Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 143 (2008) 317–333
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

A smooth nonparametric conditional quantile frontier estimator

Carlos Martins-Filho[a,*], Feng Yao[b]

[a]Department of Economics, Oregon State University, Ballard Hall 303, Corvallis, OR 97331-3612, USA
[b]Department of Economics, Chinese University of Hong Kong, Esther Lee Building, Shatin, Hong Kong

Received 18 March 2006; received in revised form 19 October 2007; accepted 29 October 2007
Available online 4 November 2007

Abstract

Traditional estimators for nonparametric frontier models (DEA, FDH) are very sensitive to extreme values/outliers. Recently, Aragon et al. [2005. Nonparametric frontier estimation: a conditional quantile-based approach. Econometric Theory 21, 358–389] proposed a nonparametric α-frontier model and estimator based on a suitably defined conditional quantile which is more robust to extreme values/outliers. Their estimator is based on a nonsmooth empirical conditional distribution. In this paper, we propose a new smooth nonparametric conditional quantile estimator for the α-frontier model. Our estimator is a kernel based conditional quantile estimator that builds on early work of Azzalini [1981. A note on the estimation of a distribution function and quantiles by a kernel method. Biometrika 68, 326–328]. It is computationally simple, resistant to outliers and extreme values, and smooth. In addition, the estimator is shown to be consistent and $\sqrt{n}$ asymptotically normal under mild regularity conditions. We also show that our estimator's variance is smaller than that of the estimator proposed by Aragon et al. A simulation study confirms the asymptotic theory predictions and contrasts our estimator with that of Aragon et al.

JEL classification: C14; C22

Keywords: Conditional quantile estimation; Nonparametric frontier; Production function

1. Introduction

The specification and estimation of production frontiers, and the measurement of the associated efficiency level of production units has been the subject of a vast and expanding literature since the seminal work of Farrell (1957). The main objective of this literature can be stated simply. Consider $(X, Y) \in \Re^d_+ \times \Re_+$ where Y describes the output of a production unit and X describes the d inputs used in production. The output set is given by $\Psi = \{(x, y) \in \Re^d_+ \times \Re_+ : x \text{ can produce } y\}$ and the production function or frontier associated with Ψ is $g(x) = \sup\{y \in \Re_+ : (x, y) \in \Psi\}$ for all $x \in \Re^d_+$. Let $(x_0, y_0) \in \Psi$ characterize the performance of a production unit and define $0 \leqslant R_0 \equiv \frac{y_0}{g(x_0)} \leqslant 1$ to be this unit's (inverse) Farrell output efficiency measure.

*Corresponding author. Tel.: +1 541 737 1476.
E-mail addresses: carlos.martins@oregonstate.edu (C. Martins-Filho), feng.yao@mail.business.und.edu (F. Yao).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2007.10.004

The main objective in production and efficiency analysis is, given a random sample of production units $\chi_n \equiv \{(X_i, Y_i)\}_{i=1}^n$ that share the set Ψ, to obtain estimates of $g(\cdot)$ and by extension $R_i = \frac{Y_i}{g(X_i)}$ for $i = 1, \ldots, n$.

Deterministic frontier models and estimators, largely represented by data envelopment analysis (DEA) and full disposal hull (FDH), have gained popularity among applied researchers because their construction relies on very mild assumptions on Ψ.[1] These models are based on the assumption that χ_n lie in Ψ, i.e., $P((X, Y) \in \Psi) = 1$, where P is the probability measure associated with the random vector (X, Y). The most appealing characteristic of such models is that there is no need to assume any restrictive parametric structure on $g(\cdot)$ or the probability measure P to perform estimation. In addition to accommodating a flexible nonparametric structure, the appeal of DEA and FDH estimators has increased since Gijbels et al. (1999) and Park et al. (2000) obtained their asymptotic distributions under some fairly reasonable assumptions.

DEA and FDH type estimators have two serious deficiencies. First, since they are based on the idea of enveloping the observed data, these estimators are very sensitive to outliers or extreme observations and are inherently biased. Second, even in cases where the production technology induces a smooth production frontier, estimated frontiers based on FDH and DEA are discontinuous or piecewise linear function, respectively. Efforts to remedy such deficiencies have appeared in different nonparametric frontier modeling contexts (Girard and Jacob, 2004; Hall et al., 1998; Knight, 2001; Martins-Filho and Yao, 2007a). Prominent among these recent developments is the contribution of Aragon et al. (2005). They propose an alternative definition for the production function,

$$h(x) = \sup\{y \in \Re_+ : F(y/x) < 1\} \equiv \inf\{y \in \Re+ : F(y/x) = 1\}, \tag{1}$$

where $F(y/x) = \frac{F(x,y)}{F_X(x)}$, $F(x, y) = P(\{(X, Y) : X \leqslant x, Y \leqslant y\})$ and $F_X(x)$ is the associated marginal distribution of X. Since $F_X(x) > 0$, they restrict attention to $\Psi^* = \{(x, y) \in \Psi : F_X(x) > 0\}$. If the frontier $g(x)$ is monotone nondecreasing, a typical assumption in economic theory, then $h(x) = g(x)$ for all x such that $(x, y) \in \Psi^*$. Note that the assumption that $g(x)$ is monotone nondecreasing is equivalent to $F(y/x)$ being monotone nonincreasing on the set $\{x \in \Re_+^d : F_X(x) > 0\}$.[2] Aragon et al. observed that $g(x)$ is the order one quantile for the conditional distribution of Y given that $X \leqslant x$, where the inequality should be understood componentwise, and therefore $g(x) \equiv q_1(x) = \inf\{y \in \Re_+ : F(y/x) = 1\}$. As a natural extension, they suggest the concept of a production function of continuous order $\alpha \in [0, 1]$ given by

$$q_\alpha(x) = \inf\{y \in \Re_+ : F(y/x) \geqslant \alpha\}. \tag{2}$$

The usefulness of this concept rests in the fact that if $F(\cdot/x)$ is strictly increasing on the support $[0, g(x)]$, then $q_\alpha(x) = F^{-1}(\alpha/x)$ where $F^{-1}(\cdot/x)$ is the inverse of $F(\cdot/x)$. In this context, any production plan $(x, y) \in \Psi^*$ belongs to some α-order conditional quantile curve, and is such that y represents an output level that is greater than 100α percent of the output of all production plans using inputs X such that $X \leqslant x$. Thus, rather than relying on $g(X_i)$ to define production efficiency of firm i, the conditional quantile function $q_\alpha(X_i)$ compares the production plan (X_i, Y_i) of firm i to all other $\{(X_j, Y_j)\}_{j \neq i}$ such that $X_j \leqslant X_i$.

Aragon et al. propose an estimator for $q_\alpha(x)$ that is based on a conditional empirical quantile obtained from inverting the empirical conditional distribution function $F_n(y/x)$. Although their estimator has desirable properties of consistency and $\sqrt{n}$ asymptotic normality, it is well known from the unconditional distribution and quantile estimation literature (Azzalini, 1981; Falk, 1985; Yang, 1985; Bowman et al., 1998) that smoothing beyond that given by the empirical distribution can produce significant gains in finite samples. Li and Racine (2005) have proposed a kernel based nonparametric conditional distribution estimator and an associated conditional quantile estimator; however, their conditioning set is $X = x$ rather than $X \leqslant x$. In this paper, we propose a smooth nonparametric kernel estimator for the α-frontier ($q_\alpha(x)$). Our estimator is an extension of the seminal idea of Nadaraya (1964) and is based on a smooth estimator of the conditional distribution $F(y/x)$. Besides having the properties of consistency and $\sqrt{n}$-asymptotic normality, the variance of our estimator is smaller than that of the estimator proposed by Aragon et al., confirming that the gains first identified by Azzalini in unconditional quantile estimation extend to conditional quantile estimation. Our simulations also confirm the superior performance of our proposed estimator.

[1] See Simar and Wilson (2006) for a review of deterministic frontiers and illustrations of their widespread use.

[2] See Aragon et al. (2005, Proposition 2.5).

Besides this introduction, this paper has five additional sections. Section 2 describes the stochastic model in detail, contrasts its assumptions with those in the past literature and describes the estimation procedure. Section 3 provides the main theorems establishing the asymptotic behavior of our estimator and discusses bandwidth selection. Section 4 contains a Monte Carlo study that implements the estimator, sheds some light on its finite sample properties and compares its performance with that of the estimator proposed by Aragon et al. Section 5 provides an empirical illustration of our estimation procedure using data on electric utilities from the United States. Lastly, section 6 provides a summary and some directions for future work.

2. Stochastic model and estimation

2.1. α frontier estimator

Consider $\chi_n = \{(X_i, Y_i)\}_{i=1}^n$ a sequence of independent random vectors taking values in Ψ^* and having the same distribution F as the vector (X, Y). Throughout the paper, X will represent a d-vector of inputs used in the production process and Y will represent a scalar measure of output. F is taken to be absolutely continuous with associated density function given by f. The marginal distribution and density functions of X are denoted by F_X and f_X, respectively. Given that our interest is on the estimation of the α-frontier, which coincides with conditional quantile $q_\alpha(x)$ for $\alpha \in [0, 1]$, we define an estimator $\hat{F}(y/x)$ for $F(y/x)$ as

$$\hat{F}(y/x) = \begin{cases} 0 & \text{if } y = 0, \\ \dfrac{\hat{F}(x, y)}{\hat{F}(x)} & \text{if } y > 0, \end{cases} \tag{3}$$

where $\hat{F}(x, y) = (nh_n)^{-1}\sum_{i=1}^n \int_0^y K(\frac{Y_i - \gamma}{h_n})\, d\gamma I(X_i \leqslant x)$ and $\hat{F}(x) = n^{-1}\sum_{i=1}^n I(X_i \leqslant x)$, $I(A)$ is the indicator function for the set A, $K(\cdot)$ is a suitably defined kernel function and h_n is a nonstochastic sequence of bandwidths such that $0 < h_n \to 0$ as $n \to \infty$. The estimator is different from that proposed by Aragon et al. in that their estimator for $F(x, y)$ is given by $F_n(x, y) = n^{-1}\sum_{i=1} I(X_i \leqslant x, Y_i \leqslant y)$. In essence, rather than estimating $F(y/x)$ by the empirical distribution of the data such that $X_i \leqslant x$ for $i = 1, \ldots, n$, we estimate $F(y/x)$ by integrating a smooth Rosenblatt density estimator constructed using the observations $\{(X_i, Y_i)\}_{i \in \{i: X_i \leqslant x\}}$. It is easy to demonstrate that $\hat{F}(y/x)$ is asymptotically a distribution function, i.e., for suitably defined kernels: (a) $\hat{F}(y/x)$ is nondecreasing in y; (b) $\hat{F}(y/x)$ is right continuous in $\Re_+$; (c) $\lim_{y\to 0}\hat{F}(y/x) = 0$; and (d) there exists $N(x)$ such that for all $n > N(x)$ we have $\lim_{y\to\infty}\hat{F}(y/x) = 1$.

Assuming that $q_\alpha(x)$ is the unique α order quantile for the conditional distribution $F(y/x)$, we define the estimator $q_{\alpha,n}(x)$ as the root of

$$\hat{F}(q_{\alpha,n}(x)/x) = \alpha \quad \text{for } \alpha \in (0, 1] \text{ and } x \in \Re_+^d. \tag{4}$$

Using the mean value theorem, absolute continuity of F and smoothness of the kernel function we can write $q_{\alpha,n}(x) - q_\alpha(x) = \frac{F(q_\alpha(x)/x) - \hat{F}(q_\alpha(x)/x)}{\hat{f}(\bar{q}_{\alpha,n}(x)/x)}$ where $\hat{f}(y/x) = \frac{\partial \hat{F}(y/x)}{\partial y} = \frac{(nh_n)^{-1}\sum_{i=1}^n K(\frac{Y_i - y}{h_n}) I(X_i \leqslant x)}{\hat{F}(x)}$ for $y \geqslant 0$ ($\hat{f}(y/x) = 0$ for $y < 0$) and $\bar{q}_{\alpha,n}(x) = \lambda q_{\alpha,n}(x) + (1 - \lambda) q_\alpha(x)$ for $\lambda \in (0, 1)$.

2.2. Assumptions

The stochastic properties of the estimator defined in (4) are obtained under the following regularity conditions:

Assumption A1. (a) $\chi_n = \{(X_i, Y_i)\}_{i=1}^n$ is a sequence of independent random vectors taking values in Ψ^* and having the same distribution F as the vector (X, Y), with support in Ψ^*; (b) Ψ^* is compact and $0 < f(x, y) < B_f$ for all $(x, y) \in \Psi^*$.

