Working Paper Series n° 614

September, 2009

"Regularization of Nonparametric Frontier Estimators"

Abdelaati DAOUIA,
Jean-Pierre FLORENS and
Léopold SIMAR

Institut d'Economie Industrielle (IDEI) – Manufacture des Tabacs
Aile Jean-Jacques Laffont – 21, allée de Brienne – 31000 TOULOUSE – FRANCE
Tél. + 33(0)5 61 12 85 89 – Fax + 33(0)5 61 12 86 37 – **www.idei.fr** – contact@cict.fr

Regularization of Nonparametric Frontier Estimators

Abdelaati Daouia
Toulouse School of Economics (GREMAQ), France
daouia@cict.fr

Jean-Pierre Florens
Toulouse School of Economics (GREMAQ, IDEI), France
florens@cict.fr

Léopold Simar*
Institut de Statistique, Université Catholique de Louvain, Belgium
and Toulouse School of Economics, (GREMAQ, IDEI), France
leopold.simar@uclouvain.be

September 22, 2009

Abstract

In production theory and efficiency analysis, we are interested in estimating the production frontier which is the locus of the maximal attainable level of an output (the production), given a set of inputs (the production factors). In other setups, we are rather willing to estimate an input (or cost) frontier that is defined as the minimal level of the input (cost) attainable for a given set of outputs (goods or services produced). In both cases the problem can be viewed as estimating a surface under shape constraints (monotonicity, ...). In this paper we derive the theory of an estimator of the frontier having an asymptotic normal distribution. The basic tool is the order-m partial frontier where we let the order m to converge to infinity when $n \to \infty$ but at a slow rate. The final estimator is then corrected for its inherent bias. We thus can view our estimator as a regularized frontier estimator which, in addition, is more robust to extreme values and outliers than the usual nonparametric frontier estimators, like FDH. The performances of our estimators are evaluated in finite samples through some Monte-Carlo experiments. We illustrate also how to provide, in an easy way, confidence intervals for the frontier function both with a simulated data set and a real data set.

Key words : Production function, Free Disposal Hull, Nonparametric frontier, Robust estimation, Extreme values, Tail index.

*Support from the "Interuniversity Attraction Pole", Phase VI (No. P6/03) of the Belgian Science Policy and from the Chair of Excellency "Pierre de Fermat", Région Midi-Pyrénées, France are acknowledged.

1 Introduction and Basic Concepts

In production theory and efficiency analysis, we are interested in estimating the production frontier which is the locus of the maximal attainable level of an output (the production), given a set of inputs (the production factors). In other setups, we are rather willing to estimate an input (or cost) frontier that is defined as the minimal attainable level of the input (cost) for a given set of outputs (goods or services produced). In both cases the problem can be viewed as estimating a surface under shape constraints (monotonicity,...). The efficiency score of a given unit is then determined by an appropriate distance (in the output direction, or in the input direction) of this unit to the optimal frontier.

Formally (we will in this paper focus the presentation in the input orientation case, where we want to estimate the minimal cost frontier[1]), let $x \in \mathbb{R}_+$ denote the input (or the cost of production) and $y \in \mathbb{R}_+^q$ be the vector of goods or services produced. The attainable set (feasible combinations of input and outputs) is defined as

$$\Psi = \{(x, y) \in \mathbb{R}_+ \times \mathbb{R}_+^q \mid y \text{ can be produced by } x\}. \tag{1.1}$$

A minimal assumption often accepted for Ψ is the free disposability of the inputs and of the outputs, namely, if $(x, y) \in \Psi$, then $(x', y') \in \Psi$ for any pairs (x', y') such that $x' \geq x$ and $y' \leq y$. This implies a monotonicity property of the frontier surface. Sometimes (not in this paper), the hypothesis of the convexity of Ψ is also assumed (see Shephard, 1970 for a comprehensive overview of the underlying economic models used in prodution theory). The efficient boundary of Ψ, in the input oriented case, is represented by the minimal input frontier function

$$\varphi(y) = \inf\{x \mid (x, y) \in \Psi\}, \tag{1.2}$$

and the Farrell-Debreu efficiency score of a unit operating at the level (x_0, y_0) is given by the ratio $\varphi(y_0)/x_0$, which gives a number between zero and one. An efficiency equal to one corresponds to an input-efficient unit (being on the minimal input frontier) and more generally $\varphi(y_0)/x_0 \leq 1$ gives the reduction of input (cost) the firm should reach to be considered as input-efficient.

A popular nonparametric estimator of the attainable set is the Free Disposal Hull (FDH) estimator proposed by Deprins, Simar and Tulkens (1984). The FDH is the smallest monotone set enveloping the data points, it relies only on the free disposability assumption and its asymptotic properties have been established (Park, Simar and Weiner, 2000 and Daouia,

[1] The presentation for the output oriented case, where we want to estimate the maximal production frontier, is a straightforward adaptation of what is done here. In the appendix, we give a summary of the notations and main results for that case.

Florens, Simar, 2008). More details will be given below. Another nonparametric estimator, the Data Envelopment Analysis (DEA), initiated by Farrell (1957) and popularized by Charnes, Cooper and Rhodes (1978), can be justified when the convexity of Ψ is moreover assumed. Its asymptotic properties have been established in Kneip et al. (2008). A recent survey of the available statistical tools for making inference in these nonparametric models can be found in Simar and Wilson (2008).

The FDH estimator: basic properties

The attainable set Ψ can be seen as the support of the random vector (X, Y) defined on an appropriate probability space. It will be useful to describe the joint distribution of (X, Y) by its joint survivor function:

$$S_{XY}(x, y) = \text{Prob}(X \geq x, Y \geq y) = S(x|y)S_Y(y), \tag{1.3}$$

where $S(x|y) = \text{Prob}(X \geq x \,|\, Y \geq y)$ and $S_Y(y) = \text{Prob}(Y \geq y)$. Notice that the conditional survivor function $S(x|y)$ is non-standard, since the condition is $Y \geq y$.

