HAL open science

Nadaraya's estimates for large quantiles and free disposal support curves

Abdelaati Daouia, Laurent Gardes, Stéphane Girard

To cite this version:

Abdelaati Daouia, Laurent Gardes, Stéphane Girard. Nadaraya's estimates for large quantiles and free disposal support curves. I. Van Keilegom and P. Wilson,. Exploring research frontiers in contemporary statistics and econometrics, Springer, pp.1-22, 2012, 978-3-7908-2348-6. 10.1007/978-3-7908-2349-3_1 . hal-00528670v2

HAL Id: hal-00528670

https://hal.science/hal-00528670v2

Submitted on 24 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nadaraya's estimates for large quantiles and free disposal support curves

Abdelaati Daouia* Laurent Gardes† Stéphane Girard‡

April 24, 2013

Abstract

A new characterization of partial boundaries of a free disposal multivariate support, lying near the true support curve, is introduced by making use of large quantiles of a simple transformation of the underlying multivariate distribution. Pointwise empirical and smoothed estimators of the full and partial support curves are built as extreme sample and smoothed quantiles. The extreme-value theory holds then automatically for the empirical frontiers and we show that some fundamental properties of extreme order statistics carry over to Nadaraya's estimates of upper quantile-based frontiers. The benefits of the new class of partial boundaries are illustrated through simulated examples and a real data set, and both empirical and smoothed estimates are compared via Monte Carlo experiments. When the transformed distribution is attracted to the Weibull extreme-value type distribution, the smoothed estimator of the full frontier outperforms frankly the sample estimator in terms of both bias and Mean-Squared Error, under optimal bandwidth. In this domain of attraction, Nadaraya's estimates of extreme quantiles might be superior to the sample versions in terms of MSE although they have a higher bias. However, smoothing seems to be useless in the heavy tailed case.

Key words : large quantile, partial boundary, kernel estimator, extreme values.
AMS 2000 subject classification: 62G08; 62G32; 62E20

1 Main results

Let $(X, Y), (X_1, Y_1), (X_2, Y_2) \ldots$ be independent random vectors from a common probability distribution on $\mathbb{R}_+^p \times \mathbb{R}_+$ whose support boundary is assumed to be nondecreasing. For $x \in \mathbb{R}_+^p$ such that $\mathbb{P}(X \leq x) > 0$, the graph of the frontier function[1]

$$\xi_1(x) = \inf\{y \geq 0 : \mathbb{P}(Y \leq y | X \leq x) = 1\}$$

*Toulouse School of Economics (GREMAQ), University of Toulouse, Manufacture des Tabacs, Aile J.J. Laffont, 21 Allée de Brienne 31000 Toulouse, France (daouia@cict.fr)

†Team Mistis, INRIA Rhône-Alpes and Laboratoire Jean Kuntzmann, 655 avenue de l'Europe, Montbonnot, 38334 Saint-Ismier cedex, France (laurent.gardes@inrialpes.fr)

‡Team Mistis, INRIA Rhône-Alpes and Laboratoire Jean Kuntzmann, 655 avenue de l'Europe, Montbonnot, 38334 Saint-Ismier cedex, France (stephane.girard@inrialpes.fr)

[1]For two vectors x_1 and x_2 with $x_1 \leq x_2$ componentwise, $\xi_1(\cdot)$ satisfies $\xi_1(x_1) \leq \xi_1(x_2)$.

coincides with the monotone surface of the joint support Ψ of (X, Y) (Cazals, Florens and Simar, 2002). As a matter of fact, the graph of $\xi_1(\cdot)$ is the lowest nondecreasing curve larger than or equal to the upper frontier of Ψ. In applied econometrics for instance, the support Ψ is interpreted as the set of all feasible production units, *i.e.*, $(x, y) \in \Psi$ in a certain sector of technology if and only if it is possible for a given firm to produce a quantity y of goods by making use of a quantity x of resources. The production set Ψ is by construction free disposal. This means that its optimal frontier which represents the set of the most efficient firms is nondecreasing. The free disposal hull (FDH) estimator of $\xi_1(\cdot)$ is given by

$$\xi_{1,n}(x) = \max\{Y_i | i : X_i \leq x\}$$

(Deprins, Simar and Tulkens, 1984). The FDH frontier is clearly the lowest step and monotone curve which envelopes all the data points (X_i, Y_i) and so it is very non-robust to extreme observations. To reduce this vexing defect, instead of estimating the frontier of the support Ψ, Aragon *et al.* (2005) have suggested to estimate a partial boundary of Ψ of order $\alpha \in (0, 1)$ lying near its true full boundary. The frontier function $\xi_1(x)$ being the quantile function of order one of the distribution of Y given $X \leq x$, they rather proposed to estimate the αth quantile function of this non-standard conditional distribution

$$q_\alpha(x) = \inf\{y \geq 0 : \mathbb{P}(Y \leq y | X \leq x) \geq \alpha\}.$$

The resulting sample quantile function, obtained by plugging the empirical version of the conditional distribution function may suffer from a lack of efficiency due to the large variation of the extreme observations involved in its construction. A smoothed variant $\hat{q}_\alpha(x)$ of this sample estimator may be then preferable as shown in Martins-Filho and Yao (2008), where $\mathbb{P}(Y \leq y | X \leq x)$ is estimated by

$$\hat{F}_{Y^x}(y) = \sum_{i=1}^{n} \mathbb{1}(X_i \leq x) H\left((y - Y_i)/h\right) / \sum_{i=1}^{n} \mathbb{1}(X_i \leq x),$$

with $h = h_n \to 0$, $H(y) = \int_{-\infty}^{y} K(u)du$ and $K(\cdot)$ being a density kernel. However no attention was devoted to the limit distribution of $\hat{q}_\alpha(x) := \hat{F}_{Y^x}^{-1}(\alpha) = \inf\{y \geq 0 : \hat{F}_{Y^x}(y) \geq \alpha\}$ when it estimates the optimal boundary itself. Daouia *et al.* (2009) have addressed this problem by specifying the asymptotic distribution of the smoothed α-frontier for fixed orders $\alpha \in (0, 1]$ as well as for sequences $\alpha = \alpha_n$ tending to one as $n \to \infty$.

It is important to note that the distribution of Y being conditioned by $X \leq x$, the estimation of the corresponding quantiles does not require a smoothing procedure in x which would be the case if the distribution was conditioned by $X = x$, see *e.g.* Girard and Menneteau (2005), Girard and Jacob (2008) or Daouia *et al.* (2011). It should be also

clear that although the simple nature of the conditioning $X \leq x$, it requires more powerful techniques of proof than the unconditional quantile setting. Our main contribution in this note is to get rid of this conditioning by exploiting the fact that

$$\xi_1(x) = \inf\{y \geq 0 : \mathbb{P}(Z^x \leq y) = 1\},$$

where $Z^x = Y \mathbb{1}(X \leq x)$. This simple formulation of the monotone frontier function was pointed out by Daouia *et al.* (2010, Appendix). Note also that the FDH estimator coincides with the maximum of the random variables $Z_i^x = Y_i \mathbb{1}(X_i \leq x)$, $i = 1, \ldots, n$. Moreover, given that the interest is also on the estimation of a concept of a partial frontier well inside the sample but near from the optimal boundary of Ψ, a natural idea is to define the alternative simple αth frontier function

$$\xi_\alpha(x) := F_{Z^x}^{-1}(\alpha) = \inf\{y \geq 0 : F_{Z^x}(y) \geq \alpha\},$$

where $F_{Z^x}(y) = \mathbb{P}(Z^x \leq y)$. In the context of productivity and efficiency analysis, when the performance of firms is measured in terms of their distance from partial frontiers rather than the full frontier, the use of the αth production frontier $q_\alpha(x)$ as a benchmark can be criticized for its divergence from the optimal frontier as x increases. Instead, Wheelock and Wilson (2008) favored the use of a hyperbolic unconditional variant of $q_\alpha(x)$. Our partial unconditional quantile-type frontier $\xi_\alpha(x)$ provides a more attractive alternative to reduce the vexing defect of the conditional version $q_\alpha(x)$ because of its construction and conceptual simplicity. The benefits of using this new class of partial support curves are demonstrated in Subsection 3.1 via two examples in the cases where $F_{Z^x}(\cdot)$ is attracted to the Weibull and Fréchet extreme value type distributions.

A natural estimator of $\xi_\alpha(x)$ is given by the sample quantile

$$\xi_{\alpha,n}(x) := F_{Z^x,n}^{-1}(\alpha) = \inf\{y \geq 0 : F_{Z^x,n}(y) \geq \alpha\},$$

where $F_{Z^x,n}(y) = n^{-1} \sum_{i=1}^n \mathbb{1}(Z_i^x \leq y)$. Therefore, the extreme-value theory holds automatically when $\alpha = 1$ and when letting $\alpha = \alpha_n \uparrow 1$ as $n \to \infty$, which is not the case for previous concepts of partial support curves such as those of Cazals et al (2002), Girard and Jacob (2004), Aragon *et al.* (2005), Wheelock and Wilson (2008) and Martins-Filho and Yao (2008). Following Nadaraya (1964), an alternative estimator to $\xi_{\alpha,n}(x)$ is given by the αth quantile

$$\hat{\xi}_\alpha(x) := \hat{F}_{Z^x}^{-1}(\alpha) = \inf\{y \geq 0 : \hat{F}_{Z^x}(y) \geq \alpha\} \tag{1}$$

of the kernel-smoothed empirical distribution function $\hat{F}_{Z^x}(y) = n^{-1} \sum_{i=1}^n H\left((y - Z_i^x)/h\right)$ based on a sequence of bandwidths h and an integrated kernel $H(\cdot)$. In the ordinary framework where the order α is a fixed constant in $(0, 1)$, Azzalini (1981) established a second-order

approximation to the mean squared error of the smoothed quantile $\hat{\xi}_\alpha(x)$ under optimal h, which makes $\hat{\xi}_\alpha(x)$ more efficient than the sample version $\xi_{\alpha,n}(x)$. The smoothed estimator may also be preferable to the sample one for the following additional respect: the construction of asymptotic confidence intervals for $\xi_\alpha(x)$ using the asymptotic normality of $\xi_{\alpha,n}(x)$ requires the estimation of the derivative $F'_{Z^x}(\xi_\alpha(x))$, whereas smoothing gives a naturally derived estimator of this quantile density function. Nadaraya (1964) has proved under mild conditions the asymptotic normality of $\hat{\xi}_\alpha(x)$ when the order α is fixed in $(0,1)$. In the present paper, we rather concentrate in Section 2 on specifying the asymptotic distributions of Nadaraya's estimates $\hat{\xi}_\alpha(x)$ when $\alpha = \alpha_n \to 1$ at different rates as $n \to \infty$, and we verify whether the benefits of smoothing are still valid when considering these extreme quantiles. This does not seem to have been appreciated before in the literature. Theorem 1 characterizes possible limit distributions of $\hat{\xi}_1(x)$ and Theorem 2 discusses moment convergence. When $n(1-\alpha_n)$ is a constant, Theorem 3 shows that $\hat{\xi}_{\alpha_n}(x)$ converges with the same centering and scaling as $\hat{\xi}_1(x)$ to a different extreme value distribution. However, when $n(1-\alpha_n) \to \infty$, Theorem 4 establishes the asymptotic normality of $\hat{\xi}_{\alpha_n}(x)$ as an estimator of $\xi_{\alpha_n}(x)$.

