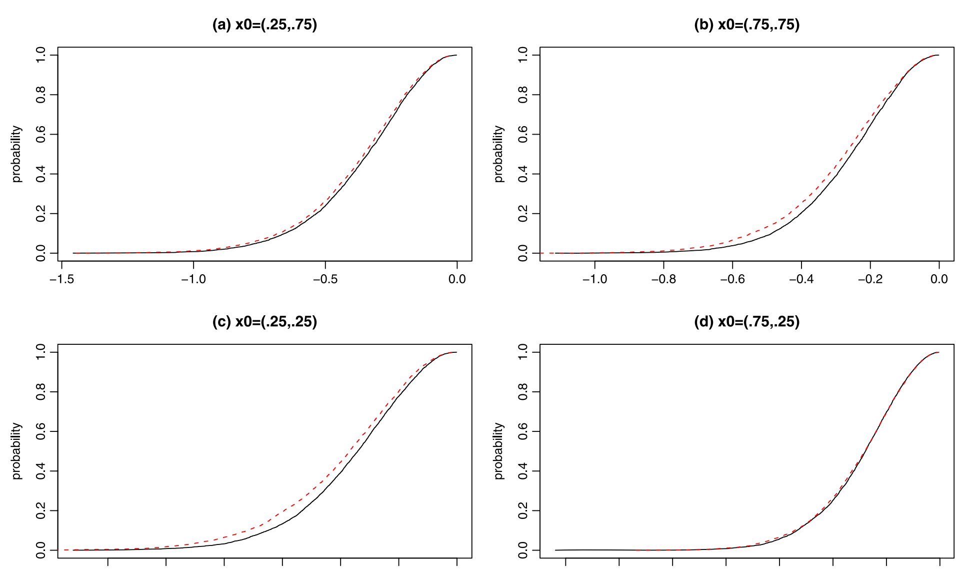

FIG. 3. *Solid curves are the empirical distribution functions of* $Z_n(0)$, *and the dotted curves are those of* $n^{2/(p+1)}\{\hat{g}(\mathbf{x}_0) - g(\mathbf{x}_0)\}$ *in the case where* $n = 400$ *and* $\lambda = 3$.

THEOREM 3. *Assume* (A1) *and* (A2). *Assume further that* $\Psi = \{(\mathbf{x}, y) \in \mathbb{R}_+^{p+1} : 0 \le y \le \mathbf{c}^\top \mathbf{x}\}$ *for some constant vector* $\mathbf{c} \neq \mathbf{0}$ *and that the density* f *of*

$(\mathbf{X}, Y)$ *is uniformly continuous on* $H_\varepsilon(\mathbf{x}_0)$ *for an arbitrarily small* $\varepsilon > 0$. *Let* L_{n1} *and* L'_{n2} *denote the distributions of* $n^{2/(p+1)}[\hat{g}(\mathbf{x}_0) - \mathbf{c}^\top \mathbf{x}_0]$ *and* $Z_n^L(\mathbf{0})$, *respectively. Then* $d(L_{n1}, L'_{n2}) \to 0$ *as* n *tends to infinity.*

PROOF. In this case we consider the following transformation:

$$T^L : (\mathbf{x}, y) \mapsto \big(T_3^L \circ T_2(\mathbf{x}), T_4^L(T_2(\mathbf{x}), y)\big), \tag{24}$$

where $T_3^L : \mathbf{z} \mapsto (z_1, n^{1/(p+1)}\mathbf{z}_2^\top)^\top$ and

$$T_4^L : (\mathbf{z}, y) \mapsto n^{2/(p+2)}\left(\frac{\|\mathbf{x}_0\|}{z_1 + \|\mathbf{x}_0\|} y - \mathbf{c}^\top \mathbf{x}_0 - \mathbf{c}^\top Q \mathbf{z}_2\right).$$

Let $(\mathbf{V}^L, W^L) = T^L(\mathbf{X}, Y)$. Then it can be shown as in the proof of Theorem 2 that the density of $(\mathbf{V}_2^L, W^L)$ is given by $n^{-1}\theta\{1 + o(1)\}$ uniformly for $\mathbf{v}_2^L$ and w^L in any compact sets of respective dimension. The rest of the proof is the same as that for Theorem 2. □