The assumption that χ_n is an independent and identically distributed sequence, and the existence of the density f as a bounded function in Ψ is standard in the deterministic frontier literature (Aragon et al., 2005; Cazals et al., 2002; Gijbels et al., 1999; Martins-Filho and Yao, 2007a; Park et al., 2000).

Assumption A2. (a) $K(\gamma): S_K \to \Re$ is a symmetric bounded function with compact support $S_K = [-B_K, B_K]$ such that: (b) $\int_{-B_K}^{B_K} K(\gamma)\,d\gamma = 1$; (c) $\int_{-B_K}^{B_K} \gamma K(\gamma)\,d\gamma = 0$, $\int_{-B_K}^{B_K} \gamma^2 K(\gamma)\,d\gamma = \sigma_K^2$; (d) for all $\gamma, \gamma' \in S_K$ we have $|K(\gamma) - K(\gamma')| \leqslant m_K |\gamma - \gamma'|$ for some $0 < m_K < \infty$; (e) for all $\gamma, \gamma' \in \Re$ we have $|\kappa(\gamma) - \kappa(\gamma')| \leqslant m_\kappa |\gamma - \gamma'|$ for some $0 < m_\kappa < \infty$, where $\kappa(\lambda) = \int_{-B_K}^{\lambda} K(\gamma)\,d\gamma$.

Assumption A2 is standard in nonparametric estimation and is satisfied by commonly used kernels such as Biweight, Epanechnikov and others.

Assumption A3. (a) f is continuous in Ψ^*; (b) for all x such that $F_X(x) > 0$ and for all $\alpha \in (0, 1]$, $f(q_\alpha(x)/x) > 0$, where $f(\cdot/x)$ is the derivative of $F(\cdot/x)$; (c) for all $(x, y), (x, y') \in \Psi^*$, $|f(x, y') - f(x, y)| \leqslant m_f |y - y'|$ for some $0 < m_f < \infty$; (d) F is twice continuously differentiable in the interior of Ψ^*.

A3(b) is assumed by Aragon et al. (2005), and the Lipschitz condition in A3(c) is also assumed by Park et al. (2000).

Assumption A4. For all $y, y' \in G$, where G is a compact subset of $(0, \infty)$, we have $|\int_{g^{-1}([y,y'])} dX| \leqslant m_{g^{-1}} |y - y'|$ for some $0 < m_{g^{-1}} < \infty$. Here, let $x = (x_1, \ldots, x_d)'$, then for any two sets $A \subseteq C_x = \times_{i=1}^{d} [0, x_i]$ and $B \subseteq [0, g(x)]$, $g(A) = \{g(x) : x \in A\}$ and $g^{-1}(B) = \{x : x \in C_x,\ g(x) \in B\}$.

Assumption A4 imposes a Lipschitz type condition on the inverse image g^{-1} of g. Note, for example, that if $g: \Re_+ \to \Re_+$ is bijective with inverse g^{-1}, Assumption A4 is equivalent to $|g^{-1}(y) - g^{-1}(y')| \leqslant m_{g^{-1}} |y - y'|$ for some $0 < m_{g^{-1}} < \infty$.

3. Asymptotic characterization of the estimator

3.1. Asymptotic properties

Theorems 1 and 2 establish consistency and asymptotic normality of $q_{\alpha,n}(x)$. The theorems depend on two auxiliary lemmas provided in the Appendix. Lemma 1 is an extension to the multivariate case of the second order results of Azzalini (1981), where the nonparametric distribution function estimator for $F(x, y)$ is given by $\hat{F}(x, y)$. Asymptotically, the difference between $\hat{F}(x, y)$ and the multivariate empirical distribution function estimator is the order at which the bias and variance converge to zero. Lemma 2 establishes conditions under which $\hat{F}(x, y)$ converges uniformly to $F(x, y)$, a necessary condition for Theorem 1. In Lemma 2 the assumption that $\min_{\{i: X_i \leqslant x\}} Y_i \geqslant h_n B_K$ implies that even as the number of observations that satisfy $\{i : X_i \leqslant x\}$ grows to infinity, the associated output levels Y_i are bounded away from zero. Although reasonable in most contexts, it is certainly an assumption that could be violated by certain data generating processes (DGPs). The lemmas, Theorems 1 and 2 and proofs can be found in Appendix A.[3] We now state.

Theorem 1. *Let* $0 < h_n \to 0$ *be a nonstochastic sequence of bandwidths with* $nh_n \to \infty$ *as* $n \to \infty$. *Assume* A1–A4 *and that for given* $x \in \Re_+^d$ *and some* $N(x)$ *we have that for all* $n > N(x)$ $\min_{\{i: X_i \leqslant x\}} Y_i \geqslant h_n B_K$. *Then,*

$$q_{\alpha,n}(x) - q_\alpha(x) = o_p(1). \tag{5}$$

Asymptotic normality of $q_{\alpha,n}(x)$ under suitable normalization is obtained in Theorem 2.

Theorem 2. *Let* $0 < h_n \to 0$ *be a nonstochastic sequence of bandwidths with* $nh_n^2 \to \infty$ *and* $nh_n^4 = O(1)$ *as* $n \to \infty$. *Assume* A1–A4 *and that for given* $x \in \Re_+^d$ *and some* $N(x)$ *we have that for all* $n > N(x) \min_{\{i: X_i \leqslant x\}} Y_i \geqslant h_n B_K$. *Then, for all* $\alpha \in (0, 1)$ *we have*

$$v_n(x)^{-1} \sqrt{n}(q_{\alpha,n}(x) - q_\alpha(x) - B_n(x)) \xrightarrow{d} N(0, 1), \tag{6}$$

[3] Additional proofs and technical details can be found in Martins-Filho and Yao (2007b).

where $B_n(x) = -\frac{1}{2}h_n^2\sigma_K^2 \frac{\int_{g^{-1}([q_\alpha(x),g(x)])} f^{(1)}(\gamma,q_\alpha(x))\,\mathrm{d}\gamma}{F_X(x)f(q_\alpha(x)/x)} + \mathrm{o}(h_n^2)$ *and* $v_n^2(x) = \frac{1}{(F_X(x)f(q_\alpha(x)/x))^2}(F(x,q_\alpha(x)) - \frac{F^2(x,q_\alpha(x))}{F_X(x)} - 2h_n\sigma_\kappa \int_{g^{-1}([q_\alpha(x),g(x)])} f(\gamma,q_\alpha(x))\,\mathrm{d}\gamma) + \mathrm{o}(h_n)$ *with* $\kappa(x) = \int_{-B_K}^{x} K(\gamma)\,\mathrm{d}\gamma$, $0<\sigma_\kappa = \int_{-B_K}^{B_K} \gamma\kappa(\gamma)\,K(\gamma)\,\mathrm{d}\gamma$, *and* $f^{(1)}(X,y)$ *denotes the first derivative of f with respect to Y.*

The conditional quantile estimator proposed by Aragon et al. (2005) is also consistent and $\sqrt{n}$ asymptotically normal under similar assumptions; however, there are some important differences between the estimators. First, we observe that although our estimator depends on kernel smoothing, and therefore a bandwidth h_n is necessary in constructing the estimator, there is no asymptotic cost as the rate of convergence to normality occurs at the parametric rate $\sqrt{n}$. Hence, the number of inputs d has no impact on the convergence rate of the estimator. Most importantly, even though there is smoothing in $\hat{F}(y/x)$, it produces no slowing on the convergence in distribution, a result obtained by Falk (1985) and Hansen (2004a) in the context of unconditional distribution functions. Second, although the extra smoothing we propose might impose modest computational costs compared to the estimator proposed by Aragon et al., Theorem 2 reveals that the extra smoothness produces a smaller variance due to the higher order terms. Note that the variance of the asymptotic distribution of their estimator is given by

$$\frac{\alpha(1-\alpha)}{f^2(q_\alpha(x)/x)F_X(x)} \equiv \frac{1}{(F_X(x)f(q_\alpha(x)/x))^2}\left(F(x,q_\alpha(x)) - \frac{F^2(x,q_\alpha(x))}{F_X(x)}\right),$$

and given that the extra term that appears in v_n^2 is nonnegative, the variance of our estimator is smaller for all n finite. Third, the extra smoothing we propose does introduce a bias term $B_n(x) = \mathrm{O}(h_n^2)$, but provided that $nh_n^4 = \mathrm{o}(1)$ the bias vanishes asymptotically. We note that this condition is consistent with the conditions on h_n necessary to obtain Theorem 2. Finally, we observe that given that $B_n(x) = \mathrm{O}(h_n^2)$ and that the variance is of order $\mathrm{O}(n^{-1} + h_n n^{-1})$ the optimal bandwidth rate for minimization of the asymptotic mean integrated squared error is $h_n \propto n^{-1/3}$.

The next theorem provides the joint asymptotic distribution of $q_{\alpha,n}(x^1), q_{\alpha,n}(x^2), \ldots, q_{\alpha,n}(x^r)$ which can be used to construct joint asymptotic confidence sets for the α-frontier for various levels of input usage. The result is similar to that in Theorem 4.2. in Aragon et al. (2005).

Theorem 3. *Let* $x^1, x^2, \ldots, x^r$ *be r levels of input X and let all assumptions in Theorem* 2 *hold. Then, for* $\alpha \in (0,1)$ *we have*

$$\sqrt{n}(q_{\alpha,n}(x^1) - q_\alpha(x^1) - B(x^1), q_{\alpha,n}(x^2) - q_\alpha(x^2) - B(x^2), \ldots, q_{\alpha,n}(x^r) - q_\alpha(x^r) - B(x^r))' \xrightarrow{d} \mathrm{N}(0,Q),$$

where $B(x^l) = -\frac{1}{f(q_\alpha(x^l)/x^l)F_X(x^l)}\sigma_K^2 \frac{h_n^2}{2}\int_{g^{-1}([q_\alpha(x^l),g(x^l)])} f^{(1)}(X,q_\alpha(x^l))\,\mathrm{d}X + \mathrm{o}(h_n^2)$, $l \in \{1,2,\ldots,r\}$ *and Q is an* $r\times r$ *matrix with* (l,m)*th element* $Q_{l,m}$ *given by*

(1) $Q_{l,l} = \frac{\alpha(1-\alpha)}{f^2(q_\alpha(x^l)/x^l)F_X(x^l)}$ *if* $l = m$,
(2) $Q_{l,m} = \frac{1}{f(q_\alpha(x^l)/x^l)F_X(x^l)f(q_\alpha(x^m)/x^m)F_X(x^m)}[F(x^{lm},q_\alpha(x^l))(1-\alpha) - \alpha F(x^{lm},q_\alpha(x^m)) + \alpha^2 F_X(x^{lm})]$ *if* $l \neq m$, *and* $q_\alpha(x^l) \leqslant q_\alpha(x^m)$,
(3) $Q_{l,m} = \frac{1}{f(q_\alpha(x^l)/x^l)F_X(x^l)f(q_\alpha(x^m)/x^m)F_X(x^m)}[F(x^{lm},q_\alpha(x^m))(1-\alpha) - \alpha F(x^{lm},q_\alpha(x^l)) + \alpha^2 F_X(x^{lm})]$ *if* $l \neq m$, *and* $q_\alpha(x^l) \geqslant q_\alpha(x^m)$, where $x^{lm} = \{\min(x_1^l,x_1^m), \min(x_2^l,x_2^m)\ldots,\min(x_d^l,x_d^m)\}'$.

As is typical in applied work, for inference purposes, the unknown higher order components of the variance terms in Theorems 2 and 3 must be estimated via consistent nonparametric estimators. $f(q_\alpha(x)/x)$ can be estimated by $\hat{f}(q_{\alpha,n}(x)/x)$ the conditional Rosenblatt density estimator, using the *rule-of-thumb* bandwidth of Silverman (1986). Note the consistency of $\hat{f}(q_{\alpha,n}(x)/x)$ has been established in the proof of Theorem 2. Furthermore, $F_X(x)$ can be consistently estimated by $\hat{F}(x) = n^{-1}\sum_{i=1}^n I(X_i \leqslant x)$, and $F(x,q_\alpha(x))$ can be consistently estimated by $\hat{F}(x,q_{\alpha,n}(x))$.

In the following theorem we turn our attention to the estimation of the true frontier $q_1(x)$.

Theorem 4. *Assume that* $\min_{\{i:X_i\leqslant x\}} Y_i \geqslant h_n B_K$, *and that* A1, A2 *hold with* Ψ^* *compact. In addition, assume that the density* f *is strictly positive on the frontier* $\{(x, g(x)) : F_X(x) > 0\}$, *and that* $g(x)$ *is continuously differentiable. Then for any* x *in the interior of the support of* X *we have that*

(a) *There exists* $N(x)>0$ *such that for all* $n>N(x)$, $q_{1,n}(x) = \max_{\{i:X_i\leqslant x\}} Y_i + h_n B_K$.
(b) $n^{1/(d+1)}(q_1(x) - q_{1,n}(x) + h_n B_K) \xrightarrow{d}$ *Weibull*$(\mu_x^{d+1}, d+1)$.