Cazals, Florens and Simar (2002) have shown that under the free disposability assumption, the minimal input function $\varphi(y)$ can equivalently be defined as

$$\varphi(y) = \inf\{x \,|\, S(x|y) < 1\}. \tag{1.4}$$

Since the attainable set is unknown, it has to be estimated from a sample of i.i.d. units $\mathcal{X}_n = \{(X_i, Y_i) \,|\, i = 1, \ldots, n\}$. The free disposal hull of $\mathcal{X}_n$ is the FDH estimator

$$\widehat{\Psi} = \{(x, y) \mid y \leq Y_i, x \geq X_i, i = 1, \ldots, n\}, \tag{1.5}$$

providing the FDH estimator of the frontier $\varphi(y)$

$$\hat{\varphi}(y) = \inf\{x \,|\, \widehat{S}(x|y) < 1\} = \min_{\{i: Y_i \geq y\}} X_i, \tag{1.6}$$

where $\widehat{S}(x|y) = \widehat{S}_{XY}(x, y)/\widehat{S}_Y(y)$ with $\widehat{S}_{XY}(x, y) = (1/n)\sum_{i=1}^n 1\!\!I(X_i \geq x, Y_i \geq y)$ and $\widehat{S}_Y(y) = (1/n)\sum_{i=1}^n 1\!\!I(Y_i \geq y)$. Park et al. (2000) have obtained the limiting distribution of FDH estimators in a full multivariate set-up under some regularity conditions. The most general asymptotic result in our setup here is given by Daouia et al. (2008) and can be summarized as follows.

Under the regularity condition (Corollary 2.2 in Daouia et al., 2008)

$$S_Y(y)(1 - S(x|y)) = \ell_y\big(x - \varphi(y)\big)^{\rho_y} + o\big((x - \varphi(y))^{\rho_y}\big), \text{ as } x \downarrow \varphi(y), \tag{1.7}$$

with $\ell_y > 0$, $\rho_y > q$ and $\varphi(y)$ being differentiable in y with strictly positive first partial derivatives, we have[2] as $n \to \infty$

$$(n\ell_y)^{1/\rho_y}\big(\hat{\varphi}(y) - \varphi(y)\big) \xrightarrow{\mathcal{L}} \text{Weibull}(1, \rho_y). \tag{1.8}$$

In addition, the joint density of (X, Y) near the frontier function can be expressed as

$$f(x, y) = c_y(x - \varphi(y))^{\beta_y} + o\big((x - \varphi(y))^{\beta_y}\big), \text{ as } x \downarrow \varphi(y), \tag{1.9}$$

where $c_y > 0$ and $\beta_y = \rho_y - (q + 1)$. Since $\beta_y > -1$, the asymptotic result covers the cases $-1 < \beta_y < 0$, where the density tends to infinity at the frontier, at a speed of the power $\rho_y - (q + 1)$, the case $\beta_y = 0$ where the density has a jump at the frontier ($\rho_y = q + 1$) and the cases $\beta_y > 0$ where the joint density decays to zero at a speed of the power $\rho_y - (q + 1)$.

Remark 1.1. *The regularity condition* (1.7) *is a particular case of the more general extreme value regularity condition (see Daouia et al., 2008 for details)*

$$S_Y(y)(1 - S(x|y)) = L_y\left(\frac{1}{x - \varphi(y)}\right)(x - \varphi(y))^{\rho_y},$$

where L_y is a slowly varying function and $\rho_y > 0$ is the tail index. For instance, if (X, Y) is uniformly distributed over $\Psi = \{(x, y)|0 \le y \le x \le 1\}$, we have $L_y(\cdot) = \ell_y = \ell = 1$ and $\rho_y = \rho = 2$ and (1.7) *is satisfied.*

If $X = Y^{1/2}\exp(U)$ where Y is uniform over $[0, 1]$ and U, independent of Y, is Exponential with parameter $\lambda = 3$, we have $\rho_y = \rho = 2$ and $L_y\left(\frac{1}{x-\varphi(y)}\right) = \ell_y + o\big((x - \varphi(y))\big)$ when $x \downarrow \varphi(y)$, with $\ell_y = \ell = 3$ and (1.7) *is satisfied.*

Order-m frontier and robust estimator of the frontier

By construction, since it envelops all the data points, the FDH estimator (and its convexified version, the DEA estimator) is very sensitive to outliers and extreme data points. Cazals et al. (2002) suggested to define a benchmark frontier that is less extreme than the full frontier function $\varphi(y)$. Indeed, the latter can be defined as the minimal achievable input level for firms producing at least the level y, see (1.4). A less extreme benchmark, based on the concept of order-m frontier, is defined as the expected minimal input value among m peers drawn at random in the population of units producing at least the level y, where m is a natural number ($m \ge 1$). Formally,

$$\varphi_m(y) = \mathbb{E}\left[\min(X_1, \ldots, X_m)|Y \ge y\right], \tag{1.10}$$

[2] The Weibull distribution is related to the Exponential distribution: $W \sim \text{Weibull}(1, c) \Leftrightarrow W^c \sim \text{Exp}(1)$.

provided the expectation exists. We have the following equivalences

$$\varphi_m(y) = \int_0^\infty S^m(u|y)\,du = \varphi(y) + \int_{\varphi(y)}^\infty S^m(u|y)\,du. \tag{1.11}$$

It can be seen that $\varphi_m(y) \to \varphi(y)$ as $m \to \infty$.