In the special case where $p = 1$, we can derive the limit distribution explicitly. In this case, the boundary function g is linear and takes the form $g(x) = cx$ for some constant $c > 0$. The transformation T^L in (24) reduces to

$$T^L(x, y) = \left(x - x_0, n\left(\frac{y}{x} x_0 - c x_0\right)\right).$$

The marginal density of W^L, where $(V^L, W^L) = T^L(X, Y)$, is approximated by the constant $n^{-1}\theta$ uniformly for w^L in any compact subset of $\mathbb{R}_-$ where θ in this case equals $x_0 \int_0^\infty u f(u x_0, u c x_0)\, du$. According to Theorem 3, the limit distribution of $n(\hat{g}(x_0) - g(x_0))$ equals the limit distribution of Z_n^L which is nothing else than $\max_{i=1}^n W_i^L$ in this simplest case where W_i^L are a random sample from the uniform distribution on $[-n\theta^{-1}, 0]$. Since $-\max_{i=1}^n W_i^L$ has the exponential distribution with mean θ^{-1} in the limit, we have

$$P\big[n\big(g(x_0) - \hat{g}(x_0)\big) \le w\big] \to 1 - \exp(-\theta w)$$

for all $w \ge 0$.

3.2. *The case where* $q > 1$. In this section we extend the results in the previous section to the case where $q > 1$ and Ψ is a conical-hull of a convex set $\mathcal{A}$ in $\mathbb{R}_+^{p+q}$. For this we make a canonical transformation on $\mathbf{y}$-space so that the problem for $q > 1$ is reduced to the case where $q = 1$. Again we fix the point $(\mathbf{x}_0, \mathbf{y}_0)$ where we want to estimate the function λ.

Let Γ be a $q \times (q-1)$ matrix whose columns form a basis for $\mathbf{y}_0^\perp$. Consider a transformation $\mathcal{T}$ that maps $\mathbf{y} \in \mathbb{R}_+^q$ to $(\mathbf{u}, \omega) \in \mathbb{R}^{q-1} \times \mathbb{R}_+$ where

$$\mathbf{u} = \Gamma^\top \mathbf{y}, \qquad \omega = \frac{\mathbf{y}_0^\top \mathbf{y}}{\|\mathbf{y}_0\|}. \tag{25}$$

Then, in the new coordinate system $(\mathbf{x}, \mathbf{u}, \omega)$, the set Ψ can be represented as

$$\Psi_{\mathcal{T}} = \left\{(\mathbf{x}, \mathbf{u}, \omega) \in \mathbb{R}_+^p \times \mathbb{R}^{q-1} \times \mathbb{R}_+ : \left(\mathbf{x}, \Gamma\mathbf{u} + \omega\frac{\mathbf{y}_0}{\|\mathbf{y}_0\|}\right) \in \Psi\right\}. \tag{26}$$

Define a $(p+q-1)$-dimensional function

$$g_{\mathcal{T}}(\mathbf{x}, \mathbf{u}) \equiv g_{\mathcal{T}}(\mathbf{x}, \mathbf{u}; \mathbf{y}_0) = \sup\left\{a > 0 : \left(\mathbf{x}, \Gamma\mathbf{u} + a\frac{\mathbf{y}_0}{\|\mathbf{y}_0\|}\right) \in \Psi\right\}.$$

This is a boundary function in the transformed space such that all points $(\mathbf{x}, \mathbf{u}, \omega)$ in $\Psi_{\mathcal{T}}$ lie below the surface represented by the equation $\omega = g(\mathbf{x}, \mathbf{u})$.

Convexity of the function $g_{\mathcal{T}}$ follows from the fact that, due to convexity of Ψ,

$$a_0 \in \left\{a > 0 : \left(\mathbf{x}, \Gamma\mathbf{u} + a\frac{\mathbf{y}_0}{\|\mathbf{y}_0\|}\right) \in \Psi\right\}$$

and

$$a_0' \in \left\{a' > 0 : \left(\mathbf{x}', \Gamma\mathbf{u}' + a'\frac{\mathbf{y}_0}{\|\mathbf{y}_0\|}\right) \in \Psi\right\},$$

together, imply

$$\begin{aligned}&\alpha a_0 + (1-\alpha)a_0' \\ &\quad \in \left\{a > 0 : \left(\alpha\mathbf{x} + (1-\alpha)\mathbf{x}', \Gamma\big(\alpha\mathbf{u} + (1-\alpha)\mathbf{u}'\big) + a\frac{\mathbf{y}_0}{\|\mathbf{y}_0\|}\right) \in \Psi\right\}.\end{aligned}$$