μ_x is a constant depending on the slope of $g(\cdot)$ and the value of f at the frontier. Park et al. (2000) provide the exact expression for μ_x as well as a consistent estimator for μ_x. We note that by their Theorem 3.3, it is a direct consequence of the assumptions in Theorem 4 that

$$\mathrm{E}(q_1(x) - q_{1,n}(x)) = \Gamma\left(\frac{d+2}{d+1}\right)\mu_x^{-1} n^{-1/(d+1)} - h_n B_K + \mathrm{o}(n^{-1/(d+1)})$$

which suggests that the bias associated with the estimation of the true frontier $q_1(x)$ via $q_{1,n}(x)$ could be smaller than that associated with the FDH estimator. We now turn our attention to bandwidth selection.

3.2. Bandwidth selection

Implementation of our α-frontier estimator requires the selection of a bandwidth. Following standard practice (Fan and Gijbels, 1995; Ruppert et al., 1995) we select the bandwidth by minimizing an asymptotic approximation of the estimator's mean integrated squared error (AMISE) over all α. Disregarding terms of order $\mathrm{o}(h_n^4)$ and $\mathrm{o}(h_n/n)$ and defining $I_1(x,\alpha) = \int_{g^{-1}([q_\alpha(x), g(x)])} f^{(1)}(\gamma, q_\alpha(x))\,\mathrm{d}\gamma$, and $I_2(x,\alpha) = \int_{g^{-1}([q_\alpha(x), g(x)])} f(\gamma, q_\alpha(x))\,\mathrm{d}\gamma$ we have

$$\begin{aligned} AMISE(q_{\alpha,n}(x); h_n) = {} & \frac{h_n^4(\sigma_K^2)^2}{4F_X^2(x)}\int_0^1 \frac{I_1^2(x,\alpha)}{f^2(q_\alpha(x)/x)}\,\mathrm{d}\alpha + \frac{1}{nF_X(x)}\int_0^1 \frac{\alpha(1-\alpha)}{f^2(q_\alpha(x)/x)}\,\mathrm{d}\alpha \\ & - \frac{h_n 2\sigma_\kappa}{nF_X^2(x)}\int_0^1 \frac{I_2(x,\alpha)}{f^2(q_\alpha(x)/x)}\,\mathrm{d}\alpha \end{aligned}$$

a function of h_n. The bandwidth that minimizes $AMISE(q_{\alpha,n}(x); h_n)$ is given by

$$h_n^* = \left(\frac{2\sigma_\kappa \int_0^1 \dfrac{I_2(x,\alpha)}{f^2(q_\alpha(x)/x)}\mathrm{d}\alpha}{(\sigma_K^2)^2 \int_0^1 \dfrac{I_1^2(x,\alpha)}{f^2(q_\alpha(x)/x)}\mathrm{d}\alpha}\right)^{1/3} n^{-1/3} = Cn^{-1/3}.$$

Since our expression for AMISE accounts for all possible values of α, h_n^* can be interpreted as a global optimal bandwidth with respect to α for given input level x. Since our α-frontier estimator is constructed as a quantile estimator which smooths only the output for the underlying conditional distribution, it is not surprising that the optimal bandwidth is of order $\mathrm{O}(n^{-1/3})$. This is the same order obtained in Azzalini (1981), Bowman et al. (1998) and Hansen (2004a) where a kernel estimator is used to estimate an unconditional distribution. However, our constant C is different from theirs. Compared with other conditional quantile estimators, both the order and the constant C in the expression for h_n^* are different from those in Hansen (2004b) and Li and Racine (2005) since the conditioning set we consider $\{X\leqslant x\}$ is different.

The practical use of h_n^* requires the estimation of the unknowns appearing in its expression, as in the traditional plug-in bandwidth selection methods. In the next section, we provide an easily implementable estimation procedure for these unknowns and shed light on the finite sample performance of our estimator via a small Monte Carlo study.

4. Monte Carlo study

In this section, we perform a Monte Carlo study which implements our smooth α-frontier estimator (S) and provides evidence on its finite sample performance. For comparison purpose we also include in the study two alternative estimators, the empirical α-frontier estimator of Aragon et al. (2005) (E) and a conditional α-quantile estimator based on a linearly interpolated empirical conditional distribution (Kincaid and Cheney, 1996)(I). The interpolated estimator is interesting in that extra smoothness is obtained without the need for bandwidth estimation.

The data are simulated according to the model $Y_i = g(X_i)R_i$, $i = 1, 2, \ldots, n$ where Y_i represents output, the univariate input X_i are pseudorandom variables generated from a uniform distribution with support given by $[b_l, b_u]$. $R_i = \exp(-Z_i)$ and Z_i are independently generated pseudorandom variables from an exponential distribution with parameter $\beta = \frac{1}{3}$, therefore the efficiency R_i has support $(0, 1]$ with global average level of efficiency $\mathrm{E}(R_i) = 0.75$. We consider two specifications for $g(\cdot)$, $g_1(x) = \sqrt{x}$ with $[b_l, b_u] = [4, 25]$ and $g_2(x) = x^3$ with $[b_l, b_u] = [1, 2]$ which are associated with convex and nonconvex production technologies, respectively. This DGP has been considered in Aragon et al. (2005), Gijbels et al. (1999), Martins-Filho and Yao (2007a,b), Park et al. (2000) and is regarded as reasonable with respect to many applications found in the econometric literature (Gijbels et al., 1999, p. 224).

For each specification of $g(x)$ we consider three sample sizes $n = 100, 200$ and 400 and perform 1000 repetitions at each experiment design. We estimate the α-frontiers for $\alpha - 0.25, 0.5, 0.75$ and 0.99. Since the estimators for $q_\alpha(x)$ are constructed using data points with input levels which are less than or equal to x, we avoid estimation with extremely small samples by evaluating the performance of the estimators over the input interval starting from the 33rd percentile to the upper bound of the support. Using 30 equally spaced points in the support interval, we obtain the averaged bias, standard deviation and root mean squared error of each estimator. We also construct 95% asymptotic confidence intervals for the α-frontiers at different α levels using the asymptotic distributions available for our estimator and the estimator proposed by Aragon et al.

4.1. Estimator implementation

The empirical α-frontier estimator is implemented as described in Aragon et al. We implement the interpolated α-frontier estimator as

$$\hat{q}_{\alpha,nI}(x) = \begin{cases} Y_{(i_1)} & \text{if } 0 \leqslant \alpha < \dfrac{1}{N_x}, \\ Y_{(i_k)} + \left(\alpha - \dfrac{k}{N_x}\right) N_x (Y_{(i_{k+1})} - Y_{(i_k)}) & \text{if } \dfrac{k}{N_x} \leqslant \alpha < \dfrac{k+1}{N_x},\ 1 \leqslant k \leqslant N_x - 1, \\ 1 & \text{if } y \geqslant Y_{(i_{N_x})}, \end{cases}$$

where $N_x = \sum_{i=1}^{n} I(X_i \leqslant x)$ and $Y_{(i_j)}$ is the jth order statistic for the observations Y_i such that $X_i \leqslant x$. We note that $\hat{q}_{\alpha,nI}(x)$ produces estimates that are identical to those given by the empirical α-frontier estimator when α coincides with the nodes $\frac{k}{N_x}$. Our estimator is implemented using the Epanechnikov kernel and the following *plug-in* bandwidth

$$\hat{h}_{PI} = \left(\frac{2\sigma_K \int_0^1 \dfrac{\hat{I}_2(x,\alpha)}{\hat{f}^2(q_\alpha(x)/x)} \mathrm{d}\alpha}{(\sigma_K^2)^2 \int_0^1 \frac{\hat{I}_1^2(x,\alpha)}{\hat{f}^2(q_\alpha(x)/x)} \mathrm{d}\alpha} \right)^{1/3} n^{-1/3},$$

where $\hat{I}_1(x,\alpha)$, $\hat{I}_2(x,\alpha)$, $\hat{f}(q_\alpha(x)/x)$ are estimators for $I_1(x,\alpha)$, $I_2(x,\alpha)$ and $f(q_\alpha(x)/x)$ appearing in h_n^*. Specifically, $\hat{f}(q_{\alpha,n}(x)/x) = \frac{\frac{1}{ng_n}\sum_{i=1}^{n} K(\frac{Y_i - q_{\alpha,n}(x)}{g_n}) I(X_i \leqslant x)}{\hat{F}(x)}$ where $\hat{F}(x)$ is the empirical distribution function. Since $\hat{f}(q_{\alpha,n}(x)/x)$ is a suitably defined Rosenblatt density estimator, we utilize the *rule-of-thumb* bandwidth of Silverman (1986) for g_n. In $I_1(x,\alpha)$ and $I_2(x,\alpha)$ the area of integration $g^{-1}([q_{\alpha,n}(x), g(x)])$ needs to be estimated.

In the case of an univariate input ($d=1$) $g^{-1}([q_{\alpha,n}(x), g(x)]) = [g^{-1}(q_{\alpha,n}(x)), x]$. To estimate $I_1(x,\alpha)$ consider $\int_{b_1}^{b_2} f^{(1)}(x,y)\,\mathrm{d}x = \int_0^{b_2} f^{(1)}(x,y)\,\mathrm{d}x - \int_0^{b_1} f^{(1)}(x,y)\,\mathrm{d}x$ for some positive bounds b_1 and b_2. Given our estimator for the conditional distribution and an arbitrary $b>0$, a natural estimator for $\theta(y) = \int_0^b f^{(1)}(x,y)\,\mathrm{d}x$ is given by $\hat{\theta}(y) = \frac{1}{ng_{n1}}\sum_{i=1}^n K^{(1)}(\frac{y-Y_i}{g_{n1}})I(X_i \leqslant b)$, where $K^{(1)}(x) = \frac{\mathrm{d}K(x)}{\mathrm{d}x}$ for a bandwidth g_{n1}. The estimation of $\theta(y)$ requires a bandwidth selection procedure for g_{n1}. Based on the bias and variance expressions in Lemma 3 (Appendix A) we obtain the optimal bandwidth that minimizes the AMISE of $\hat{\theta}(y)$ as

$$g_{n1}^* = \left(\frac{3C_{K1}\int\int_0^b f(x,y)\,\mathrm{d}x\,\mathrm{d}y}{\frac{4}{36}C_K^2\int(\int_0^b f^{(3)}(x,y)\,\mathrm{d}x)^2\,\mathrm{d}y}\right)^{1/7} n^{-1/7},$$

where C_K and C_{K1} are constants given in Lemma 3 which depend only on the kernel, and $f^{(3)}(x,y)$ is the third order partial derivative of $f(x,y)$ with respect to y. It is straightforward to verify that the Epanechnikov kernel satisfies all conditions in Lemma 3, and since $\mathrm{E}(\hat{\theta}(y) - \theta(y)) = \mathrm{O}(g_{n1}^2)$, $V(\hat{\theta}(y)) = \mathrm{O}((ng_{n1}^3)^{-1})$ and $g_{n1}^* = \mathrm{O}(n^{-1/7})$, we have that $\hat{\theta}(y) - \theta(y) = \mathrm{o_p}(1)$. We note that $g_{n1}^* \propto n^{-1/7}$ is similar to the order obtained for optimal bandwidth for kernel density derivative estimation (Jones, 1994). The bandwidth g_{n1}^* depends on additional unknowns, but at this stage we follow standard practice and utilize the standard joint normal distribution as a reference. For $I_2(x,\alpha)$ we consider an estimator for $H(y) = \int_0^b f(x,y)\,\mathrm{d}x$ for some constant $b>0$. We define the estimator $\hat{H}(y) = \frac{1}{ng_{n2}}\sum_{i=1}^n K(\frac{y-Y_i}{g_{n2}})I(X_i \leqslant b)$ for a bandwidth g_{n2}. Since $\hat{H}(y)$ is a suitably defined Rosenblatt density estimator, we utilize the *rule-of-thumb* bandwidth of Silverman (1986) forg_{n2}. Finally, estimation of $I_1(x,\alpha)$ and $I_2(x,\alpha)$ requires estimators for $g^{-1}(\cdot)$ and $q_{\alpha,n}(x)$. Here, we utilize the FDH estimator for $\hat{g}^{-1}(\cdot)$ and provide an initial estimator for $q_{\alpha,n}(x)$ by using our α-quantile frontier estimator implemented with the h_n^* derived above based on a standard joint normal distribution for (x,y) and constant returns to scale production function $g(x) = 3x$.