A nonparametric estimator of $\varphi_m(y)$ is given by pluging the empirical version of $S(u|y)$ in (1.11) to obtain

$$\hat{\varphi}_m(y) = \int_0^\infty \widehat{S}^m(u|y)\,du. \tag{1.12}$$

For fixed m, it has been shown that $\sqrt{n}\big(\hat{\varphi}_m(\cdot) - \varphi_m(\cdot)\big) \xrightarrow{\mathcal{L}} \mathcal{G}(0,\Omega)$ where $\mathcal{G}$ is a gaussian process with covariance function Ω given in Cazals et al. (2002). In particular, for any given y and a fixed value of m, we have as $n \to \infty$,

$$\frac{\sqrt{n}}{\sigma(m,y)}\big(\hat{\varphi}_m(y) - \varphi_m(y)\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \tag{1.13}$$

where

$$\sigma^2(m,y) = \mathbb{E}\left[\frac{m\,1\!\!1(Y \geq y)}{S_Y(y)} \int_0^\infty \Big(S^{m-1}(u|y)\,1\!\!1(X \geq u) - S^m(u|y)\Big)\,du\right]^2. \tag{1.14}$$

It is clear that if $m \to \infty$, $\hat{\varphi}_m(y)$ will converge to the FDH estimator $\hat{\varphi}(y)$. Cazals et al. show that if $m = m_n \to \infty$ fast enough when $n \to \infty$, the resulting estimator has the same asymptotic properties than the FDH estimator

$$(n\ell_y)^{1/\rho_y}\big(\hat{\varphi}_{m_n}(y) - \varphi(y)\big) \xrightarrow{\mathcal{L}} \text{Weibull}(1,\rho_y).$$

Of course, for finite n, the resulting estimator $\hat{\varphi}_{m_n}(y)$ does not envelop all the data points and so provides a robust version of the FDH estimator.

In this paper we address the problem of the regularization of the FDH estimator. The central question is the following: is it possible to find a sequence of m_n converging to infinity as $n \to \infty$, but slowly enough to keep an asymptotic normal distribution. We will find this sequence and so obtain a regularized nonparametric estimator of the frontier having an asymptotic normal distribution and being robust to outliers and extreme data points. This motivates the development of the theory for order-m frontiers, when m tends to infinity.

Related work is Daouia et al. (2008) where the links between frontier estimation and extreme values theory are established. By doing so, they revisit and extend former results on the asymptotic behavior of the FDH estimator. Extreme value theory allows also to extend the properties of another partial frontier, the order-α quantile frontier (see Aragon, Daouia and Thomas-Agnan, 2005) providing an alternative robust estimator of the frontier function. The duality between order-m and order-α frontier has been investigated by Daouia and Gijbels (2009). They show in particular, that even if the order-α quantile frontiers have global better robustness properties (higher breakdown value), it appears that once they breakdown, they become less resistant to outliers than the order-m frontiers.

Section 2 gives the main theoretical result of this paper: the estimation of the order-m frontier when m tends to infinity (subsection 2.1) and how to implement an estimator of the frontier $\varphi(y)$ in practice (subsection 2.2). Section 3 addresses the problem of estimating the unknown parameters of the asymptotic distribution. Section 4 illustrates how the procedure works in practice with simulated data and with a real data set. Section 5 concludes.

2 The Main Result

2.1 Estimation of the order-m frontier when $m \to \infty$

We start with a preliminary lemma which controls, as $m \to \infty$, the variance of the order-m estimator $\hat{\varphi}_m(y)$ given in (1.14).

Lemma 2.1. *Under the regularity condition* (1.7)*, we have for any y such that $S_Y(y) > 0$, as $m \to \infty$*

$$k_{1,y}\, m^{1-2/\rho_y} \leq \sigma^2(m,y) \leq k_{2,y}\, m^{2-2/\rho_y}, \tag{2.1}$$

where $k_{1,y}$ and $k_{2,y}$ are some positive constants.

Proof: We first obtain after some elementary algebraic manipulations that the variance can be expressed as

$$\sigma^2(m,y) = \frac{2m^2}{S_Y(y)} \int_{\varphi(y)}^{\infty} \int_{\varphi(y)}^{\infty} S^m(u|y) S^{m-1}(v|y)(1 - S(v|y))\, \mathbb{1}(u \geq v)\, du\, dv. \tag{2.2}$$

(i) Searching a minorant of $\sigma^2(m,y)$ when $m \to \infty$. We first notice that

$$\sigma^2(m,y) = \frac{2m^2}{S_Y(y)} \int_{\varphi(y)}^{\infty} S^{m-1}(v|y) F(v|y) \left[\int_v^{\infty} S^m(u|y)\, du \right] dv,$$

where $F(v|y) = 1 - S(v|y)$. So that for all $\delta > 0$, we have

$$\begin{aligned} \sigma^2(m,y) &\geq \frac{2m^2}{S_Y(y)} \int_{\varphi(y)}^{\varphi(y)+\delta} S^{m-1}(v|y) F(v|y) \left[\int_v^{v+\delta} S^m(u|y)\, du \right] dv, \\ &\geq \frac{2m^2\delta}{S_Y(y)} \int_{\varphi(y)}^{\varphi(y)+\delta} S^{m-1}(v|y) F(v|y) S^m(v+\delta)\, dv. \end{aligned}$$

Since $S^{m-1}(v|y) \geq S^{m-1}(v+\delta|y) \geq S^m(v+\delta|y)$, we have

$$\begin{aligned}\sigma^2(m,y) &\geq \frac{2m^2\delta}{S_Y(y)} \int_{\varphi(y)}^{\varphi(y)+\delta} S^{2m}(v+\delta|y)F(v|y)\,dv, \\ &\geq \frac{2m^2\delta}{S_Y(y)} S^{2m}(\varphi(y)+2\delta|y) \int_{\varphi(y)}^{\varphi(y)+\delta} F(v|y)\,dv.\end{aligned}$$