Also, it has the CRS property (5) since Ψ satisfies (2). Furthermore, since $(\mathbf{x}, \mathbf{y}) \in \Psi$ if and only if $(\mathbf{x}, \mathcal{T}(\mathbf{y})) \in \Psi_{\mathcal{T}}$, and $\mathcal{T}(\alpha\mathbf{y}_0) = (\mathbf{0}^\top, \alpha\|\mathbf{y}_0\|)^\top$ for all $\alpha > 0$, we obtain

$$\begin{aligned} g_{\mathcal{T}}(\mathbf{x}_0, \mathbf{0}) &= \sup\left\{a > 0 : \left(\mathbf{x}_0, a\frac{\mathbf{y}_0}{\|\mathbf{y}_0\|}\right) \in \Psi\right\} \\ &= \sup\{a > 0 : (\mathbf{x}_0, (\mathbf{0}, a)) \in \Psi_{\mathcal{T}}\} \\ &= \|\mathbf{y}_0\| \sup\{\lambda > 0 : (\mathbf{x}_0, (\mathbf{0}, \lambda\|\mathbf{y}_0\|)) \in \Psi_{\mathcal{T}}\} \\ &= \|\mathbf{y}_0\| \sup\{\lambda > 0 : (\mathbf{x}_0, \mathcal{T}(\lambda\mathbf{y}_0)) \in \Psi_{\mathcal{T}}\} \\ &= \|\mathbf{y}_0\|\lambda(\mathbf{x}_0, \mathbf{y}_0). \end{aligned} \tag{27}$$

Here and below, $\mathbf{0}$ denotes the $(q-1)$-dimensional zero vector. Thus the problem of estimating $\lambda(\mathbf{x}_0, \mathbf{y}_0)$ using $(\mathbf{X}_i, \mathbf{Y}_i)$ is reduced to that of estimating $g_{\mathcal{T}}(\mathbf{x}_0, \mathbf{0})$ in the transformed space using $(\mathbf{X}_i, \mathcal{T}(\mathbf{Y}_i))$.

We note that in the proof of Theorem 2 we use only convexity and the CRS property of g. Thus the theory we developed in the previous section is applicable to $g_{\mathcal{T}}$. Let $(\mathbf{U}_i, \Omega_i) = \mathcal{T}(\mathbf{Y}_i)$ where $\mathbf{U}_i$ is the vector of the first $(q-1)$ elements of $\mathcal{T}(\mathbf{Y}_i)$, and Ω_i is the scalar-valued random variable. The joint density of $(\mathbf{X}_i, \mathbf{U}_i, \Omega_i)$ at the point $(\mathbf{x}, \mathbf{u}, \omega)$ is given by

$$f_{\mathcal{T}}(\mathbf{x}, \mathbf{u}, \omega) = \det{}^{1/2}(\Gamma^\top \Gamma) f\left(\mathbf{x}, \Gamma \mathbf{u} + \omega \frac{\mathbf{y}_0}{\|\mathbf{y}_0\|}\right). \tag{28}$$

The constant θ defined in (10) that corresponds to the density $f_{\mathcal{T}}$ equals

$$\begin{aligned}\theta_{\mathcal{T}} &= \|(\mathbf{x}_0, \mathbf{0})\| \int_0^\infty u^{p+q-1} f_{\mathcal{T}}(u\mathbf{x}_0, \mathbf{0}, u g_{\mathcal{T}}(\mathbf{x}_0, \mathbf{0}))\, du \\ &= \det{}^{1/2}(\Gamma^\top \Gamma) \|\mathbf{x}_0\| \int_0^\infty u^{p+q-1} f\left(u\mathbf{x}_0, u g_{\mathcal{T}}(\mathbf{x}_0, \mathbf{0}) \frac{\mathbf{y}_0}{\|\mathbf{y}_0\|}\right) du \\ &= \det{}^{1/2}(\Gamma^\top \Gamma) \|\mathbf{x}_0\| \int_0^\infty u^{p+q-1} f(u\mathbf{x}_0, u\lambda(\mathbf{x}_0, \mathbf{y}_0)\mathbf{y}_0)\, du,\end{aligned}$$