The asymptotic properties of the proposed bandwidth selection rule $\hat{h}_{PI}$ are unknown. To shed some light on its finite sample performance and also to illustrate the relative performance of smooth and empirical frontier estimators without the noise introduced by bandwidth estimation, we implement the smooth frontier estimator with both the estimated bandwidth $\hat{h}_{PI}$ and the true optimal bandwidth h_n^*, which is available from the DGP. As will be discussed later, simulation results reveal that the performance of the smooth estimator with both bandwidths are similar for large sample sizes, suggesting that $\hat{h}_{PI}$ is "close" to h_n^* in probability.

Theorem 2 supports asymptotic confidence intervals for the smooth α-frontier estimator. Given that the asymptotic bias is $\mathrm{O}(h_n^2)$ and $h_n^* \propto n^{-1/3}$ we have that $\mathrm{O}(\sqrt{n}h_n^2) = \mathrm{O}(n^{-1/6}) = \mathrm{o}(1)$. Hence, the normalized bias vanishes asymptotically and for 97.5% quantile $Z_{0.975}$ of a standard normal distribution, we obtain $\lim_{n\to\infty} P(q_{\alpha,n}(x) - n^{-1/2}(\hat{S}_2^2)^{1/2}Z_{0.975} \leqslant q_\alpha(x) \leqslant q_{\alpha,n}(x) + n^{-1/2}(\hat{S}_2^2)^{1/2}Z_{0.975}) = 0.95$ where $\hat{S}_2^2 = \frac{\alpha(1-\alpha)}{\hat{F}(x)(\hat{f}(q_{\alpha,n}(x)/x))^2}$. $\hat{F}(x)$ and $\hat{f}(q_{\alpha,n}(x)/x)$ are estimated as described in the bandwidth selection procedure. The asymptotic confidence interval for the empirical α-frontier estimator is constructed in a similar manner.

4.2. Results and analysis

Fig. 1 depicts the true α-frontier with estimated smooth and empirical frontiers for α ranging over $0.02, 0.04, \ldots, 1$ for a simulated data set of size $n=50$ with $g_1(x) = g_1(25)$. As expected, our α-frontier estimate is a smooth function of α and the empirical α-frontier is not. Table 1 provides the average root mean squared error of α-frontier estimators for $\alpha = 0.25, 0.5, 0.75, 0.99$. Also included are results for the smooth α-frontier estimator implemented with the true optimal bandwidth h_n^*.[4]

First, we compare the performance of the empirical estimator with that of the smooth estimator using h_n^*. In terms of root mean squared error, when $g_1(x)$ is considered, the smooth estimator outperforms the empirical

[4]Results for other values of α, averaged bias and standard deviation of the estimators are provided in Martins-Filho and Yao (2007b).

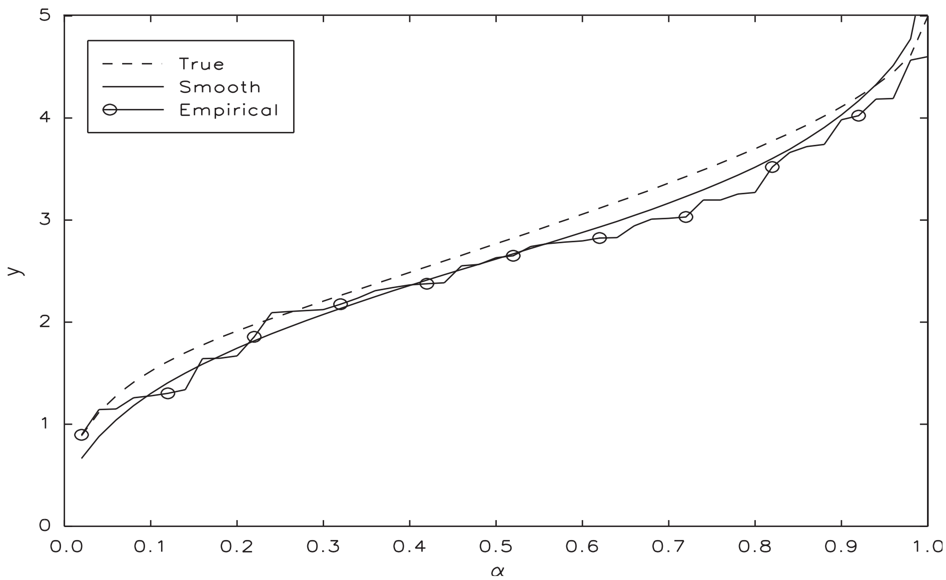

Fig. 1. Plot of true α-frontiers with estimated smooth and empirical α-frontiers, for $n = 50$, $g_1(x) = g_1(25)$ and α ranging over $0.02, 0.04, \ldots, 1$.

Table 1
Root mean squared error for α-frontier estimators

α	$g(x) = \sqrt{x}$				$g(x) = x^3$			
	S		E	I	S		E	I
	h_n^*	$\hat{h}_{PI}$			h_n^*	$\hat{h}_{PI}$		
$n = 100$								
0.25	0.126	0.128	0.136	0.137	0.102	0.103	0.110	0.109
0.5	0.118	0.120	0.137	0.137	0.159	0.162	0.174	0.173
0.75	0.121	0.124	0.142	0.143	0.211	0.215	0.229	0.228
0.99	0.248	0.303	0.133	0.161	0.266	1.037	0.257	0.307
$n = 200$								
0.25	0.095	0.096	0.101	0.101	0.074	0.075	0.079	0.078
0.5	0.089	0.089	0.099	0.099	0.111	0.113	0.120	0.119
0.75	0.091	0.093	0.103	0.104	0.150	0.151	0.161	0.161
0.99	0.159	0.177	0.096	0.107	0.165	0.295	0.187	0.205
$n = 400$								
0.25	0.065	0.066	0.069	0.069	0.051	0.051	0.054	0.054
0.5	0.065	0.064	0.071	0.071	0.079	0.079	0.084	0.084
0.75	0.066	0.066	0.073	0.073	0.106	0.106	0.132	0.113
0.99	0.105	0.123	0.069	0.073	0.116	0.131	0.132	0.139

estimator when α assumes values of $0.25, 0.5, 0.75$, with empirical estimator performing better when α is close to 1. When $g_2(x)$ is considered, the smooth estimator is superior for most experiments, where the only exception occurs in small samples when α is close to 1. When the smooth estimator is implemented with $\hat{h}_{PI}$, as we expect, the performance of the smooth estimator is slightly worse than that with h_n^* in terms of root mean squared error. However, the conclusions regarding the relative performance of the empirical and smooth estimators is largely maintained, suggesting that $\hat{h}_{PI}$ is estimating h_n^* quite well. These results are driven by the significantly smaller standard deviation of the smooth estimator, which compensates for slightly larger bias, confirming the asymptotic result in Theorem 2.

Table 2
Empirical coverage probability for α-frontier estimators smooth (s) and empirical (E)

α	$g(x)=\sqrt{x}$						$g(x)=x^3$					
	x_{10}		x_{20}		x_{30}		x_{10}		x_{20}		x_{30}	
	S	E	S	E	S	E	S	E	S	E	S	E
$n=100$												
0.25	0.961	0.929	0.970	0.946	0.964	0.944	0.973	0.963	0.985	0.960	0.978	0.957
0.5	0.957	0.916	0.960	0.932	0.968	0.940	0.954	0.899	0.950	0.919	0.942	0.914
0.75	0.963	0.921	0.958	0.927	0.976	0.938	0.940	0.903	0.943	0.913	0.926	0.910
0.99	0.995	0.879	0.992	0.903	0.988	0.772	0.981	0.832	0.958	0.892	0.979	0.754
$n=200$												
0.25	0.954	0.927	0.956	0.944	0.958	0.942	0.964	0.956	0.973	0.955	0.975	0.964
0.5	0.967	0.927	0.954	0.933	0.964	0.925	0.964	0.916	0.961	0.936	0.952	0.936
0.75	0.956	0.928	0.966	0.933	0.962	0.941	0.959	0.927	0.947	0.928	0.949	0.924
0.99	0.998	0.817	0.994	0.905	0.993	0.826	0.975	0.810	0.972	0.879	0.984	0.826
$n=400$												
0.25	0.961	0.941	0.960	0.945	0.961	0.950	0.969	0.959	0.969	0.952	0.972	0.954
0.5	0.960	0.933	0.956	0.930	0.954	0.932	0.954	0.937	0.966	0.958	0.952	0.931
0.75	0.964	0.940	0.958	0.928	0.955	0.938	0.956	0.937	0.952	0.937	0.952	0.934
0.99	0.997	0.879	0.986	0.891	0.963	0.891	0.972	0.876	0.973	0.884	0.968	0.890

A comparison between our smooth estimator and the interpolated estimator resembles that between our estimator and the empirical estimator. We also observe that interpolated estimator has larger bias, slightly smaller standard deviation, but slightly larger root mean squared error than the empirical estimator.

We find that as n increases the root mean squared error of all estimators decreases confirming the asymptotic results in the previous section.[5] This indicates that $\hat{q}_{\alpha,nI}(x)$ may be a consistent estimator of the α-frontier and that $\hat{h}_{PI}$ is likely a consistent estimator for h_n^*. We observe that for both $g_1(x)$ and $g_2(x)$ the root mean squared error for all estimators is generally larger when evaluating α-frontier with α closer to 1 than when evaluating frontiers with $\alpha = 0.25, 0.5, 0.75$. The fact that it is more difficult to estimate α-frontiers in this case is intuitively understood as there are relatively less representative data available when α is closer to 1.

The empirical coverage probability (the frequency that the estimated confidence interval contains the true α-frontier in 1000 repetitions) is given in Table 2 for the 33rd(x_{10}), 66th(x_{20}), and 100th(x_{30}) percentile of the input evaluation interval for empirical and smooth α-frontier estimators based on $\hat{h}_{PI}$. For most experiments we observe that the smooth estimator is superior to empirical estimator, i.e., the empirical coverage probability with the smooth estimator is closer to the target value 95% than that with the empirical estimator, where exceptions occur mostly for $\alpha = 0.25$. As n increases the empirical coverage probabilities from both estimators tend to get closer to 95% with some exceptions. There is also weak evidence that for the empirical estimator the coverage gets closer to 95% as α decreases. Fig. 2 provides 95% empirical coverage probability for the estimators for the $\alpha = 0.99$ frontier and a sample size $n = 100$ for 30 points across the input support. As indicated in the graph, for both $g_1(x)$ and $g_2(x)$, the smooth α-frontier estimator's empirical coverage probability slightly overestimates the 95% target. Coverage for empirical estimator is largely below the 95% target with large deviations close to the boundary of the input support.

To provide further evidence on the finite sample distribution of the two estimators, we provide kernel density estimates for the smooth and empirical α-frontier estimators centered around the true value (for $\alpha = 0.98$) $q_{0.98}(x_0)$ based on 1000 simulations of sample sizes $n = 100$ and 400 for $g_1(x)$ with $x_0 = 25$ in Fig. 3 and for $g_2(x)$ with $x_0 = 2$ in Fig. 4.

[5]This is also true for bias (with exceptions) and variance.

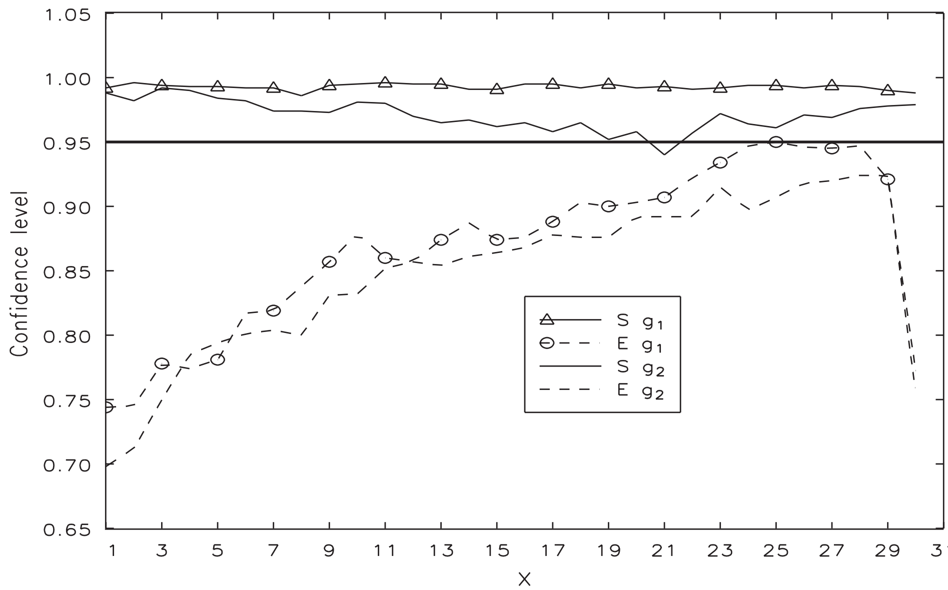

Fig. 2. Empirical 95% coverage probability for α-frontier smooth (S) and empirical (E) estimators for 30 grid points of X when $\alpha = 0.99$, $n = 100$.