Now, if $\delta \downarrow 0$, by the regularity condition (1.7) we have that

$$\int_{\varphi(y)}^{\varphi(y)+\delta} F(v|y)\,dv \geq \frac{c_y}{\rho_y+1}\frac{\delta^{\rho_y+1}}{2}, \tag{2.3}$$

where $c_y = \dfrac{\ell_y}{S_Y(y)}$. When $\delta \downarrow 0$, it is also easy to see from (1.7) that

$$S(\varphi(y)+2\delta|y) \;\; \geq \;\; 1 - 2c_y(2\delta)^{\rho_y} = \exp\left[\log\left(1-2c_y(2\delta)^{\rho_y}\right)\right].$$

Therefore $S^{2m}(\varphi(y)+2\delta|y) \geq \exp\left[2m\log\left(1-2c_y(2\delta)^{\rho_y}\right)\right]$. Since $\lim_{\delta\downarrow 0} \dfrac{\log\left(1-2c_y(2\delta)^{\rho_y}\right)}{-2c_y(2\delta)^{\rho_y}} = 1$, for sufficiently small $\delta > 0$ we have $\dfrac{\log\left(1-2c_y(2\delta)^{\rho_y}\right)}{-2c_y(2\delta)^{\rho_y}} \leq 2$. So, when $\delta \downarrow 0$ we have $S^{2m}(\varphi(y)+2\delta|y) \geq e^{-8mc_y(2\delta)^{\rho_y}}$. Pluging these results in the latter inequality for $\sigma^2(m,y)$ we have as $\delta \downarrow 0$

$$\sigma^2(m,y) \geq \frac{2m^2\delta}{S_Y(y)}\, e^{-8mc_y(2\delta)^{\rho_y}} \frac{c_y}{\rho_y+1}\frac{\delta^{\rho_y+1}}{2}.$$

Choosing $\delta = (1/m)^{1/\rho_y}$, we have as $m \to \infty$

$$\sigma^2(m,y) \geq k_{1,y}\, m^{1-2/\rho_y}. \tag{2.4}$$

(ii) Searching a majorant of $\sigma^2(m,y)$ when $m \to \infty$. From (2.2) we have

$$\begin{aligned}\sigma^2(m,y) &\leq \frac{2m^2}{S_Y(y)} \int_{\varphi(y)}^{\infty}\int_{\varphi(y)}^{\infty} S^m(u|y)S^{m-1}(v|y)(1-S(v|y))\,du\,dv, \\ &\leq \frac{2m^2}{S_Y(y)} \left[(\varphi_m(y)-\varphi(y))(\varphi_{m-1}(y)-\varphi(y)) - (\varphi_m(y)-\varphi(y))^2\right] \\ &\leq \frac{2m^2}{S_Y(y)}(\varphi_m(y)-\varphi(y))^2\left[\frac{\varphi_{m-1}(y)-\varphi(y)}{\varphi_m(y)-\varphi(y)} - 1\right]\end{aligned}$$

Now, by the regularity condition (1.7), the equation (2.5) in Daouia et al. (2008) and from the definition (1.10) of φ_m, we have as $m \to \infty$

$$\varphi_m(y) - \varphi(y) = \Gamma\left(1+\frac{1}{\rho_y}\right)\left(\frac{1}{m\,\ell_y}\right)^{1/\rho_y} + o(m^{-1/\rho_y}). \tag{2.5}$$

Therefore, as $m \to \infty$,

$$\sigma^2(m,y) \;\le\; \frac{2m^2}{S_Y(y)}\left[\Gamma^2(1+1/\rho_y)(m\ell_y)^{-2/\rho_y} + o(m^{-2/\rho_y})\right] \le k_{2,y} m^{2-2/\rho_y},$$

where $k_{2,y}$ is a positive constant. This completes the proof of the lemma. $\square$

The following theorem gives the basic results of our paper, it specifies under which condition on the sequence m_n, the asymptotic distribution of $\hat{\varphi}_{m_n}(y)$ is still a Normal distribution.

Theorem 2.1. *Under the regularity condition* (1.7), *and if* $m_n = c\, n^{1/3-\varepsilon}(\log\log n)^{-2/3}$ *for some constants* $c > 0$ *and* $\varepsilon \in (0, 1/3)$, *we have for any* y *such that* $S_Y(y) > 0$, *as* $n \to \infty$

$$\frac{\sqrt{n}}{\sigma(m_n,y)}\big(\hat{\varphi}_{m_n}(y) - \varphi_{m_n}(y)\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \tag{2.6}$$

Proof: In the proof, to simplify the notation, we will denote m_n by m, keeping in mind that $m = m_n \to \infty$ when $n \to \infty$ at the rate given by m_n. Let us define

$$R^y_{m,n} = \big(\hat{\varphi}_m(y) - \varphi_m(y)\big) - m\int_{\varphi(y)}^{\infty} S^{m-1}(u|y)\big[\widehat{S}(u|y) - S(u|y)\big]\,du.$$

So the object of interest for the theorem can be written as

$$\begin{aligned}\frac{\sqrt{n}}{\sigma(m,y)}\big(\hat{\varphi}_m(y) - \varphi_m(y)\big) &= \frac{m\sqrt{n}}{\sigma(m,y)}\int_{\varphi(y)}^{\infty} S^{m-1}(u|y)\big[\widehat{S}(u|y) - S(u|y)\big]du \\ &+ \frac{\sqrt{n}}{\sigma(m,y)}R^y_{m,n}. \end{aligned}\tag{2.7}$$