where the last identity follows from (27). The determinant that corresponds to $\det(\Lambda)$ in the definition of κ in (11) is $\det(-Q_{\mathcal{T}}^\top \ddot{g}_{\mathcal{T}}(\mathbf{x}_0, \mathbf{0}) Q_{\mathcal{T}}/2)$ where $Q_{\mathcal{T}}$ is a $(p+q-1) \times (p+q-2)$ matrix whose columns form an orthonormal basis for $(\mathbf{x}_0, \mathbf{0})^\perp$. Thus we modify the definition of κ as

$$\kappa_{\mathcal{T}} = \theta_{\mathcal{T}} \det\bigl(-Q_{\mathcal{T}}^\top \ddot{g}_{\mathcal{T}}(\mathbf{x}_0, \mathbf{0}) Q_{\mathcal{T}}/2\bigr)^{-1/2}.$$

Recall that the construction of Z_n defined in (12) depends only on κ and p. Define $Z_{n,\mathcal{T}}$ as a version of Z_n with $\kappa_{\mathcal{T}}$ and $(p+q-1)$ replacing κ and p, respectively. Also, define a $(p+q-2)$-dimensional function $g_{\mathcal{T}}^*(\mathbf{z}_2) = g_{\mathcal{T}}((\mathbf{x}_0, \mathbf{0}) + Q_{\mathcal{T}} \mathbf{z}_2)$, and $H_{\varepsilon,\mathcal{T}}(\mathbf{x}_0, \mathbf{0})$ as $H_\varepsilon(\mathbf{x}_0)$ at (13) with $(p+q-1)$, $g_{\mathcal{T}}$, $(\mathbf{x}_0, \mathbf{0})$ and $Q_{\mathcal{T}}$ replacing p, g, $\mathbf{x}_0$ and Q, respectively. Then we have the following theorem for the limit distribution of $\hat{\lambda}(\mathbf{x}_0, \mathbf{y}_0)$ for arbitrary dimensions $p, q \geq 1$.

THEOREM 4. *Assume* (A1) *and* (A2). *In addition, assume that* $-\ddot{g}_{\mathcal{T}}^*$ *is positive definite and continuous at* $\mathbf{0}$, *and that the density* $f_{\mathcal{T}}$ *given at* (28) *is uniformly continuous on* $H_{\varepsilon,\mathcal{T}}(\mathbf{x}_0, \mathbf{0})$ *for an arbitrarily small* $\varepsilon > 0$. *Let* L_{n1} *and* L_{n2} *denote the distributions of* $n^{2/(p+q)}[\hat{\lambda}(\mathbf{x}_0, \mathbf{y}_0) - \lambda(\mathbf{x}_0, \mathbf{y}_0)]$ *and* $Z_{n,\mathcal{T}}(\mathbf{0}_{p+q-2})/\|\mathbf{y}_0\|$, *respectively. Then*, $d(L_{n1}, L_{n2}) \to 0$ *as* n *tends to infinity.*

Theorem 4 excludes the case where $\Psi = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}_+^{p+q} : \mathbf{c}_1^\top \mathbf{x} - \mathbf{c}_2^\top \mathbf{y} \geq 0\}$ for some constant vectors $\mathbf{c}_1, \mathbf{c}_2 > \mathbf{0}$. Below we treat this case. When $q = 1$, this corresponds to the case where the boundary function g is linear in $\mathbf{x}$.