As a matter of fact, in this context where the underlying distribution function $F_{Z^x}(\cdot)$ has a jump at the left-endpoint of its support, we show by using simple arguments that the smoothed maximum $\hat{\xi}_1(x)$ is equal to a deterministic translation of the sample maximum $\xi_{1,n}(x)$ for all $n \geq 1$. Likewise, it turns out that a smoothed quantile of the form $\hat{\xi}_{(n-k+1)/n}(x)$ is within a fixed multiple of the bandwidth of the sample quantile $\xi_{(n-k+1)/n,n}(x)$ for all n large enough. As an immediate consequence, the asymptotic theory of the sample extremes $\xi_{1,n}(x)$ and $\xi_{(n-k+1)/n,n}(x)$ carry over to the smoothed variants.

Our Monte Carlo exercise, provided in Subsection 3.2, shows that the smoothed FDH function $\hat{\xi}_1(x)$ is a remarkable bias-corrected estimator of the frontier function $\xi_1(x) < \infty$. It outperforms frankly the sample FDH estimator $\xi_{1,n}(x)$ in terms of both bias and mean-squared error (MSE). Moreover, an explicit expression of the optimal bandwidth is derived in this case by minimizing the asymptotic MSE. Simulations seem to indicate also that, when $F_{Z^x}(\cdot)$ belongs to the maximum domain of attraction of Weibull and the bandwidth is chosen appropriately, the smoothed extreme quantile function $\hat{\xi}_{(n-k+1)/n}(x)$ is superior to the empirical version in terms of MSE although it has a higher positive bias. This result is similar to what happens in the ordinary framework. Although the naive extreme sample quantile $\xi_{(n-k+1)/n,n}(x)$ might not be so efficient, it has the advantage of not requiring the choice of the bandwidth which is not addressed here. In the heavy tailed case, it appears that $\hat{\xi}_{(n-k+1)/n}(x)$ achieves at most the same performance as the empirical version $\xi_{(n-k+1)/n,n}(x)$ in terms of both bias and MSE and so, smoothing seems to be useless in this case.

The limit theorems in Section 2 are provided in the general setting where the distribution

function of Z^x, or equivalently the conditional distribution function of Y given $X \leq x$, is attracted to the three Fisher-Tippett extreme value type distributions, whereas the previous results of *e.g.* Martins-Filho and Yao (2008) and Daouia *et al.* (2009) only cover the domain of attraction of Weibull. These results and their proofs, postponed to Appendix, are also extensions of some results and techniques described in Daouia *et al.* (2010) and Daouia *et al.* (2009). Illustrations on how the new class of unconditional quantile-based frontiers $\{\xi_\alpha(\cdot), \hat{\xi}_\alpha(\cdot)\}$ differs from the class of conditional quantile-type frontiers $\{q_\alpha(\cdot), \hat{q}_\alpha(\cdot)\}$ are provided in Section 3 through simulated examples and a real data set.

2 Limit theorems

To simplify the notation we write ξ_α and $\hat{\xi}_\alpha$, respectively, for $\xi_\alpha(x)$ and $\hat{\xi}_\alpha(x)$ throughout this section. We first show that the smooth estimator $\hat{\xi}_1$ of the endpoint ξ_1 has a similar asymptotic behaviour as the maximum $Z_{(n)} = \max\{Z_1, \ldots, Z_n\}$ under the assumption that

(A1) $\int_{-c}^{c} K(u)du = 1$ for some constant $c > 0$.

This is a standard condition in nonparametric estimation, which is satisfied by commonly used density kernels such as Biweight, Triweight, Epanechnikov, etc.

Theorem 1. *Assume that (A1) holds.*

(i) We have $\hat{\xi}_1 = Z_{(n)} + hc$, for all $n \geq 1$.

(ii) Suppose there exist $a_n > 0$, $b_n \in \mathbb{R}$, $n \geq 1$ such that

$$\mathbb{P}[a_n^{-1}(\hat{\xi}_1 - b_n) \leq z] \longrightarrow G(z) \quad \textit{as} \quad n \to \infty, \tag{2}$$

where G is assumed nondegenerate. Then G has one of the three forms

$$\begin{aligned}
&\textit{Fréchet}: \quad G(z) = \Phi_\rho(z) = \begin{cases} 0 & z < 0 \\ \exp\{-z^{-\rho}\} & z \geq 0 \end{cases} \quad \textit{for some} \quad \rho > 0 \\
&\textit{Weibull}: \quad G(z) = \Psi_\rho(z) = \begin{cases} \exp\{-(-z)^{\rho}\} & z < 0 \\ 1 & z \geq 0 \end{cases} \quad \textit{for some} \quad \rho > 0 \\
&\textit{Gumbel}: \quad G(z) = \Lambda(z) = \exp\{-e^{-z}\}, \ z \in \mathbb{R}.
\end{aligned}$$

It is clear from Theorem 1(i) that (2) holds if and only if F_Z belongs to the maximum domain of attraction[2] of an extreme value distribution $G \in \{\Phi_\rho, \Psi_\rho, \Lambda\}$. Then the characterization of a_n and b_n can be easily deduced from the classical theory of limit laws for

[2]We write $F_Z \in$ DA(G) if there exist normalizing constants $a_n > 0$, $c_n \in \mathbb{R}$ such that $a_n^{-1}(Z_{(n)} - c_n) \xrightarrow{d} G$.

maxima (see *e.g.* Resnick 1987) in conjunction with Theorem 1(i). Putting $\bar{F}_Z = 1 - F_Z$, it is well-known that $F_Z \in \mathrm{DA}(\Psi_\rho)$ iff $\xi_1 < \infty$ and[3] $\bar{F}_Z(\xi_1 - \frac{1}{t}) \in \mathrm{RV}_{-\rho}$, $t \to \infty$; in this case we may set $a_n = \xi_1 - \xi_{1-1/n}$ and $b_n = \xi_1 + hc$. Likewise $F_Z \in \mathrm{DA}(\Phi_\rho)$ iff $\bar{F}_Z \in \mathrm{RV}_{-\rho}$; in this case $\xi_1 = \infty$ and (a_n, b_n) can be taken equal to $(\xi_{1-1/n}, hc)$. Finally $F_Z \in \mathrm{DA}(\Lambda)$ iff there exists a strictly positive function g on $\mathbb{R}$ such that $\bar{F}_Z(t + g(t)z)/\bar{F}_Z(t) \to e^{-z}$ as $t \uparrow \xi_1$, for every $z \in \mathbb{R}$; in this case the constants (a_n, b_n) can be taken equal to $(g(\xi_{1-1/n}), \xi_{1-1/n} + hc)$.

On the other hand, by making use of Theorem 1(i), it is easy to show that the convergence in distribution (2) of the smoothed maximum $\hat{\xi}_1$ implies the convergence of moments in the three cases $G \in \{\Phi_\rho, \Psi_\rho, \Lambda\}$ under some condition on the left tail of F_Z.

Theorem 2. *Let (A1) and* (2) *hold and denote by* $\Gamma^{(k)}$ *the kth derivative of the gamma function* Γ.

(i) If $G = \Phi_\rho$ *in* (2) *with* $(a_n, b_n) = (\xi_{1-1/n}, hc)$ *and* $\int_{-\infty}^0 |z|^k F_Z(dz) < \infty$ *for some integer* $0 < k < \rho$, *then*

$$\lim_{n\to\infty} \mathbb{E}\{a_n^{-1}(\hat{\xi}_1 - hc)\}^k = \int_{\mathbb{R}} z^k \Phi_\rho(dz) = \Gamma(1 - k/\rho).$$

(ii) If $G = \Psi_\rho$ *with* $(a_n, b_n) = (\xi_1 - \xi_{1-1/n}, \xi_1 + hc)$ *and* $\int_{-\infty}^{\xi_1} |z|^k F_Z(dz) < \infty$ *for some integer* $k > 0$, *then*

$$\lim_{n\to\infty} \mathbb{E}\{a_n^{-1}(\hat{\xi}_1 - \xi_1 - hc)\}^k = \int_{-\infty}^0 z^k \Psi_\rho(dz) = (-1)^k \Gamma(1 + k/\rho).$$

(iii) If $G = \Lambda$ *with* $(a_n, b_n) = (g(\xi_{1-1/n}), \xi_{1-1/n} + hc)$ *and* $\int_{-\infty}^0 |z|^k F_Z(dz) < \infty$ *for some integer* $k > 0$, *then*

$$\lim_{n\to\infty} \mathbb{E}\{a_n^{-1}(\hat{\xi}_1 - b_n)\}^k = \int_{\mathbb{R}} z^k \Lambda(dz) = (-1)^k \Gamma^{(k)}(1).$$

When estimating the endpoint $\xi_1 < \infty$ of $F_Z \in \mathrm{DA}(\Psi_\rho)$, an optimal value of h can be derived by minimizing the asymptotic mean-squared error of $\hat{\xi}_1$. By making use of Theorem 2(ii), it is not hard to check that the optimal bandwidth is given by

$$h_{opt} = a_n (c\rho)^{-1} \Gamma(1/\rho). \tag{3}$$

Next we show that if $\hat{\xi}_1$ converges in distribution, then $\hat{\xi}_{1-k/n}$ converges in distribution as well, with the same centering and scaling, but a different limit distribution.

[3] A measurable function $\ell : \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying at ∞ with index γ (written $\ell \in \mathrm{RV}_\gamma$) if $\lim_{t\to\infty} \ell(tx)/\ell(t) = x^\gamma$ for all $x > 0$.

Theorem 3. *Assume that (A1) holds. If* $a_n^{-1}(\hat{\xi}_1 - b_n) \xrightarrow{d} G$ *and* $a_n^{-1}h \to 0$*, then for any integer* $k \geq 0$*,*

$$a_n^{-1}(\hat{\xi}_{\frac{n-k}{n}} - b_n) \xrightarrow{d} \mathbb{G} \quad as \quad n \to \infty$$

for the distribution function $\mathbb{G}(z) = G(z)\sum_{i=0}^{k}(-\log G(z))^i/i!$*.*

For the condition $a_n^{-1}h \to 0$ to be satisfied in the case $G = \Phi_\rho$, one only needs to suppose for instance that $\lim_{n\to\infty} h < \infty$. However, for the case $G = \Psi_\rho$, the condition $a_n^{-1}h \to 0$ holds if $h/(\xi_1 - \xi_{1-1/n}) \to 0$. The case $G = \Lambda$ is less flexible since the characterization of the normalization constant $a_n = g(\xi_{1-1/n})$ is not as explicit here than in the cases $G \in \{\Phi_\rho, \Psi_\rho\}$ (see *e.g.* Resnick (1987, p.38) for more details). It should be also clear that for $\hat{\xi}_\alpha$ to converge to the extreme-value distribution $\mathbb{G}$, it suffices to choose the sequence $\alpha = \alpha_n \uparrow 1$ such that $n(1-\alpha) = k$, with $k < n$ being an integer, whereas for $\hat{\xi}_\alpha$ to have an asymptotic normal distribution, we show in the next theorem that it suffices to choose $\alpha \to 1$ slowly so that $n(1-\alpha) \to \infty$.