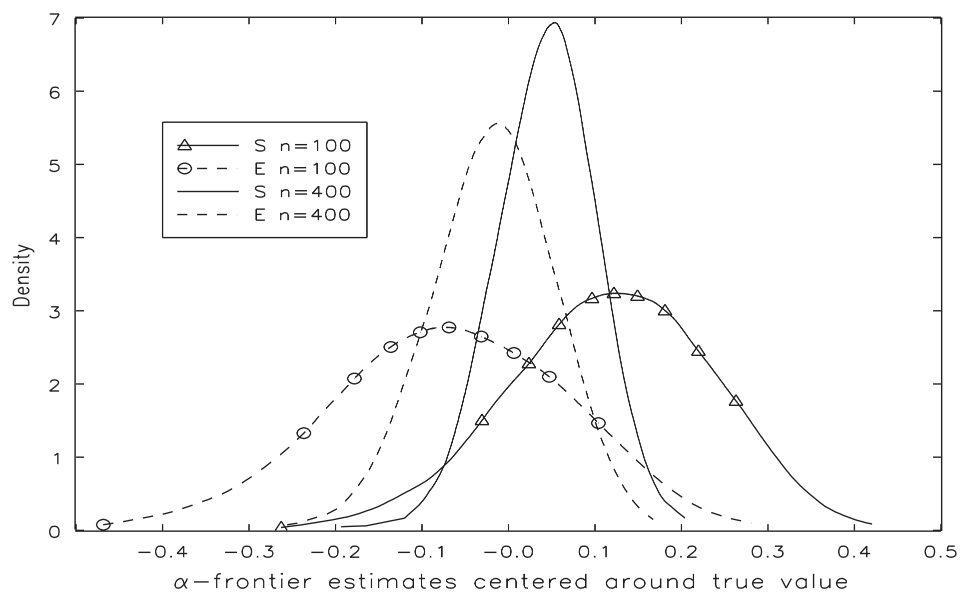

Fig. 3. Kernel density estimates for the smooth (S) and empirical (E) α-frontier estimators evaluated at $x_0 = 25$ centered around the true value $q_{0.98}(25)$, the $\alpha = 0.98$ frontier function. The kernel density estimates are based on 1000 simulations from $g_1(x)$ of sample sizes $n = 100$ and 400.

Fig. 3 shows that the kernel density for the smooth estimator is shifted to the right and more tightly centered, implying smaller variance, but larger bias compared to the empirical estimator. Fig. 4 shows a similar pattern, but here the smooth estimator exhibits significantly smaller bias, suggesting an improved performance when estimating nonconvex technologies. We note that the estimated densities have taller and more pronounced peaks as the sample size increases, confirming the asymptotic results.

Overall our simulations seem to indicate that the proposed smooth α-frontier estimator can outperform the empirical α-frontier estimator in terms of root mean squared error when n is large. This is particularly true when estimating production frontiers associated with nonconvex technologies. The simulations also reveal

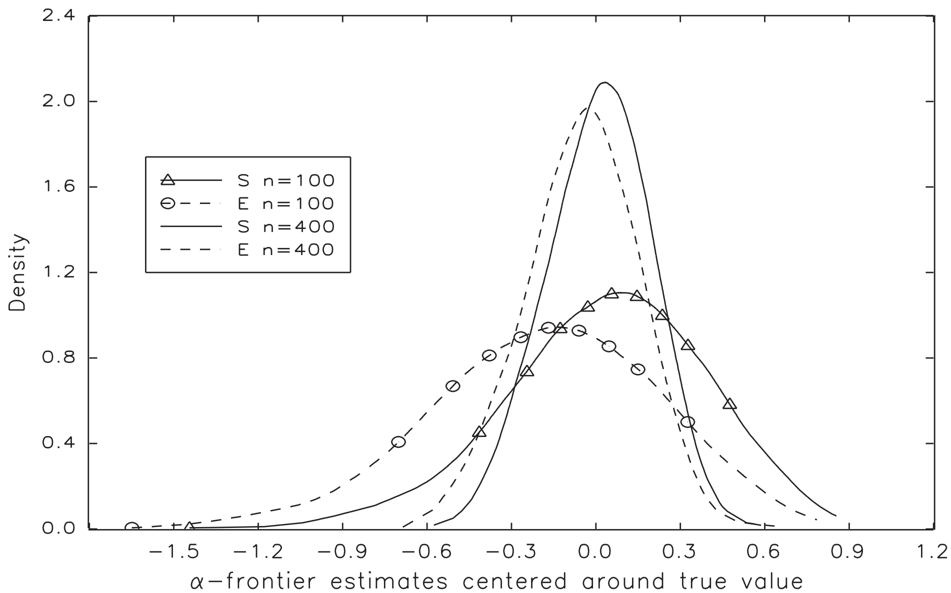

Fig. 4. Kernel density estimates for the smooth(S) and empirical(E) α-frontier estimators evaluated at $x_0 = 2$ centered around the true value $q_{0.98}(25)$, the $\alpha = 0.98$ frontier function. The kernel density estimates are based on 1000 simulations from $g_2(x)$ of sample sizes $n = 100$ and 400.

that although computationally demanding, bandwidth estimation does not significantly impact estimator performance if compared to implementation with a true bandwidth, indicating that bandwidth selection is not a significant burden in terms of estimator properties and relative performance.

5. Empirical illustration

To illustrate our methodology, we employ data on 123 utility companies from the United States reported in Greene (1990). These data consist of variables on production cost, output, input prices, and has been analyzed by Christensen and Greene (1976), Greene (1990) and Gijbels et al. (1999). Following Gijbels et al. (1999), we utilize only the measurements on the output variable with $Y = \mathrm{Ln}(Q)$ and input or cost variable defined as $X = \mathrm{Ln}(C)$, where Q is the production output for a firm, and C is the total cost involved in the production. For detailed description of the data set and analysis, see Christensen and Greene (1976) and Greene (1990).

In Fig. 5, we provide a scatterplot of the data and construct 95% confidence intervals for the $\alpha = 0.90$ frontiers using the smooth estimator following the steps outlined in the simulation section. For illustration purpose, we restrict the estimation region to be $x \in [0, 6]$, where 109 out of the 123 observations are located. The bandwidth for our smooth estimator is selected according to the *plug-in* rule $\hat{h}_{PI}$ as described in the simulation section. We note that the confidence bands are wider in regions of the input space where there are a smaller number of observations. This follows from our definition for asymptotic confidence intervals and Theorem 4.1 of Aragon et al. Indeed the width of the confidence interval depends on the density $f(q_\alpha(x)/x)$ and marginal probability $F_X(x)$. In regions of the input space where there are more data, both the density and marginal probability will be larger, and hence it is natural to observe narrower confidence intervals. Given the comments in Aragon et al. (2005) regarding the robustness of the empirical frontier to extreme observations, we conjecture that for $\alpha \in (0, 1)$ our smooth estimator should also be reasonably robust to extreme values and outliers.

6. Summary

In this paper we proposed a nonparametric α-frontier estimator based on a smooth kernel estimator of a conditional quantile of order α. Our estimator is an alternative to the conditional quantile estimator proposed

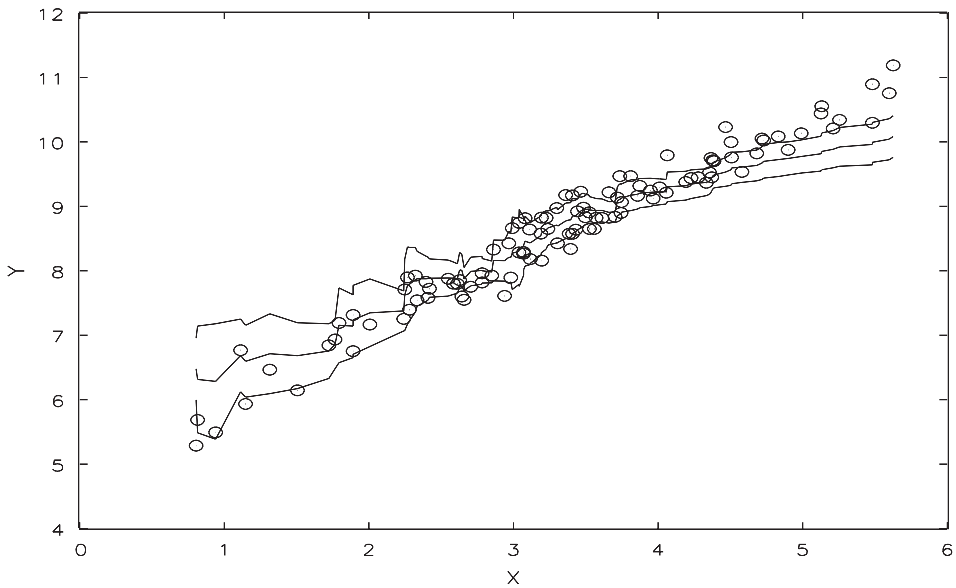

Fig. 5. 95% confidence intervals for $\alpha = 0.90$ frontiers with smooth estimate using American Electric Utility data.

by Aragon et al. (2005), which is based on empirical distribution functions. The estimator is easily implementable and we show that it is consistent and $\sqrt{n}$ asymptotically normal. In addition, the extra smoothness pays off in that our estimator's variance is smaller due to higher order terms than that of the estimator proposed by Aragon et al. (2005). Our simulation study confirms the asymptotic theory predictions and contrasts our estimator with that of Aragon et al. In most of the experiment designs in the simulations, our smooth estimator outperforms the empirical distribution based estimator of Aragon et al. (2005). Future work is needed in the context of α-frontiers, specifically estimators that can produce smooth boundaries over the input set are desirable in the applied economics literature.

Acknowledgment

We thank participants of the Econometrics in Rio Conference, Jeff Racine, Aman Ullah, an associate editor and two anonymous referees for excellent comments. Special thanks go to Professor Takeshi Amemiya for his editorship. The authors retain responsibility for any remaining errors.

Appendix A. Proofs

Lemma 1. *For all* $x \in \Re^d_+$ *and* $y \in \Re_+$ *and under Assumptions* A1, A2(a), A2(b), A2(c), *and* A3, *we have*:

(a)
$$E(\hat{F}(x,y)) = \begin{cases} F(x,y) + \frac{1}{2}h_n^2\sigma_K^2 \int_{g^{-1}([y,g(x)])} f^{(1)}(X,y)\,\mathrm{d}X + o(h_n^2) & \text{if } 0<y<g(x), \\ F(x,y) + o(h_n^2) & \text{if } y>g(x), \\ F(x,y) + o(h_n) & \text{if } y = g(x). \end{cases}$$

(b)
$$V(\hat{F}(x,y)) = \begin{cases} n^{-1}F(x,y)(1-F(x,y)) - 2n^{-1}h_n\sigma_\kappa \int_{g^{-1}([y,g(x)])} f(X,y)\,\mathrm{d}X + o(h_n/n) & \text{if } 0<y<g(x), \\ n^{-1}F(x,y)(1-F(x,y)) + o(h_n/n) & \text{if } y \geqslant g(x), \end{cases}$$

where $\kappa(x) = \int_{-B_K}^{x} K(\gamma)\,\mathrm{d}\gamma$, $\sigma_\kappa = \int_{-B_K}^{B_K} \gamma\kappa(\gamma)K(\gamma)\,\mathrm{d}\gamma$, $f^{(1)}(X,y)$ *denotes the first derivative of f with respect to* Y, *and* $0<h_n \to 0$ *is a nonstochastic sequence of bandwidths.*

Proof. (a) Let $C_x = \times_{i=1}^{d}[0, x_i]$ where x_i is the ith component of x. Since $h_n \to 0$ as $n \to \infty$, there exists $N(x) \in \Re_+$ such that for all $n > N(x)$, $\mathrm{E}(\hat{F}(x,y)) = \int_{C_x}\int_{[0,g(X)]}\kappa(\frac{y-Y}{h_n})\frac{\partial F_f(X,Y)}{\partial Y}\,\mathrm{d}Y\,\mathrm{d}X$ where $F_f(x,y) = \int_{[0,y]} f(x,\gamma)\,\mathrm{d}\gamma$. Using integration by parts

$$\int_{[0,g(X)]}\kappa\left(\frac{y-Y}{h_n}\right)\mathrm{d}F_f(X,Y) = \kappa\left(\frac{y-g(X)}{h_n}\right)F_f(X,g(X)) + \int_{y-g(X)/h_n}^{y/h_n} F_f(X, y-h_n\gamma)K(\gamma)\,\mathrm{d}\gamma.$$

By A3(d) and Taylor's theorem $F_f(X, y-h_n\gamma) = F_f(X,y) - h_n\gamma f(X,y) + \frac{1}{2}h_n^2\gamma^2 f^{(1)}(X,y) + \mathrm{o}(h_n^2)$. Hence, $\mathrm{E}(\hat{F}(x,y)) = E_{1n} + E_{2n} - E_{3n} + E_{4n} + \mathrm{o}(h_n^2)$, where $E_{1n} = \int_{C_x}\kappa(\frac{y-g(X)}{h_n})F_f(X,g(X))\,\mathrm{d}X$, $E_{2n} = \int_{C_x} F_f(X,y)$ $\int_{(y-g(X))/h_n}^{y/h_n} K(\gamma)\,\mathrm{d}\gamma\,\mathrm{d}X$, $E_{3n} = h_n\int_{C_x} f(X,y)\int_{(y-g(X))/h_n}^{y/h_n}\gamma K(\gamma)\,\mathrm{d}\gamma\,\mathrm{d}X$, $E_{4n} = \frac{h_n^2}{2}\int_{C_x} f^{(1)}(X,y)$ $\int_{(y-g(X))/h_n}^{y/h_n}\gamma^2 K(\gamma)$ $\mathrm{d}\gamma\,\mathrm{d}X$. For $(x,y) \in \Psi^*$, if $y \leqslant 0$ then $\hat{F}(x,y) = 0$. We now consider the limiting behavior of each term when: (1) $0 < y < g(x)$; (2)$y > g(x)$; (3) $y = g(x)$.