Define

$$\mathbf{c}_{\mathcal{T}} = \frac{\|\mathbf{y}_0\|}{\mathbf{c}_2^\top \mathbf{y}_0} \begin{pmatrix} \mathbf{c}_1 \\ \Gamma^\top(-\mathbf{c}_2) \end{pmatrix}.$$

Then $\Psi_{\mathcal{T}}$ defined in (26) takes the form

$$\Psi_{\mathcal{T}} = \left\{(\mathbf{x}, \mathbf{u}, w) : 0 \leq w \leq \mathbf{c}_{\mathcal{T}}^\top \begin{pmatrix} \mathbf{x} \\ \mathbf{u} \end{pmatrix}\right\},$$

and it holds that

$$\mathbf{c}_{\mathcal{T}}^{\top}\begin{pmatrix}\mathbf{x}\\ \mathbf{0}\end{pmatrix}=\|\mathbf{y}_0\|\lambda(\mathbf{x}_0,\mathbf{y}_0).$$

Thus we can apply the arguments leading to Theorem 3 with p, $\mathbf{c}$, $\mathbf{x}_0$ and Q being replaced by $(p+q-1)$, $\mathbf{c}_{\mathcal{T}}$, $(\mathbf{x}_0,\mathbf{0})$ and $Q_{\mathcal{T}}$, respectively.

Let $R_{n,\mathcal{T}}^{L}(\theta_c T)$ be the rectangle defined in (23) with θ and p being replaced by $\theta_{\mathcal{T}}$ and $(p+q-1)$. Define $Z_{n,\mathcal{T}}^{L}$ as Z_n^L using a random sample from the uniform distribution of the $(p+q-1)$-dimensional rectangle $R_{n,\mathcal{T}}^{L}(\theta_{\mathcal{T}})$. By applying the proof of Theorem 3 to $\mathbf{c}_{\mathcal{T}}$ replacing $\mathbf{c}$, we get the following theorem.

THEOREM 5. *Assume* (A1) *and* (A2). *Assume further that* $\Psi=\{(\mathbf{x},\mathbf{y})\in\mathbb{R}_+^{p+q}:\mathbf{c}_1^{\top}\mathbf{x}-\mathbf{c}_2^{\top}\mathbf{y}\ge 0\}$ *for some constant vectors* $\mathbf{c}_1,\mathbf{c}_2>\mathbf{0}$ *and that the density* $f_{\mathcal{T}}$ *given at* (28) *is uniformly continuous on* $H_{\varepsilon,\mathcal{T}}(\mathbf{x}_0,\mathbf{0})$ *for an arbitrarily small* $\varepsilon>0$. *Let* L_{n1} *and* L'_{n2} *denote the distributions of* $n^{2/(p+q)}[\hat{\lambda}(\mathbf{x}_0,\mathbf{y}_0)-\lambda(\mathbf{x}_0,\mathbf{y}_0)]$ *and* $Z_{n,\mathcal{T}}^{L}(\mathbf{0}_{p+q-2})/\|\mathbf{y}_0\|$, *respectively. Then* $d(L_{n1},L'_{n2})\to 0$ *as n tends to infinity.*

4. Estimation of κ and $\kappa_{\mathcal{T}}$. We discuss how to estimate κ as defined in (11) for the case where $q=1$. It is straightforward to extend the methods to the case where $q>1$ via the canonical transformation that we introduced in Section 3.2.

Consider the set $H_\varepsilon(\mathbf{x}_0)\subset\mathbb{R}_+^{p+1}$ defined in (13). The projection of this set on the $\mathbf{x}$-space is a conical hull around the vector $\mathbf{x}_0$, and for each direction of the ray $\mathbf{x}_0+Q\mathbf{z}_2$, determined by $\mathbf{z}_2$, its section on that direction is also a conical hull of single dimension under the boundary g. For each fixed $u\ge 0$, let

$$\begin{aligned}H_\varepsilon(u;\mathbf{x}_0)=\{&(u(\mathbf{x}_0+Q\mathbf{z}_2),y):\|\mathbf{z}_2\|\le\varepsilon,\\ &g(u(\mathbf{x}_0+Q\mathbf{z}_2))-u\varepsilon\le y\le g(u(\mathbf{x}_0+Q\mathbf{z}_2))\}.\end{aligned}$$