The three Fisher-Tippett extreme-value distributions can be defined as a one-parameter family of types

$$G_\gamma(z) = \begin{cases} \exp\{-(1+\gamma z)^{-1/\gamma}\}, & \gamma \neq 0, \quad 1+\gamma z > 0 \\ \exp\{-e^{-z}\} & \gamma = 0, \quad z \in \mathbb{R}, \end{cases}$$

where γ is the so-called extreme-value index (see, *e.g.*, Beirlant, Teugels, Goegebeur and Segers (2004)). The heavy-tailed case $F_Z \in \mathrm{DA}(\Phi_\rho)$ corresponds to $\gamma > 0$ and $\rho = 1/\gamma$. For $\gamma = 0$, it is clear that $G_\gamma = \Lambda$. The case $F_Z \in \mathrm{DA}(\Psi_\rho)$ corresponds to $\gamma < 0$ and $\rho = -1/\gamma$. In either case we give in the next theorem asymptotic confidence intervals for high quantiles ξ_α by imposing the extra condition that

(A2) The derivative U' of $U(t) = \xi_{1-1/t}$ exists so that it satisfies $U' \in \mathrm{RV}_{\gamma-1}$.

As pointed out in Dekkers and de Haan (1989), the assumption $U' \in \mathrm{RV}_{\gamma-1}$ on the inverse function $U(t) = (1/(1-F_Z))^{-1}(t)$ is equivalent to $F_Z' \in \mathrm{RV}_{-1-1/\gamma}$ for $\gamma > 0$, $F_Z'(\xi_1 - 1/t) \in \mathrm{RV}_{1+1/\gamma}$ for $\gamma < 0$ and $1/F_Z'$ is Γ-varying for $\gamma = 0$ (for the Γ-Variation, see *e.g.* Resnick 1987, p.26).

Theorem 4. *Given (A1) and (A2),*

$$\sqrt{k}\left(\hat{\xi}_{\frac{n-k+1}{n}} - \xi_{\alpha_n}\right) / \left(\hat{\xi}_{\frac{n-k+1}{n}} - \hat{\xi}_{\frac{n-2k+1}{n}}\right)$$

is asymptotically normal with mean zero and variance $2^{2\gamma}\gamma^2/(2^\gamma - 1)^2$*, provided that* $\alpha_n \uparrow 1$*,* $n(1-\alpha_n) \to \infty$ *and the integer part* k *of* $n(1-\alpha_n)$ *satisfies* $hk^{3/2}\{nU'(n/k)\}^{-1} \to 0$ *as* $n \to \infty$*.*

Note that in the case $\gamma = 0$, the asymptotic variance is understood to be the limit $\{1/\log 2\}^2$ obtained as $\gamma \to 0$. This theorem enables one to construct an asymptotic confidence interval for ξ_{α_n} by replacing the tail index γ in the asymptotic variance with a consistent estimator. One can use for example the moment's estimator introduced by Dekkers, Einmahl and de Haan (1989) or the recent proposal by Segers (2005). Note also that in frontier and efficiency analysis, econometric considerations often lead to the assumption that the joint density of the random vector $(X, Y) \in \mathbb{R}_+^{p+1}$ has a jump at its support boundary, which corresponds to the case where γ is known and equal to $-1/(p+1)$ as established in Daouia *et al.* (2010).

3 Numerical illustrations

We provide in this section some modest illustrations in the context of frontier analysis, *i.e.*, when $Z = Z^x$ and $\hat{\xi}_\alpha = \hat{\xi}_\alpha(x)$. In this case, it is important to note that the bandwidth h, the normalizing sequences (a_n, b_n), the extreme value index γ or equivalently the tail index ρ, the order α_n and the sequence k_n should depend on the fixed level $x \in \mathbb{R}_+^p$. We do not enter here into the question of how to choose in an optimal way h, α_n and k_n. Deriving asymptotically optimal values of these parameters is a tedious matter. Using for instance (3) calls for selection of subsidiary smoothing parameters (using plug-in methods requires explicit estimation of the spacing a_n and the tail index ρ, which demands optimal selection of the amount of extreme data involved in each estimate, etc). Such complexity is arguably not justified. Instead, we suggest an approximate empirical method as follows. We tune the bandwidth h involved in $\hat{\xi}_\alpha(x) = \hat{F}_{Z^x}^{-1}(\alpha)$ so that approximately a reasonable percentage $\lambda\%$ of the data points $Z_1^x, \cdots, Z_n^x$ fall into the support of $u \mapsto K((z-u)/h)$. In case of kernels with support $[-1, 1]$, as Triweight and Epanechnikov kernels, we use the explicit formula

$$h_x = \frac{\lambda}{200}(\max_{i=1,\cdots,n} Z_i^x - \min_{i=1,\cdots,n} Z_i^x).$$

Our method itself requires selection of a smoothing parameter λ, but it has the advantage to be very simple to interpret and to implement, particularly in the difficult context of nonparametric curve estimation. The same rule can be applied to the estimator $\hat{q}_\alpha(x) = \hat{F}_{Y^x}^{-1}(\alpha)$ whose computation is similar to $\hat{\xi}_\alpha(x)$. Indeed, similarly to $\hat{F}_{Z^x}(y)$ we have $\hat{F}_{Y^x}(y) = (1/N_x)\sum_{i=1}^{N_x} H\left((y - Y_i^x)/h\right)$, where $N_x = \sum_{i=1}^n \mathbb{1}(X_i \leq x)$ and $Y_1^x, \cdots, Y_{N_x}^x$ are the Y_i's such that $X_i \leq x$. As a matter of fact, $Y_1^x, \cdots, Y_{N_x}^x$ are the N_x largest statistics of the sample $(Z_1^x, \cdots, Z_n^x)$. Note also that the two families of nonparametric (unconditional and conditional) quantile-based partial frontiers $\hat{\xi}_\alpha(\cdot)$ and $\hat{q}_\alpha(\cdot)$ coincide for $\alpha = 1$, but they differ from one another when $\alpha < 1$. This difference is illustrated through a real data set and two

simulated examples in the cases where the support boundary is finite and infinite.

For our practical computations in Subsection 3.1, the smooth estimators will be evaluated only for fixed extreme orders α by using a Triweight kernel K and for a grid of values of λ. Subsection 3.2 provides a comparison between the sample extreme frontiers and their smoothed versions via Monte Carlo experiments.

3.1 Illustrative examples on one sample

3.1.1 Case of a finite support boundary

We choose (X, Y) uniformly distributed over the support $\Psi = \{(x, y)|0 \leq x \leq 1, 0 \leq y \leq x\}$. In this case, the true frontier function is $\xi_1(x) = q_1(x) = x$ and the class of conditional quantile-based frontiers $q_\alpha(x) = x(1 - \sqrt{1-\alpha})$ is different from our class of unconditional quantile-type frontiers $\xi_\alpha(x) = \max\{0, x - \sqrt{1-\alpha}\}$, for $\alpha \in (0, 1]$. Both partial order-α frontiers are graphed in Figure 1 for some large values of $\alpha = 0.9, 0.95, 0.99$.

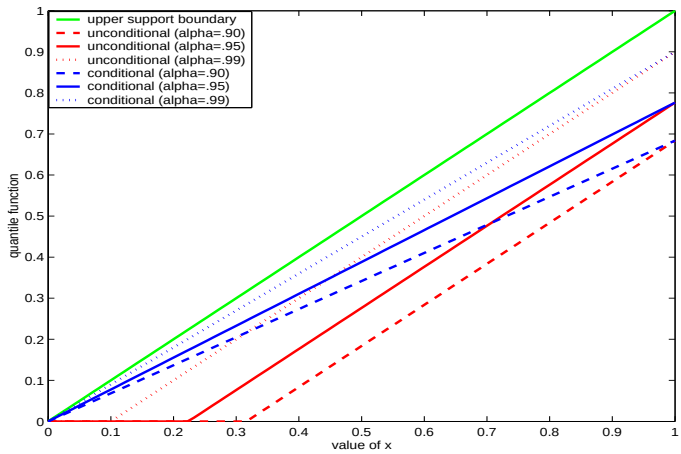

Figure 1: *The full frontier in green line and the partial frontiers $q_\alpha(\cdot)$ in blue lines and $\xi_\alpha(\cdot)$ in red lines. In dashed lines $\alpha = .9$, in solid lines $\alpha = .95$, in dotted lines $\alpha = .99$.*

Note that in traditional applied econometrics, the distance from the full support frontier is used as a benchmark to measure the production performance of firms. The economic efficiency can also be measured in terms of partial frontiers as suggested recently in the econometric literature to avoid non-robustness of the envelopment nonparametric frontier estimators. It is clear that the partial frontiers $\{(x, q_\alpha(x)) : 0 \leq x \leq 1\}$ of the support Ψ diverge from the support boundary as x increases and so, measuring efficiency relative to these curves may result in misleading efficiency measures. To reduce this defect, Wheelock and Wilson (2008) favored the use of a hyperbolic unconditional variant of $q_\alpha(x)$. The new partial frontiers $\xi_\alpha(x)$ parallel the full frontier $\xi_1(x)$ providing thus a simple alternative for measuring partial efficiencies without recourse to the hyperbolic framework. However, this

desirable property is valid only for input factors x ranging from the $(1-\alpha)$th quantile of the marginal distribution of X to its endpoint. It is easy to see that the value of $\xi_\alpha(x)$ is zero whenever $\mathbb{P}(X \not\leq x) \geq \alpha$, but the class $\{q_\alpha(x)\}$ does not take necessarily advantage from this drawback of $\{\xi_\alpha(x)\}$. Indeed, given that the interest is in estimating partial frontiers ξ_α and q_α lying close to the full support boundary, the order α shall be selected large enough in such a way that the estimates of ξ_α and q_α capture the shape of the sample's upper boundary without enveloping all the data points[4]. For such a choice of $\alpha \uparrow 1$, the $(1-\alpha)$th quantile of the distribution of X should be very small and so, the shortcoming $\hat{\xi}_\alpha(x_i) = 0$ is expected to hold only for a very few observations (x_i, y_i) at the left border of the sample. For these few observations with too small inputs-usage x_i, the estimates $\hat{q}_\alpha(x_i)$ of $q_\alpha(x_i)$ are expected by construction to coincide with the non-robust envelopment FDH estimates (as illustrated below in Figures 2, 4 and 5), which goes against the concept of partial frontier modeling. Examples are also provided below in Figure 8 where only $\hat{q}_\alpha(\cdot)$ suffers from left border defects, whereas $\hat{\xi}_\alpha(\cdot)$ is clearly the winner.