(1) For any $A \subseteq C_x$ and $B \subseteq [0, g(x)]$, let $g(A) = \{g(x) : x \in A\}$ and $g^{-1}(B) = \{x : x \in C_x,\ g(x) \in B\}$. Then, $E_{1n} = \int_{g^{-1}([0,y])}\kappa(\frac{y-g(X)}{h_n})F_f(X,g(X))\,\mathrm{d}X + \int_{g^{-1}([y,g(x)])}\kappa(\frac{y-g(X)}{h_n})F_f(X,g(X))\,\mathrm{d}X = E_{11,n} + E_{12,n}$. By A1, $|\kappa(\frac{y-g(X)}{h_n})||$ $F_f(X,g(X)| < \infty$ and by Lebesgue's dominated convergence (LDC) theorem $E_{11,n} \to \int_{g^{-1}([0,y])}\int_{[0,g(X)]}$ $f(X,Y)\,\mathrm{d}X\,\mathrm{d}Y$ since $X \in g^{-1}([0,y])$ and $\kappa(\frac{y-g(X)}{h_n}) \to 1$. Similarly, $E_{12,n} \to 0$ since $X \in g^{-1}([y,g(x)])$ and $\kappa(\frac{y-g(X)}{h_n}) \to 0$. $E_{2n} \to \int_{g^{-1}([y,g(x)])}\int_{[0,y]} f(X,Y)\,\mathrm{d}Y\,\mathrm{d}X$ since for $X \in g^{-1}([0,y])$, we have $\int_{(y-g(X))/h_n}^{y/h_n}K(\gamma)\,\mathrm{d}\gamma \to 0$, and for $X \in g^{-1}([y,g(x)])$ we have $\int_{(y-g(X))/h_n}^{y/h_n}K(\gamma)\,\mathrm{d}\gamma \to 1$. $h_n^{-1}E_{3n} \to 0$ since for $X \in g^{-1}([0,y])$ we have $\int_{(y-g(X))/h_n}^{y/h_n}\gamma K(\gamma)\,\mathrm{d}\gamma \to 0$ and by A2(c), for $X \in g^{-1}([y,g(x)])$ we have $\int_{(y-g(X))/h_n}^{y/h_n}\gamma K(\gamma)\,\mathrm{d}\gamma \to 0$. Now, $h_n^{-2}E_{4n} \to \frac{1}{2}\sigma_K^2\int_{g^{-1}([y,g(x)])} f^{(1)}(X,y)\,\mathrm{d}X$ since for $X \in g^{-1}([0,y])$ we have $\int_{(y-g(X))/h_n}^{y/h_n}\gamma^2 K(\gamma)\,\mathrm{d}\gamma \to 0$ and by A2(c), for $X \in g^{-1}([y,g(x)])$ we have $\int_{(y-g(X))/h_n}^{y/h_n}\gamma^2 K(\gamma)\,\mathrm{d}\gamma \to \sigma_K^2$. Hence, for $0 < y < g(x)$ we have $\mathrm{E}(\hat{F}(x,y)) = F(x,y) + \frac{h_n^2}{2}\sigma_K^2\int_{g^{-1}([y,g(x)])} f^{(1)}(X,y)\,\mathrm{d}X + \mathrm{o}(h_n^2)$. For cases (2) and (3) results are obtained in a similar manner. (b) Note that $V(\hat{F}(x,y)) = \frac{1}{n}(V_{1n} - V_{2n})$ where $V_{1n} = \mathrm{E}((\frac{1}{h_n}\int_0^y K(\frac{\gamma-Y}{h_n})\,\mathrm{d}\gamma)^2 I(X_i \leqslant x))$ and $V_{2n} = (\mathrm{E}(\frac{1}{h_n}\int_0^y K(\frac{\gamma-Y}{h_n})\,\mathrm{d}\gamma I(X_i \leqslant x)))^2$. The results are obtained following arguments similar to those in (a). □

Lemma 2. *Let $0 < h_n \to 0$ be a nonstochastic sequence of bandwidths with $nh_n \to \infty$ as $n \to \infty$. Assume that for a given $x \in \Re_+^d$ and some $N(x)$ we have that for all $n > N(x)\min_{\{i:X_i \leqslant x\}} Y_i \geqslant h_n B_K$ and* A1–A4. *Then,* (a) $\sup_{y\in[0,g(x)]}|\hat{F}(x,y) - \mathrm{E}(\hat{F}(x,y))| = \mathrm{o_p}(1)$ *and* (b) $\sup_{y\in[0,g(x)]}|\mathrm{E}(\hat{F}(x,y)) - F(x,y))| = \mathrm{o}(1)$.

Proof. (a) Given $\min_{\{i:X_i \leqslant x\}} Y_i \geqslant h_n B_K$, $\hat{F}(x,y) = \frac{1}{n}\sum_{i=1}^{n}\kappa(\frac{y-Y_i}{h_n})I(X_i \leqslant x)$. Since $G(x) = [0, g(x)]$ is compact, there exists $y_0 \in G(x)$ such that $G(x) \subseteq B(y_0, r_x)$ where $B(y_0, r_x) = \{y \in \Re : |y - y_0| < r_x\}$ where$r_x \in \Re_+$. By the Heine–Borel theorem there exists $\{B(y_k, (n/h_n^a)^{-1/2})\}_{k=1}^{l_n}$, $a > 0$ such that $G(x) \subset \cup_{k=1}^{l_n} B(y_k, (n/h_n^a)^{-1/2})$ for $y_k \in G(x)$ with $l_n < r_x(n/h_n^a)^{1/2}$. For $y \in B(y_k, (n/h_n^a)^{-1/2})$ we have

$$\begin{aligned}|\hat{F}(x,y) - \hat{F}(x,y_k)| &\leqslant \frac{1}{n}\sum_{i=1}^{n}\left|\kappa\left(\frac{y-Y_i}{h_n}\right) - \kappa\left(\frac{y_k - Y_i}{h_n}\right)\right| I(X_i \leqslant x)\\ &\leqslant m_\kappa h_n^{-1}|y - y_k| \leqslant m_\kappa(nh_n^{2-a})^{-1/2} \quad \text{by A2(e) and the fact that } I(X_i \leqslant x) \leqslant 1.\end{aligned}$$

Also, $|\mathrm{E}(\hat{F}(x,y)) - \mathrm{E}(\hat{F}(x,y_k))| \leqslant \int_{C_x}\int_{[0,g(X)]}|\kappa(\frac{y-Y}{h_n}) - \kappa(\frac{y_k-Y}{h_n})| f(X,Y)\,\mathrm{d}Y\,\mathrm{d}X \leqslant m_\kappa F_X(x)(nh_n^{2-a})^{-1/2}$ by A2(e). Hence, $|\hat{F}(x,y) - \mathrm{E}(\hat{F}(x,y))| \leqslant |\hat{F}(x,y_k) - \mathrm{E}(\hat{F}(x,y_k))| + m_\kappa(nh_n^{2-a})^{-1/2}(1 + F_X(x))$ and $\sup_{y\in G(x)}|\hat{F}(x,y) - \mathrm{E}(\hat{F}(x,y))| \leqslant \max_{1\leqslant k\leqslant l_n}|\hat{F}(x,y_k) - \mathrm{E}(\hat{F}(x,y_k))| + m_\kappa(nh_n^{2-a})^{-1/2}(1 + F_X(x))$. Taking $a = 1$ and given that $nh_n \to \infty$, we have $m_\kappa(nh_n^{2-a})^{-1/2} \to 0$. Hence, we need only show that for all $\varepsilon_n > 0$, $\lim_{n\to\infty}\mathrm{P}(\max_{1\leqslant k\leqslant l_n}|\hat{F}(x,y_k) - \mathrm{E}(\hat{F}(x,y_k))| \geqslant \varepsilon_n) = 0$. It suffices to establish that $\lim_{n\to\infty}\sum_{l=1}^{l_n}\mathrm{P}(|\hat{F}(x,y_k) - \mathrm{E}(\hat{F}(x,y_k))| \geqslant \varepsilon_n) = 0$. Note that, $|\hat{F}(x,y_k) - \mathrm{E}(\hat{F}(x,y_k))| = |n^{-1}\sum_{i=1}^{n} W_{in}|$ where $W_{in} = \kappa(\frac{y_k - Y_i}{h_n})I(X_i \leqslant x) -$

$\mathrm{E}(\kappa\left(\frac{y_k-Y_i}{h_n}\right)I(X_i\leqslant x))$, with $\mathrm{E}(W_{in})=0$ and $|W_{in}|\leqslant 2$ given that $I(\cdot),\kappa(\cdot)\leqslant 1$. Given A1, by Bernstein's inequality we have $\mathrm{P}(|\hat{F}(x,y_k)-\mathrm{E}(\hat{F}(x,y_k))|\geqslant\varepsilon)<2\exp(\frac{-n\varepsilon_n^2}{2\bar{\sigma}^2+\frac{4}{3}\varepsilon_n})$, with $\bar{\sigma}^2=n^{-1}\sum_{i=1}^n V(W_{in})=V_{1n}(x,y_k)-V_{2n}(x,y_k)\to F(x,y_k)(1-F(x,y_k))$ for $y_k\in G(x)$. Let $c_n=2\bar{\sigma}^2+\frac{4}{3}\varepsilon_n$ and $\varepsilon_n=\left(\frac{\ln(n)}{n}\right)^{1/2}\Delta$ for some $\Delta>0$. Then, $\mathrm{P}(|\hat{F}(x,y_k)-\mathrm{E}(\hat{F}(x,y_k))|\geqslant\varepsilon)\leqslant 2r_x(\frac{n}{h_n})^{1/2}n^{-\Delta/c_n}\leqslant r_x(nh_n)^{-1/2}$ for Δ sufficiently large. Hence, $\lim_{n\to\infty}\mathrm{P}((\frac{n}{\ln(n)})^{1/2}\sup_{y\in G(x)}|\hat{F}(x,y)-\mathrm{E}(\hat{F}(x,y))|\geqslant\Delta)=0$ and consequently $\sup_{y\in G(x)}|\hat{F}(x,y)-\mathrm{E}(\hat{F}(x,y))|=o_{\mathrm{p}}(1)$. (b) Note that

$$\sup_{y\in[0,g(x)]}|\mathrm{E}(\hat{F}(x,y))-F(x,y))|\leqslant\sup_{y\in[0,g(x)]}\left|E_{1n}(y)-\int_{g^{-1}([0,y])}\int_{[0,g(X)]}f(X,Y)\,\mathrm{d}Y\,\mathrm{d}X\right| + \sup_{y\in[0,g(x)]}\left|E_{2n}(x,y)-\int_{g^{-1}([y,g(x)])}\int_{[0,y]}f(X,Y)\,\mathrm{d}Y\,\mathrm{d}X\right|+\sup_{y\in[0,g(x)]}|E_{3n}(x,y)|,$$

where $E_{1n}(y)=\int_{g^{-1}([0,y])}\int_{[0,g(X)]}\kappa(\frac{y-Y}{h_n})f(X,Y)\,\mathrm{d}Y\,\mathrm{d}X$, $E_{2n}(x,y)=\int_{g^{-1}([y,g(x)])}\int_{[0,y]}\kappa(\frac{y-Y}{h_n})f(X,Y)\,\mathrm{d}Y\,\mathrm{d}X$ and $E_{3n}(x,y)=\int_{g^{-1}([y,g(x)])}\int_{[y,g(X)]}\kappa(\frac{y-Y}{h_n})f(X,Y)\,\mathrm{d}Y\,\mathrm{d}X$. To complete the proof we show that each supremum on the r.h.s. is o(1). For the first term we have: (a) $X\in g^{-1}([0,y])$ which implies that $g(X)\leqslant y$ and as $n\to\infty$, $\frac{y-Y}{h_n}>B_K$ and $\kappa(\frac{y-Y}{h_n})\to 1$, hence by LDC theorem $E_{1n}(y)\to\int_{g^{-1}([0,y])}\int_{[0,g(X)]}f(X,Y)\,\mathrm{d}Y\,\mathrm{d}X$ for every $y\in[0,g(x)]$; (b) For all $y\in[0,g(x)]\,E_{1n}(y)\leqslant E_{1,n+1}(y)$, hence given A1(b), κ satisfies a Lipschitz condition, which together with A4 shows that $E_{1n}(y)$ is continuous. Since $\int_{g^{-1}([0,y])}\int_{[0,g(X)]}f(X,Y)\,\mathrm{d}Y\,\mathrm{d}X$ is continuous in y, $\sup_{y\in[0,g(x)]}|E_{1n}(y)-\int_{g^{-1}([0,y])}\int_{[0,g(X)]}f(X,Y)\,\mathrm{d}Y\,\mathrm{d}X|=o(1)$. Following a similar argument $\sup_{y\in[0,g(x)]}|E_{2n}(x,y)-\int_{g^{-1}([y,g(x)])}\int_{[0,y]}f(X,Y)\,\mathrm{d}Y\,\mathrm{d}X|=o(1)$. □