This is a section of $H_\varepsilon(\mathbf{x}_0)$ obtained by cutting $H_\varepsilon(\mathbf{x}_0)$ perpendicular to $\mathbf{x}_0$ at the distance $u\|\mathbf{x}_0\|$ from the origin. Its volume in the cutting hyperplane $u\mathcal{P}^{\dagger}(\mathbf{x}_0)$, where $\mathcal{P}^{\dagger}(\mathbf{x}_0)$ is defined between (14) and (15), equals

$$v_\varepsilon(u)=c_{p-1}u^p\varepsilon^p,$$

where c_r denote the volume of the r-dimensional unit ball, that is, $c_r=\frac{\pi^{r/2}}{\Gamma(r/2+1)}$ with $\Gamma(z)=\int_0^\infty t^{z-1}e^{-t}\,dt$. Thus, as $\varepsilon\to 0$ we have

$$\begin{aligned}P[(\mathbf{X},Y)\in H_\varepsilon(\mathbf{x}_0)]&=\int_0^\infty\int_{(\mathbf{x},y)\in H_\varepsilon(u;\mathbf{x}_0)}f(\mathbf{x},y)\,d\mathbf{x}\,dy\,du\\ &=\int_0^\infty f(u\mathbf{x}_0,ug(\mathbf{x}_0))v_\varepsilon(u)\,du\,\{1+o(1)\}\\ &=c_{p-1}\varepsilon^p\int_0^\infty u^p f(u\mathbf{x}_0,ug(\mathbf{x}_0))\,du\,\{1+o(1)\}.\end{aligned}$$

This consideration motivates the following estimator of θ:

$$\hat{\theta} = \|\mathbf{x}_0\| c_{p-1}^{-1} n^{-1} \varepsilon^{-p} \sum_{i=1}^{n} I\big((\mathbf{X}_i, Y_i) \in \widehat{H}_\varepsilon(\mathbf{x}_0)\big), \tag{29}$$

where $\hat{H}_\varepsilon(\mathbf{x}_0)$ is the sample version of $H_\varepsilon(\mathbf{x}_0)$ with g replaced by $\hat{g}$ in its definition. Note that, for implementing $\hat{\theta}$, it is convenient to use the fact,

$$(\mathbf{X}_i, Y_i) \in \hat{H}_\varepsilon(\mathbf{x}_0) \quad \Leftrightarrow \quad \|\mathbf{Z}_{2i}\| \le \varepsilon, \qquad \hat{g}^*(\mathbf{Z}_{2i}) - \varepsilon \le Y_i' \le \hat{g}^*(\mathbf{Z}_{2i}).$$

It is straightforward to see that $\hat{\theta}$ is a consistent estimator of θ under the conditions of Theorem 2.

For estimating $\det(\Lambda)$, one can apply local polynomial fitting to $\{(\mathbf{Z}_{2i}, \hat{g}^*(\mathbf{Z}_{2i}))\}$. For a small $\delta > 0$, perform a second-order polynomial regression on the set of the points

$$\{(\mathbf{Z}_{2i}, \hat{g}^*(\mathbf{Z}_{2i})) : \|\mathbf{Z}_{2i}\| \le \delta, i = 1, 2, \dots, n\} \cup \{(\mathbf{0}, \hat{g}^*(\mathbf{0}))\},$$

to get

$$\breve{g}^*(\mathbf{z}) = \breve{g}_0 + \breve{\mathbf{g}}_1' \mathbf{z} + \mathbf{z}' \breve{\mathbf{g}}_2 \mathbf{z}. \tag{30}$$

Use $\det(\breve{\mathbf{g}}_2)$ as an estimator of $\det(\Lambda)$. An estimator of κ is then defined by $\hat{\kappa} = \hat{\theta} \det(\breve{\mathbf{g}}_2)^{-1/2}$.

Using the estimator of κ one can obtain a bias-corrected estimator of the function g^*. For this, one generates Z_n repeatedly as described at (12) using the estimated κ. Call them $Z_{n,1}, Z_{n,2}, \dots, Z_{n,B}$. A bias-corrected estimator is then defined by

$$\hat{g}^*(\mathbf{0}) - n^{-2/(p+1)} \bar{Z}_{n,\cdot}(\mathbf{0}),$$

where $\bar{Z}_{n,\cdot}(\mathbf{0}) = B^{-1} \sum_{b=1}^{B} Z_{n,b}(\mathbf{0})$. Also, a $100 \times (1 - \alpha)\%$ confidence interval is given by

$$\big[\hat{g}^*(\mathbf{0}) - n^{-2/(p+1)} Z_{n,(B(1-\alpha/2))}(\mathbf{0}), \hat{g}^*(\mathbf{0}) - n^{-2/(p+1)} Z_{n,(B\alpha/2)}(\mathbf{0})\big],$$

where $Z_{n,(j)}(\mathbf{0})$ are the ordered values $Z_{n,j}(\mathbf{0})$ such that $Z_{n,(1)}(\mathbf{0}) > Z_{n,(2)}(\mathbf{0}) > \cdots > Z_{n,(B)}(\mathbf{0})$.