(i) We first prove that $\frac{\sqrt{n}}{\sigma(m,y)}R^y_{m,n} \xrightarrow{a.s.} 0$ *as* $n \to \infty$. Since $\hat{\varphi}(y) \overset{a.s.}{\ge} \varphi(y)$, we have

$$R^y_{m,n} \overset{a.s.}{=} \int_{\varphi(y)}^{\infty}\big(\widehat{S}^m(u|y) - S^m(u|y)\big)\,du - m\int_{\varphi(y)}^{\infty} S^{m-1}(u|y)\big[\widehat{S}(u|y) - S(u|y)\big]\,du.$$

Now, consider the following Taylor expansion

$$\begin{aligned}\int_{\varphi(y)}^{\infty}\big(\widehat{S}^m(u|y) - S^m(u|y)\big)\,du &= m\int_{\varphi(y)}^{\infty} S^{m-1}(u|y)\big[\widehat{S}(u|y) - S(u|y)\big]\,du \\ &+ \frac{1}{2}m(m-1)\int_{\varphi(y)}^{\infty}\big[\widehat{S}(u|y) - S(u|y)\big]^2 b_y^{m-2}(u)\,du,\end{aligned}$$

where, $\widehat{S}(u|y) \wedge S(u|y) \le b_y(u) \le \widehat{S}(u|y) \vee S(u|y)$. So, we obtain:

$$R^y_{m,n} \overset{a.s.}{=} \frac{1}{2}m(m-1)\int_{\varphi(y)}^{\infty}\big[\widehat{S}(u|y) - S(u|y)\big]^2 b_y^{m-2}(u)\,du.$$

By the Law of Iterated Logarithms, we know that $\sup_u \left|\widehat{S}(u|y) - S(u|y)\right| \overset{a.s.}{\leq} C\Big(\frac{\log\log n}{n}\Big)^{1/2}$ for some constant C, so we have

$$\frac{\sqrt{n}}{\sigma(m,y)}\left|R^y_{m,n}\right| \overset{a.s.}{\leq} \frac{1}{2}\,\frac{m(m-1)}{\sigma(m,y)}\,\frac{C^2\log\log n}{\sqrt{n}}\,\int_{\varphi(y)}^{\infty} b_y^{m-2}(u)\,du. \tag{2.8}$$

Let us now analyze the behavior of $\int_{\varphi(y)}^{\infty} b_y^m(u)\,du$ when $m\to\infty$. We can write

$$\int_{\varphi(y)}^{\infty} b_y^m(u)\,du = \int_{\varphi(y)}^{\infty}\big(S(u|y)+r_y(u)\big)^m\,du,$$

for some $r_y(u)$ such that $\widehat{S}(u|y)\wedge S(u|y) - S(u|y) \leq r_y(u) \leq \widehat{S}(u|y)\vee S(u|y) - S(u|y)$. Note that $|r_y(u)| \leq C\Big(\frac{\log\log n}{n}\Big)^{1/2}$. Since $\frac{(S(u|y)+r_y(u))^m - S^m(u|y)}{r_y(u)} = m\big(S(u|y)+g_y(u)\big)^{m-1}$, for some $g_y(u)$ such that $|g_y(u)|\leq |r_y(u)|$, we obtain

$$\begin{aligned}\int_{\varphi(y)}^{\infty}\big(S(u|y)+r_y(u)\big)^m\,du \;&\leq\; \int_{\varphi(y)}^{\infty} S^m(u|y)\,du\\ &+\; mC\Big(\frac{\log\log n}{n}\Big)^{1/2}\int_{\varphi(y)}^{\infty}\big(S(u|y)+g_y(u)\big)^{m-1}\,du.\end{aligned}$$

Applying the same argument for the exponent $m-1$, one can find

$$\begin{aligned}\int_{\varphi(y)}^{\infty}\big(S(u|y)+g_y(u)\big)^{m-1}\,du \;&\leq\; \int_{\varphi(y)}^{\infty} S^{m-1}(u|y)\,du\\ &+\; (m-1)C\Big(\frac{\log\log n}{n}\Big)^{1/2}\int_{\varphi(y)}^{\infty}\big(S(u|y)+h_y(u)\big)^{m-2}\,du.\end{aligned}$$

for some $h_y(u)$ such that $|h_y(u)|\leq|g_y(u)|\leq|r_y(u)|$. It is clear that

$$\begin{aligned}&\int_{\varphi(y)}^{\infty}\big(S(u|y)+h_y(u)\big)^{m-2}\,du \leq \int_{\varphi(y)}^{\infty}\big(\widehat{S}(u|y)\vee S(u|y)\big)^{m-2}\,du\\ &\quad\leq \int_{\varphi(y)}^{\infty}\big(\widehat{S}^{m-2}(u|y)\vee S^{m-2}(u|y)\big)\,du \leq \int_{\varphi(y)}^{\infty}\widehat{S}^{m-2}(u|y)\,du + \int_{\varphi(y)}^{\infty} S^{m-2}(u|y)\,du.\end{aligned}$$

So, when $m\to\infty$, $\int_{\varphi(y)}^{\infty}\big(S(u|y)+h_y(u)\big)^{m-2}\,du \overset{a.s.}{=} o(1)$. So finally we obtain when $m\to\infty$,

$$\begin{aligned}\int_{\varphi(y)}^{\infty} b_y^m(u)\,du \;&\overset{a.s.}{\leq}\; \big(\varphi_m(y)-\varphi(y)\big) + mC\Big(\frac{\log\log n}{n}\Big)^{1/2}\big(\varphi_{m-1}(y)-\varphi(y)\big)\\ &+m(m-1)C^2\Big(\frac{\log\log n}{n}\Big)\,o(1).\end{aligned} \tag{2.9}$$