Proof of Theorem 1. From Nadaraya (1964), for all $\varepsilon>0$, we have $F(q_\alpha(x)+\varepsilon/x)>F(q_\alpha(x)/x)>F(q_\alpha(x)-\varepsilon/x)$. If $\omega\in A\equiv\{\omega:|q_{\alpha,n}(x)-q_\alpha(x)|>\varepsilon\}$ we have that either $F(q_{\alpha,n}(x)/x)\geqslant F(q_\alpha(x)+\varepsilon/x)$ or $F(q_{\alpha,n}(x)/x)\leqslant F(q_\alpha(x)-\varepsilon/x)$. Hence, there exists $0<\delta(\varepsilon,x)$ where $\delta(\varepsilon,x)=\min\{F(q_\alpha(x)+\varepsilon/x)-F(q_\alpha(x)/x),F(q_\alpha(x)/x)-F(q_\alpha(x)-\varepsilon/x)\}$ such that $\omega\in B=\{\omega:|F(q_{\alpha,n}(x)/x)-F(q_\alpha(x)/x)|>\delta(\varepsilon,x)\}$ so $A\subseteq B$ and $P(|q_{\alpha,n}(x)-q_\alpha(x)|>\varepsilon)\leqslant P(|F(q_{\alpha,n}(x)/x)-F(q_\alpha(x)/x)|>\delta(\varepsilon,x))$. Since, $|F(q_{\alpha,n}(x)/x)-F(q_\alpha(x)/x)|=|F(q_{\alpha,n}(x)/x)-\hat{F}(q_{\alpha,n}(x)/x)|\leqslant\sup_{y\in\Re_+}|\hat{F}(y/x)-F(y/x)|$ and we write $\sup_{y\in\Re_+}|\hat{F}(y/x)-F(y/x)|\leqslant\frac{1}{\hat{F}(x)}\sup_{y\in\Re_+}|\hat{F}(x,y)-F(x,y)|+|\frac{1}{F_X(x)}-\frac{1}{\hat{F}(x)}|F_X(x)$ since $F(x,y)\leqslant F_X(x)$. Now, we have $\sup_{y\in\Re_+}|\hat{F}(x,y)-F(x,y)|\leqslant\sup_{y\in[0,g(x)]}|\hat{F}(x,y)-F(x,y)|+\sup_{(g(x),\infty)}|\hat{F}(x,y)-F(x,y)|$. From Lemma 2, $\sup_{y\in[0,g(x)]}|\hat{F}(x,y)-F(x,y)|=o_{\mathrm{p}}(1)$. For all $y\in(g(x),\infty)$ we have that $F(x,y)=F(x,g(x))=\int_{C_x}\int_{[0,g(X)]}f(X,Y)\,\mathrm{d}Y\,\mathrm{d}X=F_X(x)$. In addition, given $\min_{\{i:X_i\leqslant x\}}Y_i\geqslant h_nB_K$ and $0<Y\leqslant g(x)$, we have that for all $y\in(g(x),\infty)$, $y-Y>0$. Hence, there exists $N(x)$ such that for all $n>N(x)$ we have that $\hat{F}(x,y)=n^{-1}\sum_{i=1}^n\int_{-B_K}^{B_K}K(\gamma)\,\mathrm{d}\gamma I(X_i\leqslant x)=n^{-1}\sum_{i=1}^n I(X_i\leqslant x)=\hat{F}(x)$. Therefore, $\sup_{(g(x),\infty)}|\hat{F}(x,y)-F(x,y)|=\sup_{(g(x),\infty)}|\hat{F}(x)-F_X(x)|=o_{\mathrm{p}}(1)$ given Chebyshev's inequality and $\hat{F}(x)-F_X(x)=o_{\mathrm{p}}(1)$. To complete the proof, note that $\hat{F}(x)=O_{\mathrm{p}}(1)$, and $F_X(x)>0$, hence by Slutsky theorem $\hat{F}(x)^{-1}-F_X(x)^{-1}=o_{\mathrm{p}}(1)$. □

Proof of Theorem 2. $q_{\alpha,n}(x)-q_\alpha(x)=(A_n+C_n)(\frac{1}{f(q_\alpha(x)/x)}+\beta_n)$, where $A_n=F(q_\alpha(x)/x)-\frac{\mathrm{E}(\hat{F}(x,q_\alpha(x)))}{\mathrm{E}(\hat{F}(x))}$, $\beta_n=\hat{f}^{-1}(\bar{q}_{\alpha,n}(x)/x)-f^{-1}(q_\alpha(x)/x)$ and $C_n=\frac{\mathrm{E}(\hat{F}(x,q_\alpha(x)))}{\mathrm{E}(\hat{F}(x))}-\hat{F}(q_\alpha(x)/x)$. The theorem follows if: (a) $\beta_n=o_{\mathrm{p}}(1)$; (b) $A_n=-\frac{1}{2}h_n^2\sigma_K^2\frac{\int_{g^{-1}([q_\alpha(x),g(x)])}f^{(1)}(\gamma,q_\alpha(x))\,\mathrm{d}\gamma}{F_X(x)}+o(h_n^2)$; (c) $(\frac{s_n(x)}{\hat{F}(x)})^{-1}\sqrt{n}C_n\xrightarrow{\mathrm{d}}N(0,1)$ where $s_n^2(x)=F(x,q_\alpha(x))-\frac{F(x,q_\alpha(x))^2}{F_X(x)}-2h_n\sigma_\kappa\int_{g^{-1}([q_\alpha(x),g(x)])}f(X,q_\alpha(x))\,\mathrm{d}X+o(h_n)$. (a) It suffices to show that $\hat{f}(\bar{q}_{\alpha,n}(x)/x)-f(q_\alpha(x)/x)=o_{\mathrm{p}}(1)$ for all $\alpha\in(0,1)$. Since $q_{\alpha,n}(x)-q_\alpha(x)=o_{\mathrm{p}}(1)$ it suffices to show that $\sup_{y\in G}|\hat{f}(y/x)-f(y/x)|=o_{\mathrm{p}}(1)$, where

$G \subset (0, g(x))$, G compact. Note that

$$\sup_{y\in G}|\hat{f}(y/x) - f(y/x)| \leqslant \frac{1}{\hat{F}(x)}\sup_{y\in G}\left|\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{Y_i - y}{h_n}\right)I(X_i \leqslant x) - \int_{g^{-1}([y,g(x)])} f(X,y)\,\mathrm{d}X\right| + \left|\frac{1}{F_X(x)} - \frac{1}{\hat{F}(x)}\right|\sup_{y\in G}\int_{g^{-1}([y,g(x)])} f(X,y)\,\mathrm{d}X.$$