5. Numerical study. In this section we investigate, by a Monte Carlo experiment, the behavior of the sampling distribution of the DEA–CRS estimator in finite samples. To be more specific we will compare if the bias-corrected estimator suggested above has better properties than the original DEA–CRS estimator in terms of median squared error.

For our Monte Carlo scenario, we adapted the scenario proposed in Kneip, Simar and Wilson (2008) to our setup. The efficient frontier is defined with a CRS generalized Cobb–Douglas production function,

$$Y_{1e} = X_1^{0.4} X_2^{0.6} \cos\omega,$$
$$Y_{2e} = X_1^{0.5} X_2^{0.5} \sin\omega,$$

TABLE 1
Ratio $\mathrm{R}_{\varepsilon,\delta}$ *of the median of the squared errors of the bias-corrected estimator over the median of the squared errors of the original DEA–CRS estimator*

($n = 100$)		**($n = 400$)**	
$\varepsilon = \delta$	**Ratio of median of squared errors**	$\varepsilon = \delta$	**Ratio of median of squared errors**
3.50	0.7123	3.25	0.6500
3.75	0.6863	3.50	0.6402
4.00	0.7264	3.75	0.6965
4.25	0.8081	4.00	0.7026
4.50	0.8213	4.25	0.7734

where the random rays are generated through $\omega \sim \text{Uniform}(\frac{1}{9}\frac{\pi}{2}, \frac{8}{9}\frac{\pi}{2})$ and the values of the inputs $\mathbf{X}$ by $(X_1, X_2) \sim \text{Uniform}[10, 20]^2$. Then inefficient firms are generated below the efficient frontier by

$$(Y_1, Y_2) = (Y_{1e}, Y_{2e})e^{-V/3} \qquad \text{where } V \sim \text{Exp}(1).$$

So we are in a situation with $p = q = 2$, and we will analyze the estimation of the efficiency score of the fixed point $\mathbf{x}_0 = (15, 15)$, $\mathbf{y}_0 = (10, 10)$. It is easy to see that the true value of the parameter to estimate is $\lambda_0 = \lambda(\mathbf{x}_0, \mathbf{y}_0) = 1.0607$. We analyze the cases $n = 100$ and $n = 400$.

We performed 500 Monte Carlo simulations and computed the squared errors of the original DEA–CRS estimator and of the bias-corrected estimator. Table 1 summarizes the results. It gives the ratios of the median of the squared error of the two estimators,

$$\mathrm{R}_{\varepsilon,\delta} = \frac{\text{med}\{(\tilde{\lambda}_{0,j} - \lambda_0)^2, j = 1, 2, \ldots, 500\}}{\text{med}\{(\hat{\lambda}_{0,j} - \lambda_0)^2, j = 1, 2, \ldots, 500\}},$$

where $\hat{\lambda}_{0,j}$ and $\tilde{\lambda}_{0,j}$ denote the original DEA–CRS estimate and the bias-corrected estimate computed in the jth Monte Carlo replication, respectively. Note that the bias-corrected estimator relies on the values of the smoothing parameters (ε, δ) which appear in the definitions (29) and (30), respectively.

It is observed from the table that the bias-correction works very well for a wide range of the smoothing parameters, even though the smoothing parameters were taken to be equal in the simulation study for saving computational costs. We see also that the performance of the bias-corrected estimator gets better when compared to the original DEA–CRS as the sample size increases.

6. Discussion. In this paper we developed the theoretical properties of the DEA estimator defined in (3) in the case where the support Ψ of the data $(\mathbf{X}_i, \mathbf{Y}_i)$

satisfies the CRS condition (2). The assumption of CRS may be tested. In fact, whether the underlying technology exhibits CRS or VRS is a crucial question in studying productive efficiency. The question has important economic implications. If the technology does not exhibit CRS, then some production units may be found to be either too large or too small. Using the estimator at (3) in the case where the true technology displays nonconstant returns to scale results in statistically inconsistent estimates of efficiency and seriously distorts measures of efficiency.