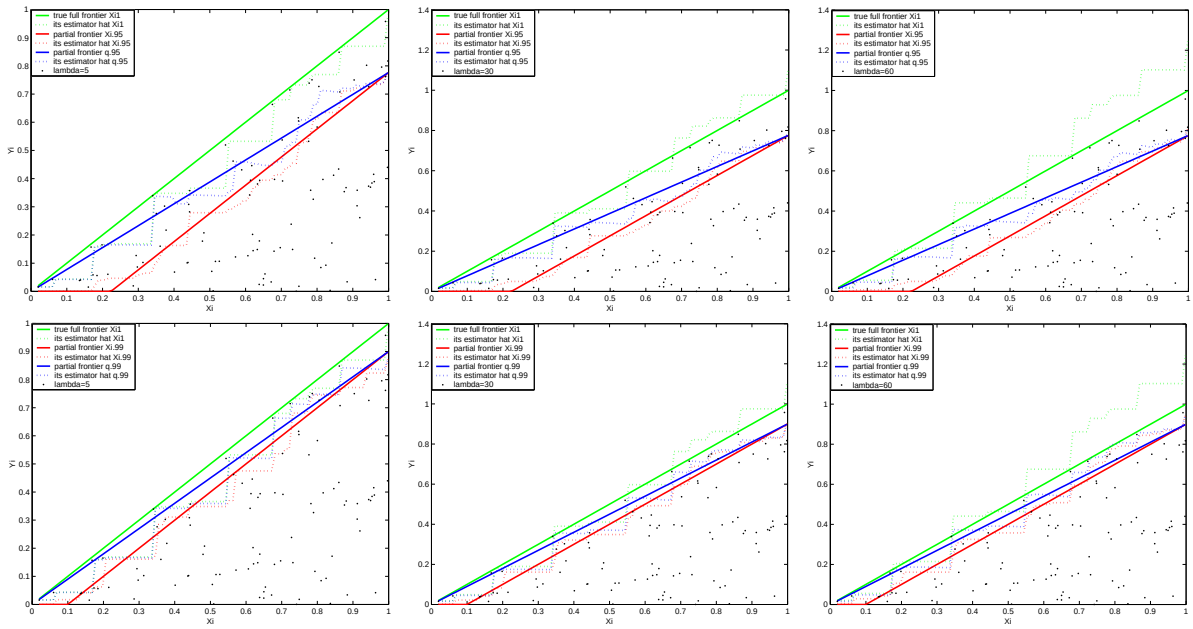

Figure 2: *In solid lines the frontiers ξ_1 (green), ξ_α (red) and q_α (blue). In dotted lines the estimators $\hat{\xi}_1$ (green), $\hat{\xi}_\alpha$ (red) and $\hat{q}_\alpha$ (blue). $n = 100$ and from left to right $\lambda = 5, 30, 60$. From top to bottom $\alpha = .95, .99$.*

Figure 2 (top) depicts the true α-frontiers $q_\alpha(x)$ and $\xi_\alpha(x)$ with estimated smooth frontiers $\hat{q}_\alpha(x)$ and $\hat{\xi}_\alpha(x)$ for α ranging over $\{0.95, 1\}$ for a simulated data set of size $n = 100$. The

[4]Once a reasonable large value of α is picked out, the idea in practice is then to interpret the observations left outside the αth frontier estimator as highly efficient and to assess the performance of the points lying below the estimated partial frontier by measuring their distances from this frontier in the output-direction.

three pictures correspond to the values $5, 30, 60$ of the parameter λ. The kernel estimator $\hat{\xi}_1$ of the full frontier ξ_1 is clearly sensible to the choice of the smoothing parameter λ. The worse behavior of this estimator for too large and too small values of λ is explained by Theorem 1(i) which states that $\hat{\xi}_1(x)$ is nothing else than a shifted value of the conventional FDH estimator. The choice of h_x according to a high percentage $\lambda\%$ generates a very large bandwidth, which may result in over-estimations of the true frontier. On the other hand, when $\lambda \downarrow 0$ we have $h_x \downarrow 0$, and so $\hat{\xi}_1(x)$ converges to the FDH estimator which underestimates the frontier $\xi_1(x)$. In contrast, $\hat{\xi}_{.95}$ and $\hat{q}_{.95}$ are less sensible to the choice of λ in this particular case. Figure 2 (bottom) corresponds to the same exercise with α ranging over $\{0.99, 1\}$. Again we obtain the same results as before. Reasonable values of the smooth frontier $\hat{\xi}_1$ require, via computer simulation, the choice of a moderate parameter λ, say $\lambda = 20, 25, 30, 35$.

3.1.2 Case of an infinite support boundary

We consider the standard case where the distribution of Y given $X = x$ is Pareto of parameter $\beta > 0$, that is

$$\mathbb{P}(Y \leq y | X = x) = \begin{cases} 1 - (x/y)^\beta & \text{if} \quad y \geq x \\ 0 & \text{if} \quad y < x. \end{cases}$$

Here we choose X uniform on $(0, 1)$. The two partial αth frontiers are then given by

$$\begin{aligned} q_\alpha(x) &= \max\{x, x[(1-\alpha)(1+\beta)]^{-1/\beta}\}, \\ \xi_\alpha(x) &= \max\{x, x^{1+1/\beta}[(1-\alpha)(1+\beta)]^{-1/\beta}\}. \end{aligned}$$

Both families of partial frontiers differ following the values of the distribution parameter β. A graphical illustration is displayed in Figure 3. In each picture we superimpose the lower support boundary, the quantile functions $q_\alpha(x)$ and $\xi_\alpha(x)$ and the regression αth quantiles of Y given $X = x$, for a fixed $\beta \in \{1; 10\}$ and for two large values of $\alpha = 0.95, 0.99$. First note that $\xi_\alpha(\cdot)$ is overall smaller than $q_\alpha(\cdot)$ and that $q_\alpha(\cdot)$ itself is overall smaller than the αth regression quantile function, for any $\alpha \in (0, 1)$. Second note that for small values of β (*e.g.* $\beta = 1$), the use of extreme regression quantiles (*e.g.* $\alpha = 0.99$) to capture the most efficient firms seems in this particular case to be less justified than the use of $q_\alpha(\cdot)$ and $\xi_\alpha(\cdot)$. Indeed, from an economic point of view, it is not reasonable for optimal dominating firms to be too far in the output direction from the set of relatively inefficient firms (lower support boundary). The use of the three types of quantile-based frontiers seems to be more justified from an economic viewpoint for large values of the parameter β (*e.g.* $\beta = 10$). In particular the parabolic shape of the function $\xi_\alpha(\cdot)$ diminishes as β increases. The three αth quantile functions converge to the linear lower support frontier as $\beta \to \infty$.

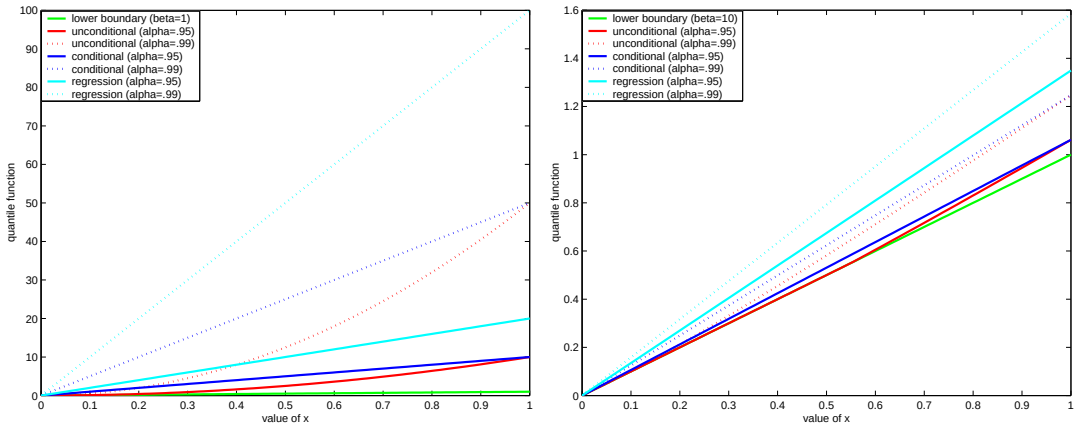

Figure 3: *The lower support boundary in green line, the quantile functions* $q_\alpha(\cdot)$ *in blue lines and* $\xi_\alpha(\cdot)$ *in red lines and the regression quantiles in cyan lines. On the left-hand side* $\beta = 1$*, on the right-hand side* $\beta = 10$*. In solid lines* $\alpha = .95$*, in dotted lines* $\alpha = .99$*.*

For a simulated data set of size $n = 100$ using $\beta = 3$, the true frontiers $q_\alpha(x)$ and $\xi_\alpha(x)$ with the smooth estimators $\hat{q}_\alpha(x)$ and $\hat{\xi}_\alpha(x)$ are graphed in Figure 4 for $\alpha \in \{0.95, 1\}$ and $\lambda = 5, 30, 60$. We obtain the same conclusions as in the preceding example. Here also, while $\hat{q}_{.95}(x)$ diverges from the extreme smooth frontier $\hat{\xi}_1(x)$ as x increases, the partial frontier $\hat{\xi}_{.95}$ parallels $\hat{\xi}_1(x)$ in much the same way as the partial frontiers $\hat{\xi}_\alpha$ do in Figure 2. In this particular example, it is clear that $\hat{q}_{.95}$ is more attracted by extreme data points with small X_i's. In general, for $\alpha < 1$, the frontiers $\hat{\xi}_\alpha$ are by construction more robust to extremes than $\hat{q}_\alpha$ and are less sensible to the border effects from which the frontiers $\hat{q}_\alpha$ suffer due to the conditioning $X \leq x$.

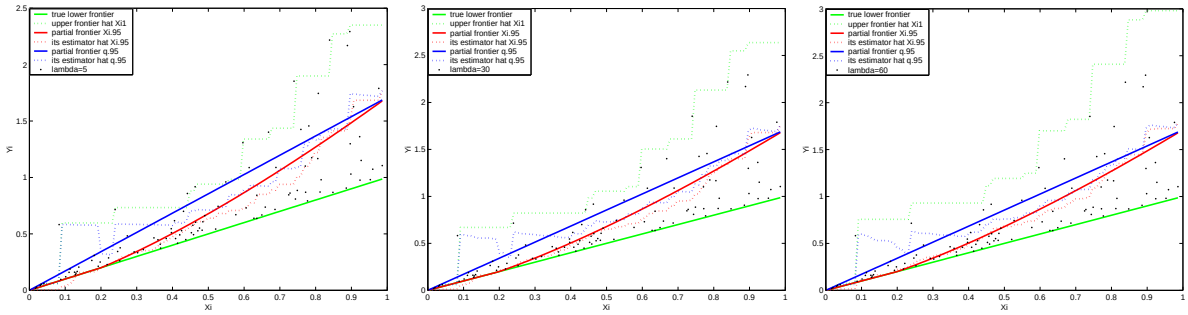

Figure 4: *Here* $\beta = 3$*. In solid lines: true lower frontier (green),* $\xi_{.95}$ *(red),* $q_{.95}$ *(blue). In dotted lines:* $\hat{\xi}_1$ *(green),* $\hat{\xi}_{.95}$ *(red) and* $\hat{q}_{.95}$ *(blue).* $n = 100$ *and from left to right* $\lambda = 5, 30, 60$*.*

3.1.3 Frontier analysis of French post offices

To illustrate our methodology, we employ data on 4000 post offices from France reported in Cazals *et al.* (2002). These data consist of the cost variable X_i which represents the quantity

of labor and the output variable Y_i defined as the volume of delivered mail.