By A1, $\sup_{y\in G}\int_{g^{-1}([y,g(x)])} f(X,y)\,\mathrm{d}X \leqslant B_f \int_{g^{-1}([y,g(x)])}\mathrm{d}X = \mathrm{O}(1)$ for all finite x, and since $\hat{F}(x)^{-1} - F_X(x)^{-1} = \mathrm{o_p}(1)$ the second term on the r.h.s. is $\mathrm{o_p}(1)$. We now establish that the first term on the r.h.s. is $\mathrm{o_p}(1)$. From Lemma 1 in Martins-Filho and Yao (2007a), if $nh_n^2 \to \infty$, $\sup_{y\in G}|s_{0,x}(y) - \mathrm{E}(s_{0,x}(y))| = \mathrm{O_p}((\frac{\ln(n)}{nh_n})^{1/2})$ where $s_{0,x}(y) = \frac{1}{nh_n}\sum_{i=1}^{n}K(\frac{Y_i - y}{h_n})I(X_i \leqslant x)$. Now, $\mathrm{E}(s_{0,x}(y)) = \int_{C_x}\int_{[-y/h_n,(g(X)-y)/h_n]} K(\gamma)f(X, y + h_n\gamma)\,\mathrm{d}\gamma\,\mathrm{d}X$ and by A3(c) $|\mathrm{E}(s_{0,x}(y)) - \int_{C_x}\int_{[-y/h_n,(g(X)-y)/h_n]} K(\gamma)f(X,y)\,\mathrm{d}\gamma\,\mathrm{d}X| \leqslant m_f h_n \int_{C_x}\int_{[-B_K,B_K]}|\gamma|K(\gamma)\,\mathrm{d}\gamma\,\mathrm{d}X = \mathrm{O}(h_n)$. Given that $y \in G \subset (0, g(x))$, there exists $N(x)$ such that for all $n > N(x)$ we have $\int_{C_x}\int_{[-y/h_n,(g(X)-y)/h_n]} K(\gamma)f(X,y)\,\mathrm{d}\gamma\,\mathrm{d}X = H_{1n}(x,y) + H_{2n}(x,y)$ where $H_{1n}(x,y) = \int_{g^{-1}([0,y])}\kappa(\frac{g(X)-y}{h_n})f(X,y)\,\mathrm{d}X$ and $H_{2n}(x,y) = \int_{g^{-1}([y,g(x)])}\kappa(\frac{g(X)-y}{h_n}) f(X,y)\,\mathrm{d}X$. Following the proof for Lemma 2, we obtain $\sup_{y\in G}|H_{1n}(x,y)| = \mathrm{o}(1)$ and $\sup_{y\in G} |H_{2n}(x,y) - \int_{g^{-1}([y,g(x)])} f(X,y)\,\mathrm{d}X| = \mathrm{o}(1)$. Consequently, we have $\sup_{y\in G}|\int_{C_x}\kappa(\frac{g(X)-y}{h_n})f(X,y)\,\mathrm{d}X - \int_{g^{-1}([y,g(x))} f(X,y)\,\mathrm{d}X| = \mathrm{o}(1)$ and also $\sup_{y\in G}|\frac{1}{nh_n}\sum_{i=1}^{n}K(\frac{Y_i-y}{h_n})I(X_i \leqslant x) - \int_{g^{-1}([y,g(x)])} f(X,y)\,\mathrm{d}X| = \mathrm{o_p}(1)$. (b) $A_n = (\mathrm{E}(\hat{F}(x)))^{-1}(A_{1n}(x) + A_{2n}(x))$ where $A_{1n}(x) = F(q_\alpha(x)/x)\mathrm{E}(\hat{F}(x)) - F(x, q_\alpha(x))$ and $A_{2n}(x) = F(x, q_\alpha(x)) - \mathrm{E}(\hat{F}(x, q_\alpha(x)))$. Since $\mathrm{E}(\hat{F}(x)) = F_X(x)$, $A_{1n}(x) = 0$. Given that $0 < \alpha < 1$, we have $0 < q_\alpha(x) < g(x)$ and from Lemma 1, $A_{2n}(x) = -\frac{1}{2}h_n^2\sigma_K^2\int_{g^{-1}([q_\alpha(x),g(x)])} f^{(1)}(X, q_\alpha(x))\,\mathrm{d}X + \mathrm{o}(h_n^2)$. Thus, $A_n = -\frac{1}{F_X(x)}\frac{h_n^2}{2}\sigma_K^2\int_{g^{-1}([q_\alpha(x),g(x)])} f^{(1)} (X, q_\alpha(x))\,\mathrm{d}X + \mathrm{o}(h_n^2)$. (c) $\sqrt{n}C_n = -\frac{1}{\hat{F}(x)}\sum_{i=1}^{n}Z_{in}$ where $Z_{in} = \frac{1}{\sqrt{n}}(\frac{1}{h_n}\int_{[0,q_\alpha(x)]} K(\frac{Y_i-\gamma}{h_n})I(X_i \leqslant x)\,\mathrm{d}\gamma - I(X_i \leqslant x) \frac{\mathrm{E}(\hat{F}(x,q_\alpha(x)))}{F_X(x)})$ with $\mathrm{E}(Z_{in}) = 0$, $s_n^2 = \sum_{i=1}^{n}\mathrm{E}(Z_{in}^2)$. By A1, $s_n^2 = s_{1n} + s_{2n} + s_{3n}$ where $s_{1n} = \mathrm{E}(\frac{1}{h_n}\int_{[0,q_\alpha(x)]} K(\frac{Y_i-\gamma}{h_n}) I(X_i \leqslant x)\,\mathrm{d}\gamma)^2$, $s_{2n} = \mathrm{E}(I(X_i \leqslant x))\frac{(\mathrm{E}(\hat{F}(x,q_\alpha(x)))^2}{F_X(x)^2}$ and $s_{3n} = -2\frac{\mathrm{E}(\hat{F}(x,q_\alpha(x))}{F_X(x)}\mathrm{E}(\frac{1}{h_n}\int_{[0,q_\alpha(x)]} K(\frac{Y_i-\gamma}{h_n})I(X_i \leqslant x)\,\mathrm{d}\gamma)$. From Lemma 1, $s_{1n} = F(x, q_\alpha(x)) - 2h_n\sigma_\kappa\int_{g^{-1}([q_\alpha(x),q(x)])} f(X, q_\alpha(x))\,\mathrm{d}X + \mathrm{o}(h_n)$, and $s_{2n} = F_X(x)^{-1}(F(x, q_\alpha(x)) + \sigma_K^2\frac{h_n^2}{2}\int_{g^{-1}([q_\alpha(x),q(x)])} f^{(1)}(X, q_\alpha(x))\,\mathrm{d}X + \mathrm{o}(h_n^2))^2$. Hence, $s_{2n} = F_X(x)^{-1}F(x, q_\alpha(x))^2 + \mathrm{o}(h_n)$ and $s_{3n} = -2s_{2n} = -2\frac{(F(x,q_\alpha(x)))^2}{F_X(x)} + \mathrm{o}(h_n)$ which gives $s_n^2(x) = F(x, q_\alpha(x)) - \frac{F(x,q_\alpha(x))^2}{F_X(x)} - 2h_n\sigma_\kappa\int_{g^{-1}([q_\alpha(x),g(x)])} f(X, q_\alpha(x))\,\mathrm{d}X + \mathrm{o}(h_n)$. By Liapounov's CLT $\sum_{i=1}^{n}\frac{Z_{in}}{s_n(x)} \xrightarrow{\mathrm{d}} N(0,1)$ provided that $\lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{E}(|\frac{Z_{in}}{s_n(x)}|^{2+\delta}) = 0$ for some $\delta > 0$. By the c_r inequality and given the order of $s_n^2(x)$, it suffices to show that $a_n = n^{-\delta/2}\mathrm{E}(|\frac{1}{h_n}\int_{[0,q_\alpha(x)]} K(\frac{Y_i-\gamma}{h_n}) I(X_i \leqslant x)\,\mathrm{d}\gamma|^{2+\delta}) = \mathrm{o}(1)$ and $b_n = n^{-\delta/2}\mathrm{E}(|I(X_i \leqslant x)\frac{\mathrm{E}(\hat{F}(x,q_\alpha(x)))}{F_X(x))}|^{2+\delta}) = \mathrm{o}(1)$. First, note that $a_n = n^{-\delta/2}\int_{C_x} \int_{[0,g(X)]} f(X,Y)\ (\int_{-Y/h_n}^{(q_\alpha(x)-Y)/h_n} K(\gamma)\,\mathrm{d}\gamma)^{2+\delta}\,\mathrm{d}Y\,\mathrm{d}X \leqslant n^{-\delta/2}\int_{C_x}\int_{[0,g(X)]} f(X,Y)\,\mathrm{d}Y\,\mathrm{d}X \to 0$ as $n \to \infty$. Second, note that $b_n = n^{-\delta/2}\mathrm{E}(I(X_i \leqslant x))\frac{(\mathrm{E}(\hat{F}(x,q_\alpha(x)))^{2+\delta})}{F_X(x)^{2+\delta}} \to 0$ since $\mathrm{E}(I(X_i \leqslant x)) = F_X(x) > 0$, and by Lemma 1 $\mathrm{E}(\hat{F}(x, q_\alpha(x))) \to F(x, q_\alpha(x))$. Hence, $(\frac{s_n(x)}{\hat{F}_X(x)})^{-1}\sqrt{n}C_n \xrightarrow{\mathrm{d}} N(0,1)$ since $\hat{F}(x) \xrightarrow{\mathrm{p}} F_X(x)$. □

Proof of Theorem 3. The proof is similar to that of Theorem 2 by using the Cramer–Wold device. □

Proof Theorem 4. (a) $q_{1,n}(x) \equiv \inf\{y \in \Re_+ : (nh_n)^{-1}\sum_{i=1}^{n}\int_0^y K(\frac{Y_i-\gamma}{h_n})\,\mathrm{d}\gamma I(X_i \leqslant x) = n^{-1}\sum_{i=1}^{n}I(X_i \leqslant x)\}$. Given $\min_{\{i:X_i\leqslant x\}} Y_i \geqslant h_n B_K$, there exists $N(x) \in R_+$ such that for all $n > N(x)$, we have that the equality (in the set) holds for all $y \geqslant \max_{\{i:X_i\leqslant x\}} Y_i + h_n B_K$, and it is false for all $y < \max_{\{i:X_i\leqslant x\}} Y_i + h_n B_K$. Hence, $q_{1,n}(x) = \max_{\{i:X_i\leqslant x\}} Y_i + h_n B_K$ for all $n > N(x)$. (b) The FDH estimator is defined as $\theta_{\mathrm{FDH}}(x) = \max_{\{i:X_i\leqslant x\}} Y_i$ and from Park et al. (2000) $n^{1/(d+1)}(q_1(x) - \theta_{\mathrm{FDH}}(x)) \xrightarrow{\mathrm{d}} \mathit{Weibull}(\mu_x^{d+1}, d+1)$. Hence, if $nh_n^{d+1} = \mathrm{O}(1)$, $n^{1/(d+1)}(q_1(x) - q_{1,n}(x) + h_n B_K \theta_{\mathrm{FDH}}(x)) \xrightarrow{\mathrm{d}} \mathit{Weibull}(\mu_x^{d+1}, d+1)$. □

Lemma 3. *Let* $\theta(y) = \int_0^b f^{(1)}(x, y)\,\mathrm{d}x$ *and* $\hat{\theta}(y) = \frac{1}{ng_{n1}}\sum_{i=1}^{n} K^{(1)}(\frac{y-Y_i}{g_{n1}})I(X_i \leqslant b)$. *Suppose third order partial derivatives of* $f(x, y)$ *with respect to* y-$f^{(3)}(x, y)$ *exists around* (x, y) *and the first order derivative of* $K(\cdot)$ – $K^{(1)}(\cdot)$ *satisfies the following conditions*: (K1) $\int_{-B_K}^{B_K} K^{(1)}(\psi)\,\mathrm{d}\psi = 0$; (K2) $\int_{-B_K}^{B_K} \psi K^{(1)}(\psi)\,\mathrm{d}\psi = -1$; (K3) $\int_{-B_K}^{B_K} \psi^2 K^{(1)}(\psi)\,\mathrm{d}\psi = 0$; (K4) $\int_{-B_K}^{B_K} \psi^3 K^{(1)}(\psi)\,\mathrm{d}\psi = C_K$; (K5) $\int_{-B_K}^{B_K} K^{2(1)}(\psi)\,\mathrm{d}\psi = C_{K1}$. Then, (a) $\mathrm{E}(\hat{\theta}(y) - \theta(y)) = -C_K \frac{g_n^2}{6}\int_0^b f^{(3)}(x, y)\,\mathrm{d}x + \mathrm{o}(g_n^2)$, (b) $V(\hat{\theta}(y)) = C_{K1}\frac{1}{ng_{n1}^3}\int_0^b f(x, y)\,\mathrm{d}x + \mathrm{o}((ng_n^3)^{-1})$, *and* (c) $AMIS\mathrm{E}(\hat{\theta}(y), g_n) = C_K^2 \frac{g_n^4}{36}\int(\int_0^b f^{(3)}(x, y)\,\mathrm{d}x)^2\,\mathrm{d}y + C_{K1}\frac{1}{ng_n^3}\int\int_0^b f(x, y)\,\mathrm{d}x\,\mathrm{d}y$.

Proof. By considering a third order Taylor expansion of $f(x, y)$ around y in the expression for $\mathrm{E}(\hat{\theta}(y))$, we obtain (a) in a straightforward manner given (K1)–(K4). (b) results from applying the LDC theorem to $V(\hat{\theta}(y))$ together with (K5). (c) follows directly from (a) and (b). □

References

Aragon, Y., Daouia, A., Thomas-Agnan, C., 2005. Nonparametric frontier estimation: a conditional quantile-based approach. Econometric Theory 21, 358–389.

Azzalini, A., 1981. A note on the estimation of a distribution function and quantiles by a kernel method. Biometrika 68, 326–328.

Bowman, A., Hall, P., Prvan, T., 1998. Bandwidth selection for the smoothing of distribution functions. Biometrika 85, 799–808.

Cazals, C., Florens, J.P., Simar, L., 2002. Nonparametric frontier estimation: a robust approach. Journal of Econometrics 106, 1–25.

Christensen, L.R., Greene, W.H., 1976. Economies of scale in U.S. electric power generation. Journal of Political Economy 84, 655–676.

Falk, M., 1985. Asymptotic normality of the kernel quantile estimator. The Annals of Statistics 13, 428–433.

Fan, J., Gijbels, I., 1995. Data driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. Journal of the Royal Statistical Society B 57, 371–394.

Farrell, M., 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society A 120, 253–290.

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

Girard, S., Jacob, P., 2004. Extreme values and kernel estimates of point processes boundaries. ESAIM: Probability and Statistics 8, 150–168.

Greene, W.H., 1990. A gamma-distributed stochastic frontier model. Journal of Econometrics 46, 141–163.

Hall, P., Park, B., Stern, S., 1998. On polynomial estimators of frontiers and boundaries. Journal of Multivariate Analysis 66, 71–98.

Hansen, B., 2004a. Bandwidth selection for nonparametric distribution estimation. Working paper, Department of Economics, University of Wisconsin, Madison.

Hansen, B., 2004b. Nonparametric estimation of smooth conditional distributions. Working paper, Department of Economics, University of Wisconsin, Madison.

Jones, M.C., 1994. On kernel density derivative estimation. Communications in Statistics—Theory and Methods 23 (8), 2133–2139.

Kincaid, D., Cheney, W., 1996. Numerical Analysis, second ed. Brooks/Cole Publishing Company, Pacific Grove, CA.

Knight, K., 2001. Limiting distributions of linear programming estimators. Extremes 4, 87–103.

Li, Q., Racine, J., 2005. Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. Journal of Business and Economic Statistics, in press.

Martins-Filho, C., Yao, F., 2007a. Nonparametric frontier estimation via local linear regression. Journal of Econometrics 141, 283–319.

Martins-Filho, C., Yao, F., 2007b. A smooth nonparametric conditional quantile frontier estimator: proofs and extended Monte Carlo simulations, Department of Economics, Oregon State University, ⟨http://oregonstate.edu/~martinsc/martins-filho-yao(06).pdf⟩.

Nadaraya, E., 1964. Some new estimate for distribution functions. Theory of Probability and Applications 15, 497–500.

Park, B.U., Simar, L., Weiner, C., 2000. The FDH estimator for productivity efficiency scores. Econometric Theory 16, 855–877.

Ruppert, D., Sheather, S.J., Wand, M.P., 1995. An effective bandwidth selector for local least squares regression. Journal of the American Statistical Association 90, 1257–1270.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London.

Simar, L., Wilson, P., 2006. Statistical inference in nonparametric frontier models: recent developments and perspectives. In: Fried, H., Lovell, C.A.K., Schmidt, S.S. (Eds.), The Measurement of Productive Efficiency, second ed. Oxford University Press, Oxford.

Yang, S., 1985. A smooth nonparametric estimator of a quantile function. Journal of the American Statistical Association 80, 1004–1011.