One way to test CRS against VRS is to use the test statistic defined as

$$\rho_n = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{\lambda}(\mathbf{X}_i, \mathbf{Y}_i)}{\hat{\lambda}_{\mathrm{VRS}}(\mathbf{X}_i, \mathbf{Y}_i)} - 1\right),$$

where $\hat{\lambda}_{\mathrm{VRS}}$ is a version of $\hat{\lambda}$ for the case of VRS defined as in (3) but with $\widehat{\Psi}$ replaced by the convex-hull of $\{(\mathbf{X}_i, \mathbf{Y}_i)\}_{i=1}^{n}$. By construction,

$$\hat{\lambda}(\mathbf{X}_i, \mathbf{Y}_i) \geq \hat{\lambda}_{\mathrm{VRS}}(\mathbf{X}_i, \mathbf{Y}_i) > 0$$

so that $\rho_n \geq 0$. A larger value of ρ_n gives a stronger evidence against the null hypothesis of CRS in favor of the alternative hypothesis of VRS. The test statistic was considered by Simar and Wilson (2002). One may compute p-values or critical values using a bootstrap method. For example, a subsampling scheme with the subsample size determined by the procedure described in Politis, Romano and Wolf (2001) might work for this problem. For testing CRS against nonconstant returns-to-scale, which is broader than VRS, one may use the estimators analyzed by Hall, Park and Stern (1998) and Park (2001) instead of $\hat{\lambda}_{\mathrm{VRS}}$. Theoretical and numerical properties of these testing procedures are yet to be developed.

REFERENCES

CHARNES, A., COOPER, W. W. and RHODES, E. (1978). Measuring the inefficiency of decision making units. *European J. Oper. Res.* **2** 429–444. MR0525905

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

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

GIJBELS, I., MAMMEN, E., PARK, B. U. and SIMAR, L. (1999). On estimation of monotone and concave frontier functions. *J. Amer. Statist. Assoc.* **94** 220–228. MR1689226

HALL, P., PARK, B. U. and STERN, S. (1998). On polynomial estimators of frontiers and boundaries. *J. Multivariate Anal.* **66** 71–98. MR1648521

JEONG, S.-O. (2004). Asymptotic distribution of DEA efficiency scores. *J. Korean Statist. Soc.* **33** 449–458. MR2126372

JEONG, S.-O. and PARK, B. U. (2006). Large sample approximation of the distribution for convex-hull estimators of boundaries. *Scand. J. Statist.* **33** 139–151. MR2255114

KNEIP, A., PARK, B. U. and SIMAR, L. (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory* **14** 783–793. MR1666696

KNEIP, A., SIMAR, L. and WILSON, P. W. (2008). Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models. *Econometric Theory* **24** 1663–1697. MR2456542

KOROSTELEV, A., SIMAR, L. and TSYBAKOV, A. (1995). On estimation of monotone and convex boundaries. *Publ. Inst. Statist. Univ. Paris* **39** 3–18. MR1744393

PARK, B. U. (2001). On nonparametric estimation of data edges. *J. Korean Statist. Soc.* **30** 265–280. MR1892209

POLITIS, D. N., ROMANO, J. P. and WOLF, M. (2001). On the asymptotic theory of subsampling. *Statist. Sinica* **11** 1105–1124. MR1867334

SIMAR, L. and WILSON, P. W. (2002). Nonparametric test of return to scale. *European J. Oper. Res.* **139** 115–132. MR1888265

B. U. PARK
DEPARTMENT OF STATISTICS
SEOUL NATIONAL UNIVERSITY
SOUTH KOREA
E-MAIL: bupark@stats.snu.ac.kr

S.-O. JEONG
DEPARTMENT OF STATISTICS
HANKUK UNIVERSITY OF FOREIGN STUDIES
SOUTH KOREA
E-MAIL: seokohj@hufs.ac.kr

L. SIMAR
INSTITUT DE STATISTIQUE
UNIVERSITÉ CATHOLIQUE DE LOUVAIN
BELGIUM
E-MAIL: leopold.simar@uclouvain.be