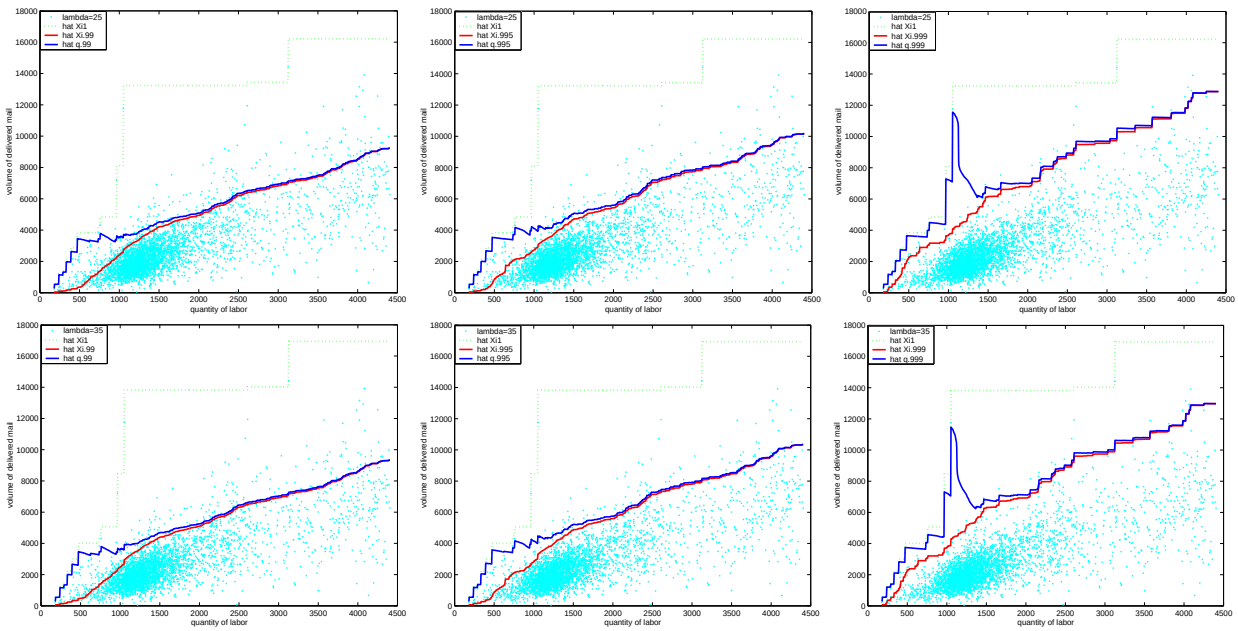

Figure 5: $n = 4000$ *French post offices. In dotted green line* $\hat{\xi}_1$. *In solid lines* $\hat{\xi}_\alpha$ *(red) and* $\hat{q}_\alpha$ *(blue). From left to right* $\alpha = .99, .995, .999$. *From top to bottom* $\lambda = 25, 35$.

In Figure 5, we provide in each picture a scatterplot of the data and plot the frontiers $\hat{\xi}_1$, $\hat{\xi}_\alpha$ and $\hat{q}_\alpha$, for $\alpha = 0.99$, 0.995 and 0.999, from left to right. From top to bottom, we used $\lambda = 25$ and $\lambda = 35$ in our computations. We first observe that both αth frontiers $\hat{\xi}_\alpha$ and $\hat{q}_\alpha$ are not influenced by the choice of the smoothing parameter λ, whereas the full extreme frontier $\hat{\xi}_1$ changes slightly. We also see that large frontiers $\hat{\xi}_\alpha$ suggest better capability of fitting efficient post offices than large $\hat{q}_\alpha$. It is apparent that the estimates $\hat{\xi}_\alpha$ (red lines) are less sensitive to the choice of extreme orders α than the frontiers $\hat{q}_\alpha$ are (blue lines). Figure 6 plots the αth frontier estimates $\hat{\xi}_\alpha$ (respectively, $\hat{q}_\alpha$) for the three values $\alpha = 0.99, 0.995, 0.999$ against each other: each blue (respectively, cyan) panel compares estimates $\{(\hat{\xi}_{\alpha_1}(X_i), \hat{\xi}_{\alpha_2}(X_i)) : i = 1, \ldots, n\}$ (respectively $\{(\hat{q}_{\alpha_1}(X_i), \hat{q}_{\alpha_2}(X_i)) : i = 1, \ldots, n\}$) for a pair (α_1, α_2) of values for α. Unlike cyan panels, most points fall on or near a straight line for all blue panels, confirming thus the impression from Figure 5, *i.e.*, while the frontier estimates $\hat{\xi}_\alpha$ obtained with the three extreme values of α are somewhat similar, one sees substantial differences for the frontiers $\hat{q}_\alpha$.

Note that the full extreme frontier $\hat{\xi}_1(x)$ (green line) is far from large partial frontiers $\hat{\xi}_\alpha(x)$ (red lines) even when α increases (this is not the case for $\hat{q}_\alpha(x)$). This might suggest the heavy-tailed case $F_{Z^x} \in \mathrm{DA}(G_{\gamma_x})$ with $\gamma_x > 0$. However, this assumption can hardly be

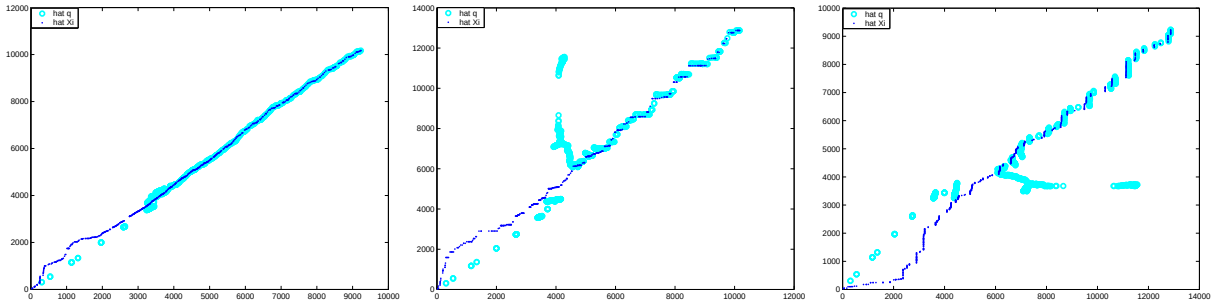

Figure 6: *The blue (respectively, cyan) plot consists of the points* $\{(\hat{\xi}_{\alpha_1}(X_i), \hat{\xi}_{\alpha_2}(X_i)) : i = 1, \ldots, n\}$ *(respectively* $\{(\hat{q}_{\alpha_1}(X_i), \hat{q}_{\alpha_2}(X_i)) : i = 1, \ldots, n\}$*). From left to right* $(\alpha_1, \alpha_2) = (.99, .995), (.995, .999), (.999, .99)$*. Here* $\lambda = 25$*.*

accepted by looking to the two first moments[5] plots displayed in Figure 7 and consisting of the sets of points $\{(k, \hat{\gamma}_x(k)) : 1 \leq k \leq N_x\}$, for $x = 1000, 2000$ (top panels). In contrast, when the bottom graphs (for $x = 3000, 4000$) look stable, they correspond to values $\hat{\gamma}_x \geq 0$.

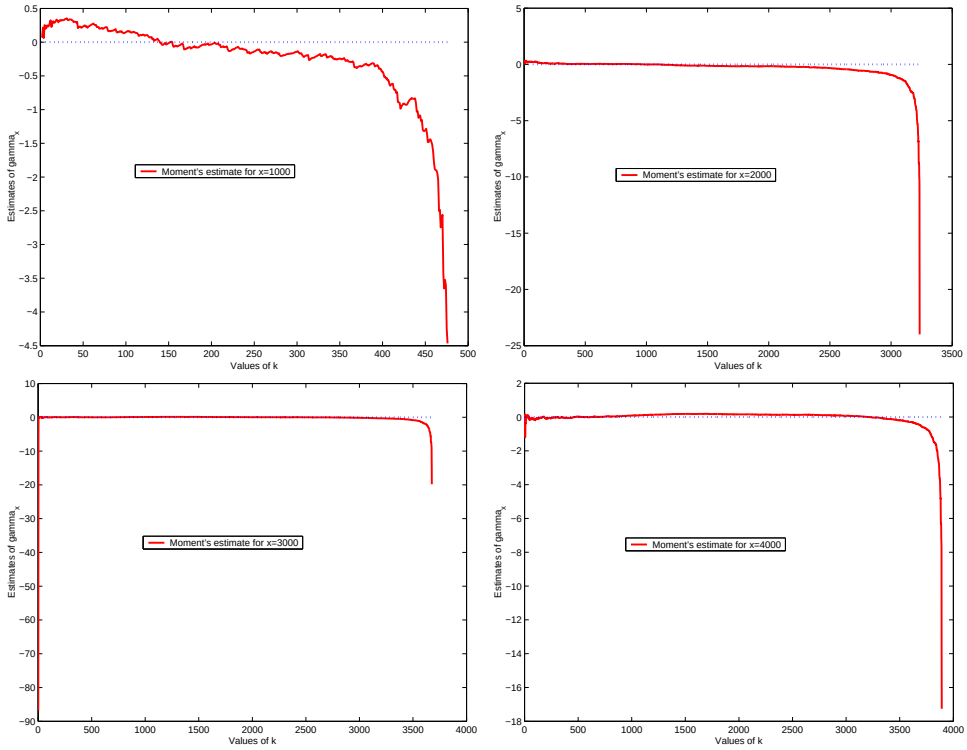

Figure 7: *Moments plots for* $x = 1000, 2000, 3000, 4000$*, respectively from left to right and from top to bottom.*

The frontier $\hat{\xi}_1(x)$ being by construction a shifted variant of the FDH estimator, it is

[5]The moment's estimator $\hat{\gamma}_x$ of the tail index γ_x is defined as $\hat{\gamma}_x = H_n^{(1)} + 1 - \frac{1}{2}\{1 - (H_n^{(1)})^2/H_n^{(2)}\}^{-1}$, with $H_n^{(j)} = (1/k)\sum_{i=0}^{k-1}(\log Z_{(n-i)}^x - \log Z_{(n-k)}^x)^j$ for $k < n$ and $j = 1, 2$ (Dekkers *et al.* (1989)).

more sensitive to extreme values. This frontier is clearly determined by a very few outlying post offices. At the opposite, the partial frontier $\hat{\xi}_\alpha(x)$ is more resistant to these outliers than $\hat{\xi}_1(x)$ and $\hat{q}_\alpha(x)$ even for too high values of α. This is the reason which explains the substantial difference between $\hat{\xi}_\alpha(x)$ and $\hat{\xi}_1(x)$ even when $\alpha = 0.9993$, 0.9995 and 0.9997 as shown in Figure 8.