Plugging in (2.8) the results (2.9) and (2.5) and using Lemma 2.1, we obtain for $m\to\infty$,

$$\begin{aligned}\frac{\sqrt{n}}{\sigma(m,y)}\left|R^y_{m,n}\right| \;&\overset{a.s.}{\leq}\; \frac{C^2m^2}{2\sqrt{k_{1,y}}m^{1/2-1/\rho_y}}\,\frac{\log\log n}{\sqrt{n}}\Bigg\{\big[\Gamma(1+1/\rho_y)\ell^{-1/\rho_y}+o(1)\big]\\ &\times\;\Big(m^{-1/\rho_y} + mC\big(\frac{\log\log n}{n}\big)^{1/2}m^{-1/\rho_y}\Big) + m^2C^2\frac{\log\log n}{n}\,o(1)\Bigg\},\end{aligned}$$

so that

$$\frac{\sqrt{n}}{\sigma(m,y)}\left|R^y_{m,n}\right| \overset{a.s.}{\leq} m^{3/2}\frac{\log\log n}{\sqrt{n}}(K_1+o(1)) + m^{5/2}\frac{(\log\log n)^{3/2}}{n}(CK_1+o(1)) + m^{7/2+1/\rho_y}\frac{(\log\log n)^2}{n^{3/2}}\,o(1), \tag{2.10}$$

where K_1 is some positive constant. Since under the condition of the theorem $m = m_n = c\,n^{1/3-\varepsilon}\,(\log\log n)^{-2/3}$ all the terms in the r.h.s. of the last inequality converges to 0 when $n \to \infty$, we obtain

$$\frac{\sqrt{n}}{\sigma(m,y)}\,R^y_{m,n} \overset{a.s.}{\longrightarrow} 0 \ \text{ as } n \to \infty. \tag{2.11}$$

(ii) We now will prove the leading term of (2.7) *converges to a standard normal.* We can rewrite this leading term as

$$\frac{\sqrt{n}\;m}{\sigma(m,y)}\int_{\varphi(y)}^{\infty} S^{m-1}(u|y)\big[\widehat{S}(u|y) - S(u|y)\big]\,du = \frac{S_Y(y)}{\widehat{S}_Y(u)}\sum_{i=1}^{n}\frac{W_{n,i}}{\sqrt{n}\,\sigma(m,y)},$$

where $W_{n,i} = \big(m/S_Y(y)\big)\int_{\varphi(y)}^{\infty} S^{m-1}(u|y)\big[1\!\!I(X_i \geq u, Y_i \geq y) - S(u|y)\,1\!\!I(Y_i \geq y)\big]\,du$. It is easy to see that $\mathbb{E}(W_{n,i}) = 0$ and $\mathbb{V}(W_{n,i}) = \sigma^2(m,y)$. By the Lindberg-Feller theorem (Serfling, 1980, p. 29) we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{W_{n,i}}{\sigma(m,y)} \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0,1), \ \text{ as } n \to \infty, \tag{2.12}$$

under the Liapounoff condition, *i.e.* if

$$\frac{n\mathbb{E}\big(|W_{n,i}|^3\big)}{\big[n\mathbb{V}(W_{n,i})\big]^{3/2}} \longrightarrow 0, \ \text{ as } n \to \infty. \tag{2.13}$$

The Liapounoff condition is easy to check under the assumptions of the theorem. Indeed, $\mathbb{E}\big(|W_{n,i}|^3\big) = \mathbb{E}\big(W_{n,i}^2|W_{n,i}|\big)$ and since $\big|1\!\!I(X_i \geq u, Y_i \geq y) - S(u|y)\,1\!\!I(Y_i \geq y)\big| \leq 1$, we have

$$|W_{n,i}| \leq \frac{m}{S_Y(y)}\int_{\varphi(y)}^{\infty} S^{m-1}(u|y)du = \frac{m}{S_Y(y)}\big(\varphi_{m-1}(y) - \varphi(y)\big).$$

So, $\mathbb{E}\big(|W_{n,i}|^3\big) \leq \big(m/S_Y(y)\big)\big(\varphi_{m-1}(y) - \varphi(y)\big)\sigma^2(m,y)$ and we obtain

$$\frac{n\mathbb{E}\big(|W_{n,i}|^3\big)}{\big[n\mathbb{V}(W_{n,i})\big]^{3/2}} \leq \frac{m}{\sqrt{n}\,S_Y(y)}\;\frac{\varphi_{m-1}(y) - \varphi(y)}{\sigma(m,y)}.$$

Under the regularity condition (1.7), Lemma 2.1 and (2.5), we have, as $m \to \infty$, $\sigma^2(m,y) \geq k_{1,y} m^{1-2/\rho_y}$ and $\varphi_{m-1}(y) - \varphi(y) \sim \Gamma(1+1/\rho_y)\big(\frac{1}{\ell_y(m-1)}\big)^{1/\rho_y}$, so that

$$\frac{n\mathbb{E}\big(|W_{n,i}|^3\big)}{\big[n\mathbb{V}(W_{n,i})\big]^{3/2}} \leq K_2 \frac{m^{1/2}}{\sqrt{n}},$$

where K_2 is some poistive constant. The r.h.s. of the latter inequality tends to zero if $n \to \infty$ and $m \to \infty$ such that $m/n \to 0$ which is the case for the sequence $m = m_n$ given in the assumption of the theorem. Finally, since $\big(S_Y(y)/\widehat{S}_Y(y)\big) \xrightarrow{a.s.} 1$, as $n \to \infty$, we obtain the desired result. □

Rate of convergence

It is interesting to analyze the resulting rate of convergence of the estimator as a function of n. We have as $n \to \infty$, $\tau_n\big(\hat{\varphi}_m(y) - \varphi_m(y)\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ with $\tau_n = \sqrt{n}/\sigma(m,y)$ and $m = m_n = c\, n^{1/3-\varepsilon}\,(\log\log n)^{-2/3}$. We know by Lemme 2.1 that as $n \to \infty$,

$$k_{1,y} c^{1-2/\rho_y} n^{(1/3-\varepsilon)(1-2/\rho_y)-1} (\log\log n)^{-(2/3)(1-2/\rho_y)} \leq \tau_n^{-2}$$
$$\leq k_{2,y} c^{2-2/\rho_y} n^{(1/3-\varepsilon)(2-2/\rho_y)-1} (\log\log n)^{-(2/3)(2-2/\rho_y)}.$$