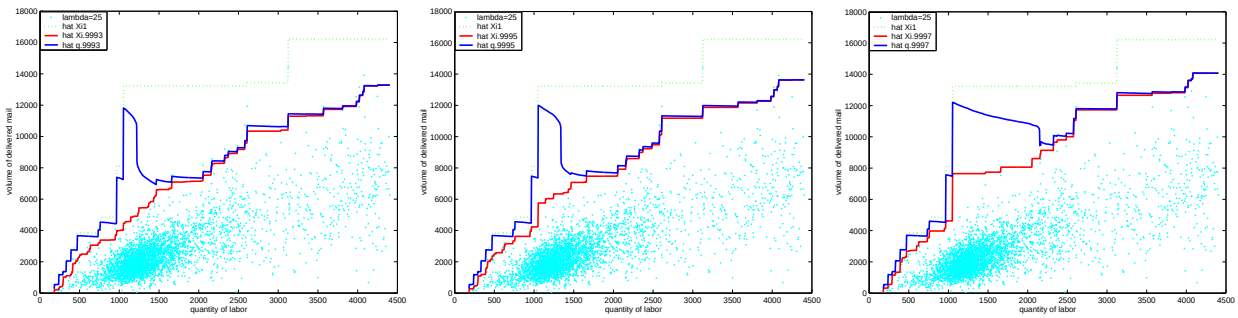

Figure 8: *As in Figure 5 with* $\lambda = 25$ *and, from left to right,* $\alpha = .9993, .9995, .9997$.

3.2 Monte Carlo experiments

3.2.1 Comparison of the full frontier estimators

Let us first compare the performance of both estimators $\hat{\xi}_1(\cdot)$ and $\xi_{1,n}(\cdot)$ of the full frontier function $\xi_1(\cdot)$. In the particular example described above in Paragraph 3.1.1 where $\xi_1(x) = x$, we have $F_{Z^x} \in \mathrm{DA}(\Psi_{\rho_x})$ with $\rho_x = -1/\gamma_x = 2$ and $a_n(x) = x - \max\{0, x - \sqrt{1/n}\}$. For the computation of the smoothed estimator $\hat{\xi}_1(x)$, we use here the true value of the optimal bandwidth $h_{opt}(x) = a_n(x)\rho_x^{-1}\Gamma(\rho_x^{-1})$ derived in (3), as well as the plug-in values $h_{opt,1}(x) = \hat{a}_n(x)\rho_x^{-1}\Gamma(\rho_x^{-1})$ and $h_{opt,2}(x) = \hat{a}_n(x)\hat{\rho}_x^{-1}\Gamma(\hat{\rho}_x^{-1})$ obtained by replacing $a_n(x)$ and ρ_x, respectively, with the empirical counterpart $\hat{a}_n(x) = Z^x_{(n)} - Z^x_{(n-1)}$ and the moment's estimator $\hat{\rho}_x = -1/\hat{\gamma}_x$. The computation of $\hat{\rho}_x$ depends on the sample fraction k whose choice is difficult in practice. By definition of the moment's estimator, the sequence $k = k_n(x)$ must be chosen as a function of both n and x such that $Z^x_{(n-k)} > 0$, which is equivalent to selecting k in $\{1, \ldots, N_x - 1\}$, with $N_x = \sum_{i=1}^n \mathbb{1}(X_i \leq x)$ being the number of strictly positive Z_i^x's. Here, we only use the values $k \in \{[N_x^{0.5}], [N_x^{0.7}], [N_x^{0.9}]\}$ to illustrate how much the estimates $\hat{\xi}_1(x)$ based on $h_{opt,2}(x)$ differ from those based on $h_{opt,1}(x)$ and $h_{opt}(x)$.

Figure 9 provides the Monte Carlo estimates of the Bias and the Mean-Squared Error (MSE) of $\hat{\xi}_1(x)$ and $\xi_{1,n}(x)$ computed over 2000 random replications with $n = 1000$ (what is important is not the sample size n itself but the number N_x of observations X_i smaller than or equal to x). The results are displayed for the FDH estimator $\xi_{1,n}(x)$ in solid blue line and for the smoothed version $\hat{\xi}_1(x)$ computed with the true bandwidth $h_{opt}(x)$ in solid

red line. We see clearly that the resulting smoothed maximum $\hat{\xi}_1(x)$ outperforms the sample maximum $\xi_{1,n}(x)$ in terms of both Bias (top picture) and MSE (bottom), for every x.

When only the tail-index ρ_x is supposed to be known and equal to $-1/2$ as it is typically the case in the econometric literature on nonparametric frontier estimation, the results for $\hat{\xi}_1(x)$ obtained with the plug-in bandwidth $h_{opt,1}(x)$ are displayed in dotted red line. We observe that the smoothed maximum's performance deteriorates when using the naive spacing $\hat{a}_n(x)$ in place of the theoretical scaling $a_n(x)$, but it remains still appreciably better than the sample maximum's performance in terms of both Bias and MSE, say, for all $x \geq 0.05$ (for too small values of $x < 0.05$, $\hat{\xi}_1(x)$ performs at least as $\xi_{1,n}(x)$).

When ρ_x is estimated by $\hat{\rho}_x$, the results for $\hat{\xi}_1(x)$ obtained by using the plug-in bandwidth $h_{opt,2}(x)$ with $k = [N_x^{0.5}], [N_x^{0.7}], [N_x^{0.9}]$, are displayed respectively in dotted green, cyan and magenta lines. These three lines are graphed only for the values of x where $\hat{\xi}_1(x)$ behaves better than $\xi_{1,n}(x)$ in terms of Bias and MSE. We see that, when N_x is large enough ($x > 0.5$), the three different selected values of k used for the computation of $\hat{\rho}_x$ in $h_{opt,2}(x)$ give very similar results to the "benchmarked" case $h_{opt,1}(x)$ (the idea is that for a properly chosen value k, both cases $h_{opt,1}(x)$ and $h_{opt,2}(x)$ should approximately yield similar values of $\hat{\xi}_1(x)$). In contrast, the results are all the more sensitive to the choice of k as N_x becomes small : large values of k seem to be needed as N_x decreases in order to get sensible results.

It should be also clear that the minimal value of N_x computed over the 2000 realizations is given, for instance, by $N_{0.3} = 60$ for $x = 0.3$ and by $N_{0.1} = 1$ for $x = 0.1$. While the estimation of ρ_x from the sample $\{Z^x_{(n-N_x+1)}, \ldots, Z^x_{(n)}\}$ of size N_x as small as $N_{0.3}$ can hardly result in satisfactory moment's estimate (or any other extreme-value based estimates), it is not even feasible when $N_x < 2$. This is a recurrent problem in extreme-value theory.

Apart from this vexing border defect, we could say in view of the results described above that the smoothed estimator's performance may be improved in terms of Bias and MSE by deriving a more efficient estimate for $a_n(x)$ than the naive spacing $\hat{a}_n(x)$, and by providing an appropriate choice of the sample fraction $k = k_n(x)$ involved in the tail-index estimates.

3.2.2 Comparison of extreme partial frontiers

Let us now compare the asymptotic normal estimators $\hat{\xi}_{(n-k+1)/n}$ and $\xi_{(n-k+1)/n,n} = Z^x_{(n-k+1)}$ of the extreme partial frontier $\xi_{(n-k+1)/n}$, for two values of $k = k_n \in \{[n^{0.5}], [n^{0.75}]\}$ and two sample sizes $n \in \{100, 1000\}$.

Case of a finite frontier : In the scenario of Paragraph 3.1.1 above, the derivative of the quantile tail function $U_x(t) := \xi_{1-1/t}(x)$ exists for all $t > x^{-2}$ and is given by $U'_x(t) = t^{\gamma_x - 1}/2 \in \mathrm{RV}_{\gamma_x - 1}$. Hence Condition (A2) holds in this case and the assumption of Theorem 4

that $hk^{3/2}\{nU_x'(n/k)\}^{-1} \to 0$ reduces to the simple condition that $n^{1/2}h \to 0$. For the computation of the smoothed estimator $\hat{\xi}_{(n-k+1)/n}$, we use here $h = n^{-r}$ with three values of $r \in \{0.5, 1, 1.5\}$: the choice $r = 0.5$ (for which $n^{1/2}h \not\to 0$ and $nh \to \infty$) corresponds rather to the asymptotic normality of ordinary smoothed quantiles, whereas the cases with $r > 0.5$ (for which $n^{1/2}h \to 0$ and $nh \not\to \infty$) correspond to the asymptotic normality of extreme smoothed quantiles. The Monte Carlo estimates of the Bias and MSE of the empirical partial frontier $\xi_{(n-k+1)/n,n}$ and the resulting smoothed three variants $\hat{\xi}_{(n-k+1)/n}$ are shown in Figure 10 for $k = [n^{0.5}]$ and in Figure 11 for $k = [n^{0.75}]$.

It may be seen that the smoothed extreme frontier $\hat{\xi}_{(n-k+1)/n}$ achieves its best performance in terms of MSE as h approaches to $n^{-1/2}$ and behaves in this case (*i.e.*, for $r \approx 0.5$) better than $\xi_{(n-k+1)/n,n}$, but not by much, as is to be expected from their asymptotic behavior. This appears to be true uniformly in x except for the too small values : a plausible explanation is that what is important when estimating $\xi_{(n-k+1)/n}(x)$ is not the sample size n itself, but the number of non-null transformed observations $Z_i^x = Y_i \mathbb{1}(X_i \leq x)$, which becomes negligible for x too small and so, smoothing fails in this case.

It is also interesting to note that when $\hat{\xi}_{(n-k+1)/n}$ attains its greatest gains in terms of MSE (as h approaches to $n^{-1/2}$), its bias becomes however considerably larger, as it is the case for ordinary smoothed quantiles.

Case of an infinite frontier : In the scenario of Paragraph 3.1.2 above, where $F_{Z^x} \in \mathrm{DA}(\Phi_{\rho_x})$ with $\rho_x = 1/\gamma_x = \beta$, Condition (A2) holds and the assumption of Theorem 4 that $hk^{3/2}\{nU_x'(n/k)\}^{-1} \to 0$ is equivalent to $hn^{-1/\beta}k^{3/2+1/\beta-1} \to 0$. When $\beta = 3$ as chosen above, this assumption reduces to $n^{1/12}h \to 0$ for $k = [n^{0.5}]$ and to $n^{7/24}h \to 0$ for $k = [n^{0.75}]$. Here, our Monte Carlo experiments are not in favor of the smoothed extreme quantiles and so we do not reproduce the figures for saving place : $\hat{\xi}_{(n-k+1)/n}$ performs at most as well as the sample version $\xi_{(n-k+1)/n,n}$ in terms of both Bias and MSE.

Appendix: Proofs

Proof of Theorem 1 (i) For any $x < Z_{(n)} + hc$, we have $\frac{x-Z_{(n)}}{h} < c$, then $H\left(\frac{x-Z_{(n)}}{h}\right) < 1$ and so $\hat{F}_Z(x) < 1$. Whence $Z_{(n)} + hc \leq \hat{\xi}_1$. On the other hand, for any $x \geq Z_{(n)} + hc$ we have $\frac{x-Z_i}{h} \geq c$ for each $i = 1, \cdots, n$. Whence $H\left(\frac{x-Z_i}{h}\right) = 1$ for each $i = 1, \cdots, n$. Consequently $\hat{F}_Z(x) = 1$ for any $x \geq Z_{(n)} + hc$. Thus $Z_{(n)} + hc \geq \hat{\xi}_1$.

(ii) Since $\mathbb{P}[a_n^{-1}(\hat{\xi}_1 - b_n) \leq x] \to G(x)$ and $\hat{\xi}_1 = Z_{(n)} + hc$, we have $\mathbb{P}[a_n^{-1}(Z_{(n)} - c_n) \leq x] \to G(x)$ for the sequence $c_n := b_n - hc$. As an immediate consequence, the nondegenerate distribution function G has one of the three forms described in Theorem 1(ii) according to the Fisher-Tippett Theorem (see *e.g.* Resnick 1987, Proposition 0.3, p.9). $\square$

Proof of Theorem 2 By Theorem 1(i) the convergence in distribution (2) implies $a_n^{-1}(Z_{(n)}-c_n) \xrightarrow{d} G$, that is $F_Z \in \mathrm{DA(G)}$ with $c_n = b_n - hc$. For the first assertion of the theorem we have $G = \Phi_\rho$ with $a_n = (1/(1-F_Z))^{-1}(n)$ and $c_n = 0$. Then by Proposition 2.1(i) in Resnick (1987, p.77), we have

$$\lim_{n\to\infty} \mathbb{E}\{a_n^{-1} Z_{(n)}\}^k = \int_{\mathbb{R}} z^k \Phi_\rho(dz) = \Gamma(1-k/\rho)$$

and the assertion (i) follows by using $(\hat{\xi}_1 - hc) = Z_{(n)}$. Assertions (ii) and (iii) follow in the same way by an easy application of Proposition 2.1(ii)-(iii) of Resnick (1987). $\square$

Proof of Theorem 3 First note that for any fixed integer $k \geq 0$ we have $k < n$ for all n large enough. Denote by $Z_{(n-k)}$ the $(k+1)$-th largest order statistic and let us show that

$$Z_{(n-k)} - hc < \hat{\xi}_{\frac{n-k}{n}} \leq Z_{(n-k)} + hc \quad \text{for all } n \text{ large enough.} \tag{A.1}$$

For any $x \geq Z_{(n-k)} + hc$, we have $\frac{x - Z_{(i)}}{h} \geq c$ for each $i \leq n-k$. Then $\hat{F}_Z(x) \geq (1/n)\sum_{i \leq n-k} H\left(\frac{x-Z_{(i)}}{h}\right) = \frac{n-k}{n}$. Therefore $Z_{(n-k)} + hc \geq \hat{\xi}_{\frac{n-k}{n}}$. On the other hand, $\hat{F}_Z(Z_{(n-k)} - hc) = (1/n)\sum_{i<n-k} H\left(\frac{Z_{(n-k)}-Z_{(i)}}{h} - c\right) \leq \frac{n-k-1}{n} < \frac{n-k}{n} = \hat{F}_Z(\hat{\xi}_{\frac{n-k}{n}})$. Which implies $Z_{(n-k)} - hc < \hat{\xi}_{\frac{n-k}{n}}$. Now we can turn to the desired assertion of the theorem. If $a_n^{-1}(\hat{\xi}_1 - b_n) \xrightarrow{d} G$ we have $a_n^{-1}(Z_{(n)} + hc - b_n) \xrightarrow{d} G$ by Theorem 1(i). Then following Theorem 21.18 in van der Vaart (1998, p.313), we obtain $a_n^{-1}(Z_{(n-k)} + hc - b_n) \xrightarrow{d} H$. Since

$$a_n^{-1}(Z_{(n-k)} + hc - b_n) - 2hca_n^{-1} < a_n^{-1}(\hat{\xi}_{\frac{n-k}{n}} - b_n) \leq a_n^{-1}(Z_{(n-k)} + hc - b_n),$$

the desired conclusion follows immediately from the condition $a_n^{-1}h \to 0$. $\square$

Proof of Theorem 4 Let $\sigma = (Z_{(n-k+1)} - Z_{(n-2k+1)})/\sqrt{2k}$. First, we know from Dekkers and de Haan (1989, Theorem 3.1) that $\sigma^{-1}(Z_{(n-k+1)} - \xi_\alpha)$ is asymptotically normal with mean zero and variance $2^{2\gamma+1}\gamma^2/(2^\gamma - 1)^2$, provided that $\alpha \to 1$, $n(1-\alpha) \to \infty$ and k is the integer part of $n(1-\alpha)$. We also know from Dekkers and de Haan (1989, Corollary 3.1) that $(Z_{(n-k+1)} - Z_{(n-2k+1)})/(n/2k)U'(n/2k) \xrightarrow{p} (2^\gamma - 1)/\gamma$ and so, it follows from the condition $h(2k)^{3/2}\{nU'(n/2k)\}^{-1} \to 0$ that $\sigma^{-1}h \xrightarrow{p} 0$. Thus we conclude by making use of (A.1) that $\sigma^{-1}(\hat{\xi}_{\frac{n-k+1}{n}} - \xi_\alpha)$ has the same asymptotic distribution as $\sigma^{-1}(Z_{(n-k+1)} - \xi_\alpha)$. We also have by (A.1) for all n sufficiently large,

$$(Z_{(n-k+1)} - Z_{(n-2k+1)}) - 2hc < \left(\hat{\xi}_{\frac{n-k+1}{n}} - \hat{\xi}_{\frac{n-2k+1}{n}}\right) < (Z_{(n-k+1)} - Z_{(n-2k+1)}) + 2hc.$$

Since $\sigma^{-1}h \xrightarrow{p} 0$, we get $h/(Z_{(n-k+1)} - Z_{(n-2k+1)}) \xrightarrow{p} 0$. Therefore $(\hat{\xi}_{\frac{n-k+1}{n}} - \hat{\xi}_{\frac{n-2k+1}{n}})/(Z_{(n-k+1)} - Z_{(n-2k+1)}) \xrightarrow{p} 1$, which completes the proof by using Slutsky Lemma. $\square$

Acknowledgements. The authors thank an anonymous reviewer for his valuable comments which led to a considerable improvement of the manuscript.

References

[1] Aragon, Y., A. Daouia and C. Thomas-Agnan (2005), Nonparametric Frontier Estimation: A Conditional Quantile-based Approach, *Econometric Theory*, 21, 358–389.

[2] Azzalini, A. (1981), A note on the estimation of a distribution function and quantiles by a kernel method, *Biometrika*, 68, 326–328.

[3] Beirlant, J., J. Teugels, Yu. Goegebeur and J. Segers (2004). Statistics of Extremes: Theory and Applications. Wiley Series in Probability and Statistics.

[4] Cazals, C., J-P. Florens and L. Simar (2002), Nonparametric frontier estimation: a robust approach, *Journal of Econometrics*, 106, 1–25.

[5] Daouia, A., J-P. Florens and L. Simar (2010), Frontier Estimation and Extreme Value Theory. *Bernoulli*, 16, 1039–1063.

[6] Daouia, A., L. Gardes and S. Girard (2009), Large Sample Approximation of the Distribution for Smooth Monotone Frontier Estimators, preprint, submitted for publication, `http://hal.archives-ouvertes.fr/hal-00409447`.

[7] Daouia, A., L. Gardes, S. Girard and A. Lekina (2011), Kernel estimators of extreme level curves, *Test*, 20, 311–333.

[8] Dekkers, A.L.M. and L. de Haan (1989), On the estimation of extreme-value index and large quantiles estimation, *The Annals of Statistics*, 17(4), 1795–1832.

[9] Dekkers, A.L.M., Einmahl, J.H.J. and L. de Haan (1989), A moment estimator for the index of an extreme-value distribution, *The Annals of Statistics*, 17(4), 1833–1855.

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

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

[12] Girard, S. and Jacob, P. (2008), Frontier estimation via kernel regression on high power-transformed data. *Journal of Multivariate Analysis*, 99, 403–420.

[13] Girard, S. and Menneteau, L. (2005), Central limit theorems for smoothed extreme value estimates of point processes boundaries. *Journal of Statistical Planning and Inference*, 135, 433–460.

[14] Martins-Filho, C. and Yao, F. (2008), A smooth nonparametric conditional quantile frontier estimator, *Journal of Econometrics*, 143, 317–333.

[15] Nadaraya, E.A. (1964), Some new estimates for distribution functions, *Theory Prob. Applic.* 15, 497–500.

[16] Pickands, J. (1975), Statistical inference using extreme order statistics, *The Annals of Statistics*, 3, 119–131.

[17] Resnick, S.I. (1987), Extreme Values, Regular Variation, and Point Processes. Springer-Verlag.

[18] Segers, J. (2005), Generalized Pickands estimators for the extreme value index, *Journal of Statististical Planning and Inference*, 128, 381–396.

[19] van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics, 3, Cambridge University Press, Cambridge.

[20] Wheelock, D.C., and Wilson, P.W. (2008), Non-parametric, Unconditional Quantile Estimation for Efficiency Analysis with an Application to Federal Reserve Check Processing Operations, *Journal of Econometrics*, 145 (1-2), 209–225.