We remember that $\rho_y = \beta_y + q + 1$, where $q \geq 1$ and $\beta_y > -1$ (see the discussion after (1.9) above). In the particular case where the extreme value index $\rho_y \geq 2$ we get as $n \to \infty$

$$c_1 n^{-(1/3)(1-1/\rho_y)+1/2} (\log\log n)^{(1/3)(2-2/\rho_y)} \leq \tau_n \leq c_2 n^{1/2} (\log\log n)^{(1/3)(1-2/\rho_y)}.$$

This case is of particular interest when the joint density of (X,Y) has a jump at the frontier (i.e. $\beta_y = 0$, an often used assumption in the econometric literature). We have clearly in this case as $q \downarrow 1$,

$$c_1\big(n\log\log n\big)^{1/3} \leq \tau_n \leq c_2 n^{1/2},$$

and as $q \uparrow \infty$,

$$c_1 n^{1/6} (\log\log n)^{2/3} \leq \tau_n \leq c_2 n^{1/2} (\log\log n)^{1/3}.$$

So, even if the data dimension explodes, the convergence rate does not deteriorate too much avoiding thus, in a sense and partly, the “curse of dimensionality” that is typical of many nonparametric estimators.

2.2 Estimation of the frontier $\varphi(y)$

Since $\varphi_m(y) \to \varphi(y)$ as $m \to \infty$, the result of the preceding section can be used to define an estimator of the "full" frontier itself. From Theorem 2.1, if $m_n < n^{1/3}(\log\log n)^{-2/3}$, we have

$$\frac{\sqrt{n}}{\sigma(m_n, y)}\big(\hat{\varphi}_{m_n}(y) - \varphi(y) - B_{m_n}(y)\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \tag{2.14}$$

where from (2.5),

$$B_{m_n}(y) = \varphi_{m_n}(y) - \varphi(y) = \Gamma\left(1 + \frac{1}{\rho_y}\right)\left(\frac{1}{m_n\,\ell_y}\right)^{1/\rho_y} + o(m_n^{-1/\rho_y}). \tag{2.15}$$

We see that the value of the bias introduced by using the order-m_n estimator to estimate the full frontier is bounded below $\big(\sqrt{n}/\sigma(m_n, y)\big)B_{m_n}(y) > K_3 n^{1/3}(\log\log n)^{1/3}$ for some constant K_3, and this does not vanish when $n \to \infty$.

So, in practice for large values of n (and so of m), we will rather use the following asymptotic approximation:

$$\hat{\varphi}_m(y) - \varphi(y) \approx \mathcal{N}(B_m(y), \frac{\sigma^2(m,y)}{n}), \tag{2.16}$$

where for doing practical inference $B_m(y)$ and $\sigma(m, y)$ have to be consistently estimated. A consitent estimator of $\sigma(m, y)$ is provided by a plugging version of (2.2), whereas, a consistent estimator of $B_m(y)$ can be obtained through the leading part of (2.15) once ρ_y and ℓ_y are known or consistently estimated. The next section suggest a way for estimating these two parameters, using the properties of order-m frontiers.

3 Consistent estimators of ρ_y and ℓ_y

We will use here an approach inspired by the classical Pickand's tail index estimator, analyzed and developed in our frontier setup in Daouia et al. (2008). The Pickand's estimator is based on comparing different quantile-type estimators of the frontier. As well known from the literature, and illustrated in Daouia et al., the estimator is rather unstable and provide disappointing results unless the sample size is larger than, say 1000. Daouia et al. (2008) also analyze a moment estimator providing slightly better behavior in moderated sample sizes (say larger than 500).

In this paper, we adapt the approach by using the order-m estimator of the frontier instead of the order-α quantile estimator of the frontier. Indeed, when considering the asymptotic expression for $\varphi_m(y) - \varphi(y)$ given by (2.5) for the values m, am and a^2m, where a is some fixed integer with $a \geq 2$, we see that

$$\lim_{m\to\infty} \frac{\varphi_m(y) - \varphi_{am}(y)}{\varphi_{am}(y) - \varphi_{a^2m}(y)} = a^{1/\rho_y}.$$

This suggests the following estimator

$$\hat{\rho}_y = \log(a) \left\{ \log\left(\frac{\hat{\varphi}_m(y) - \hat{\varphi}_{am}(y)}{\hat{\varphi}_{am}(y) - \hat{\varphi}_{a^2m}(y)} \right) \right\}^{-1}. \tag{3.1}$$

It is also easy to see that

$$\lim_{m\to\infty} \frac{1}{m} \left[\frac{\Gamma(1 + 1/\rho_y)\big(1 - a^{-1/\rho_y}\big)}{\varphi_m(y) - \varphi_{am}(y)} \right]^{\rho_y} = \ell_y,$$

that can lead to the estimator of ℓ_y

$$\hat{\ell}_y = \frac{1}{m} \left[\frac{\Gamma(1 + 1/\hat{\rho}_y)\big(1 - a^{-1/\hat{\rho}_y}\big)}{\hat{\varphi}_m(y) - \hat{\varphi}_{am}(y)} \right]^{\hat{\rho}_y}. \tag{3.2}$$

The consistency of these estimators is provided by the following theorems.