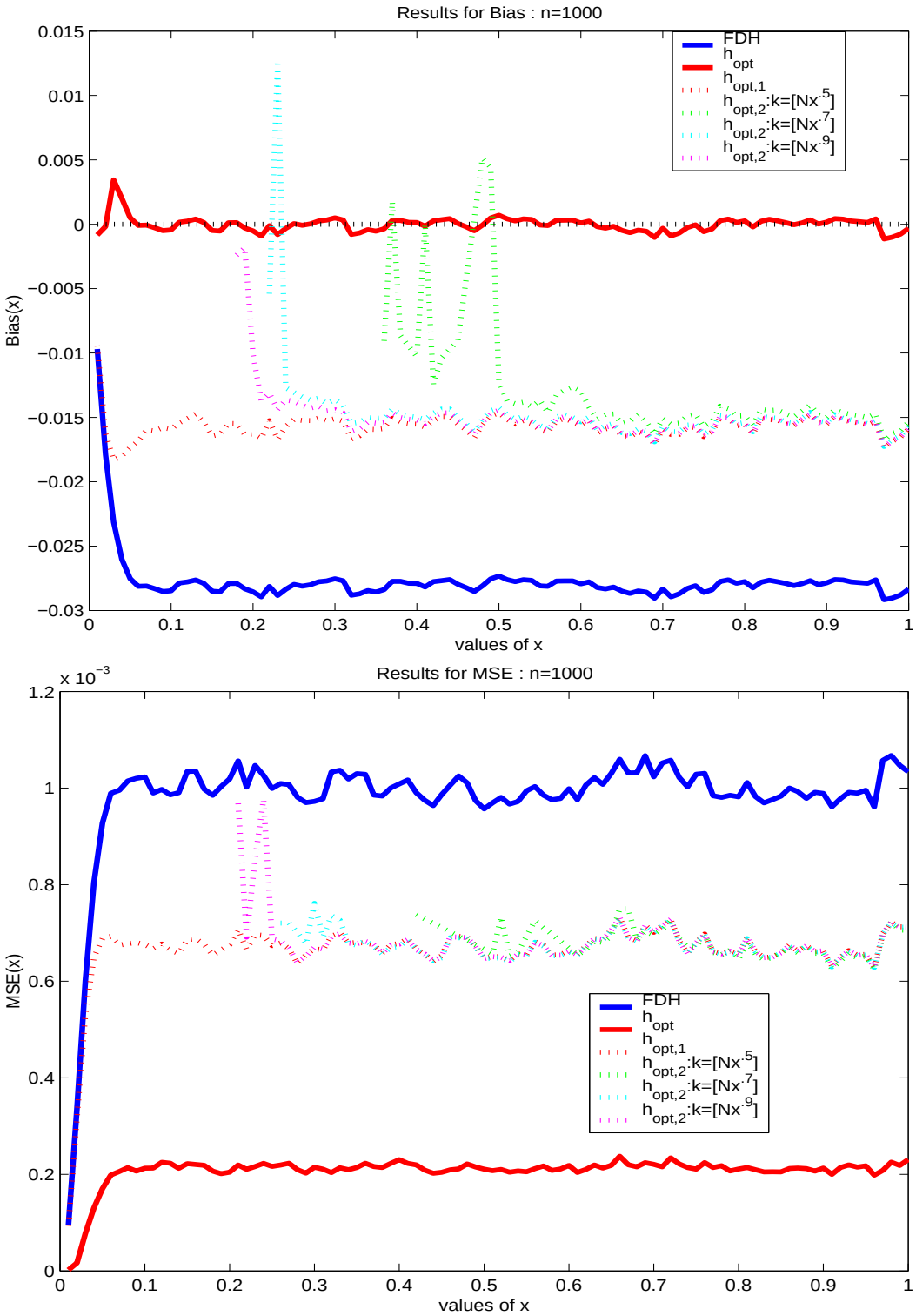

Figure 9: *Bias (top) and MSE (bottom) of the FDH estimator* $\xi_{1,n}$ *(solid blue line) and of the smoothed estimates* $\hat{\xi}_1$ *computed with :* h_{opt} *(solid red line),* $h_{opt,1}$ *(dotted red line) and* $h_{opt,2}$ *with* $k = [N_x^{0.5}], [N_x^{0.7}], [N_x^{0.9}]$ *(respectively in dotted green, cyan and magenta lines). 2000 Monte-Carlo simulations with* $n = 1000$.

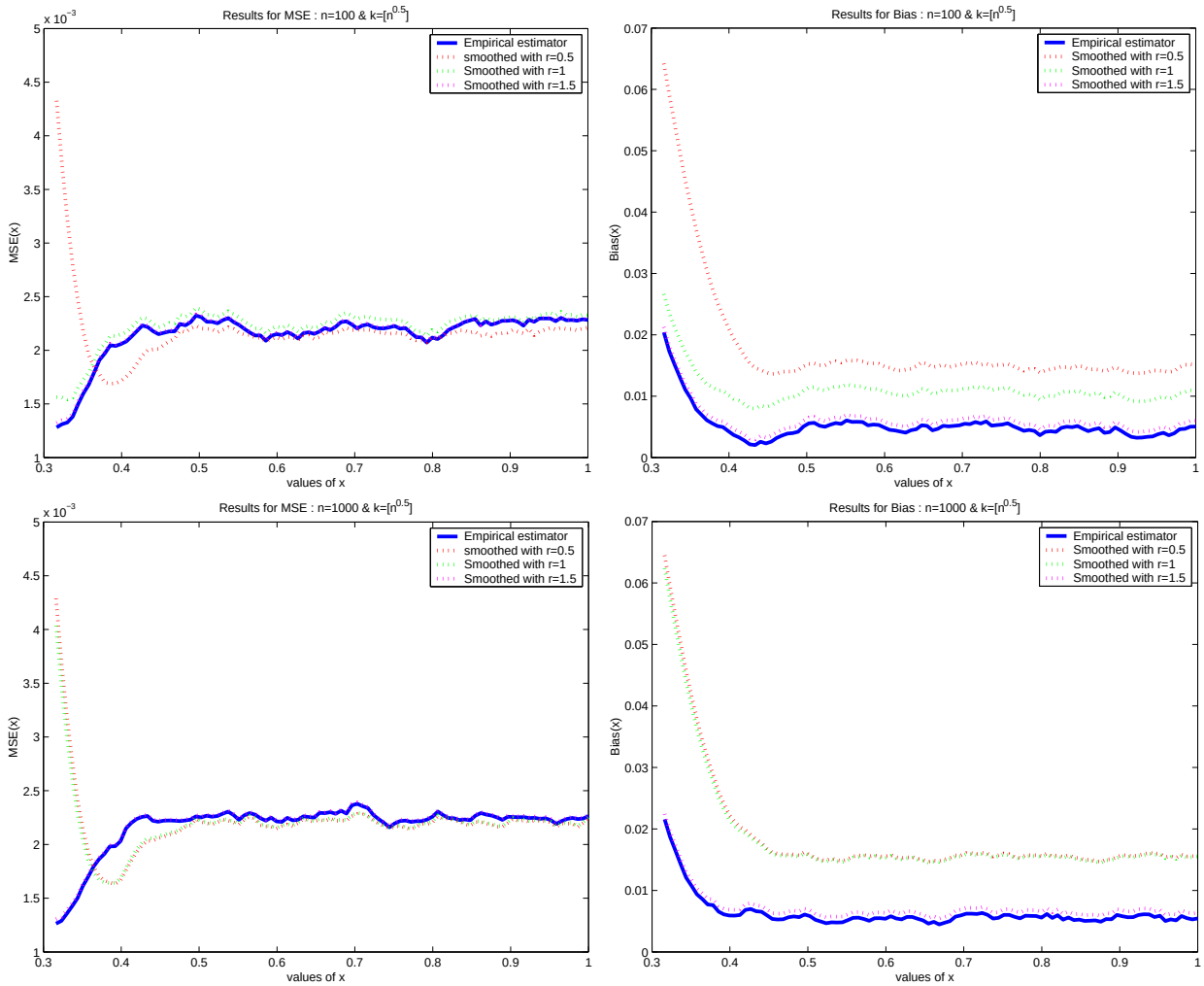

Figure 10: *Results for* $k = [n^{0.5}]$. *Bias (right panels) and MSE (left panels) of the estimates* $\xi_{(n-k+1)/n,n}$ *(blue) and* $\hat{\xi}_{(n-k+1)/n}$ *(red, green and magenta, respectively, for* $r = 0.5, 1, 1.5$*) over 2000 Monte-Carlo simulations, sample size* $n = 100$ *(top) and* $n = 1000$ *(bottom).*

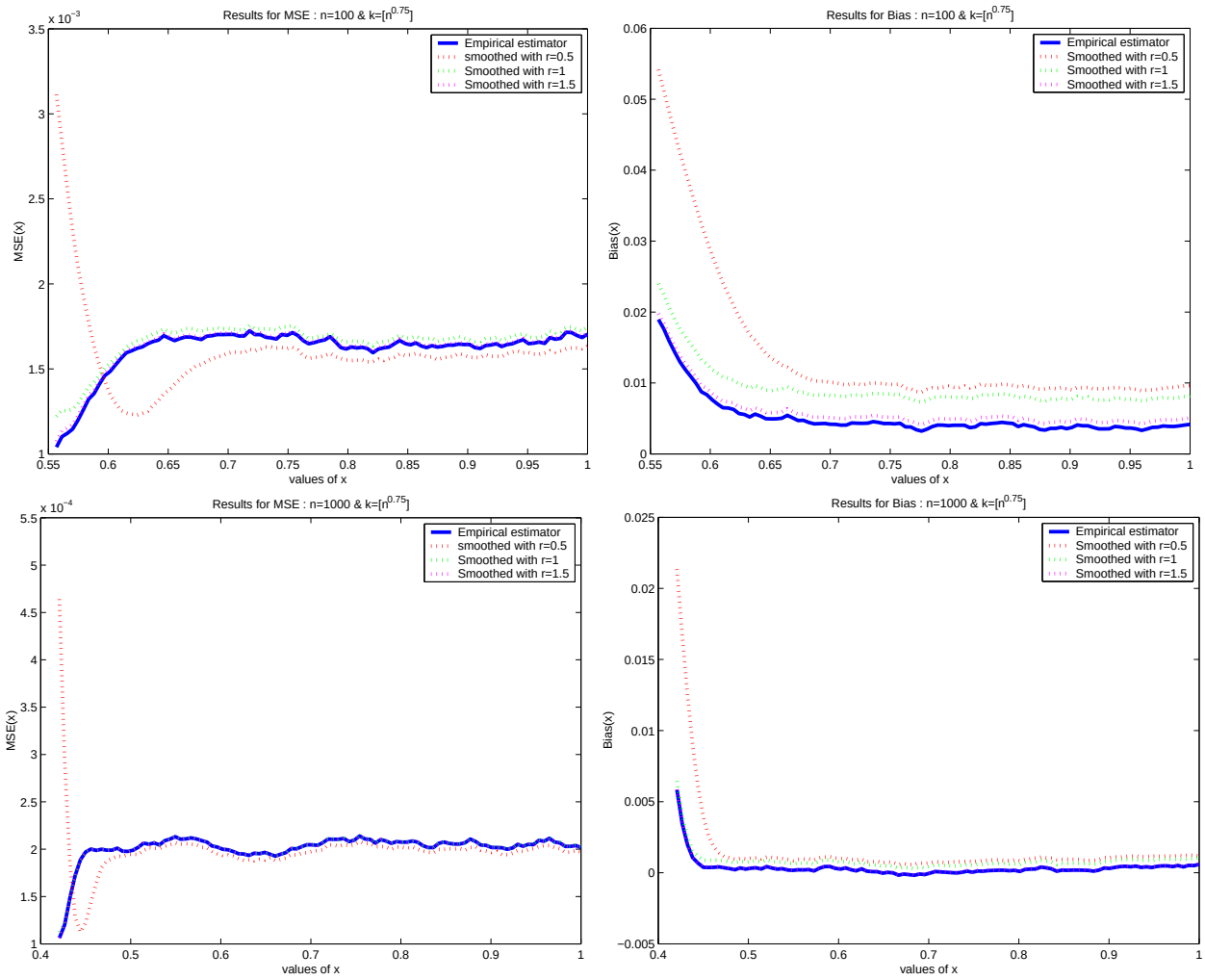

Figure 11: *As above with* $k = [n^{0.75}]$.