Theorem 3.1. *Under the regularity conditions of Theorem 2.1,*

$$\hat{\rho}_y \xrightarrow{P} \rho_y \text{ and } \hat{\ell}_y \xrightarrow{P} \ell_y \text{ as } n \to \infty, \tag{3.3}$$

for any y such that $S_Y(y) > 0$,

Proof: By Theorem 2.1, we have $\hat{\varphi}_m(y) - \varphi_m(y) = O_p\big(\sigma(m,y)/\sqrt{n}\big)$. Now, by (2.5), and by Lemma 2.1, we obtain

$$\hat{\varphi}_m(y) - \varphi(y) = C_y \left(\frac{1}{m}\right)^{1/\rho_y} + o\big(m^{-1/\rho_y}\big) + O_p\Big(\frac{m^{1-1/\rho_y}}{\sqrt{n}}\Big)$$

where $C_y = \Gamma\left(1 + \frac{1}{\rho_y}\right)\left(\frac{1}{\ell_y}\right)^{1/\rho_y}$. Similarly we have for all $a \geq 2$

$$\hat{\varphi}_{am}(y) - \varphi(y) = C_y \left(\frac{1}{am}\right)^{1/\rho_y} + o\big(m^{-1/\rho_y}\big) + O_p\Big(\frac{m^{1-1/\rho_y}}{\sqrt{n}}\Big)$$
$$\hat{\varphi}_{a^2m}(y) - \varphi(y) = C_y \left(\frac{1}{a^2m}\right)^{1/\rho_y} + o\big(m^{-1/\rho_y}\big) + O_p\Big(\frac{m^{1-1/\rho_y}}{\sqrt{n}}\Big).$$

Now by doing the differences we have

$$m^{1/\rho_y}\big(\hat{\varphi}_m(y) - \hat{\varphi}_{am}(y)\big) = C_y\big(1 - 1/a^{1/\rho_y}\big) + o(1) + O_p\Big(\frac{m}{\sqrt{n}}\Big)$$
$$(am)^{1/\rho_y}\big(\hat{\varphi}_{am}(y) - \hat{\varphi}_{a^2m}(y)\big) = C_y\big(1 - 1/a^{1/\rho_y}\big) + o(1) + O_p\Big(\frac{m}{\sqrt{n}}\Big),$$

leading to

$$\frac{\hat{\varphi}_m(y) - \hat{\varphi}_{am}(y)}{\hat{\varphi}_{am}(y) - \hat{\varphi}_{a^2m}(y)} = a^{1/\rho_y} \frac{C_y\big(1 - 1/a^{1/\rho_y}\big) + o(1) + O_p\Big(\frac{m}{\sqrt{n}}\Big)}{C_y\big(1 - 1/a^{1/\rho_y}\big) + o(1) + O_p\Big(\frac{m}{\sqrt{n}}\Big)}.$$

As $m/\sqrt{n} \to 0$ as $n \to \infty$, the ratio on the right hand side converges in probability to 1, so that

$$\frac{\hat{\varphi}_m(y) - \hat{\varphi}_{am}(y)}{\hat{\varphi}_{am}(y) - \hat{\varphi}_{a^2m}(y)} \xrightarrow{P} a^{1/\rho_y},$$

which gives $\hat{\rho}_y \xrightarrow{P} \rho_y$. On the other hand, since

$$m^{1/\rho_y}\big(\hat{\varphi}_m(y) - \hat{\varphi}_{am}(y)\big) = \Gamma(1 + 1/\rho_y)(1 - 1/a^{1/\rho_y})(\ell_y)^{-1/\rho_y} + o(1) + O_p\Big(\frac{m}{\sqrt{n}}\Big),$$

we have by using $m/\sqrt{n} \to 0$ and $\hat{\rho}_y \xrightarrow{P} \rho_y$ as $n \to \infty$,

$$\frac{1}{m}\left[\frac{\Gamma(1 + 1/\hat{\rho}_y)\big(1 - a^{-1/\hat{\rho}_y}\big)}{\hat{\varphi}_m(y) - \hat{\varphi}_{am}(y)}\right]^{\hat{\rho}_y} \xrightarrow{P} \ell_y,$$

which gives $\hat{\ell}_y \xrightarrow{P} \ell_y$. $\square$

Practical choice of a and m

The choice of an optimal a and m is an open theoretical issue, but in practice, in the examples and simulations below, we have chosen for m, $m_n = N_y^{1/3}$, where $N_y = \sum_{i=1}^{n} 1\!\!I(Y_i \geq y)$ is the number of observations with $Y_i \geq y$. This choice guarantees by Theorem 2.1 the regular behavior of the estimator $\hat{\varphi}_{m_n}(y)$ as $n \to \infty$ and as seen above, it guarantees also the consistency of the estimators $\hat{\rho}_y$ and $\hat{\ell}_y$. The choice of $a \geq 2$ is much less important: the results are rather stable relative to this choice. Higher values of a will give more weights to extreme data points. It turns out that in all the Monte-Carlo experiments below, the choice $a = 10$ provided quite reasonable estimates with nice behavior of the estimators. When working with particular samples, and for the estimation of ρ_y, we have to tune the choice of a and m more carefully to obtain sensible results (see below), but even in these cases, for estimating the frontier function, in a second step, the choice $a = 10$ and $m = N_y^{1/3}$ provided always nice results.

For the final evaluation of the confidence intervals for $\varphi(y)$, we use the normal approximation centered at $\tilde{\varphi}(y)$, the bias-corrected order-m estimate:

$$\tilde{\varphi}(y) = \hat{\varphi}_m(y) - \widehat{B}_m(y), \tag{3.4}$$

where $\widehat{B}_m(y)$ is the plug-in version of $B_m(y)$, replacing ρ_y and ℓ_y by their consistent estimators derived above. Of course by doing so, we will increase the variance of the estimator $\hat{\varphi}_m(y)$, so we estimate this variance by a bootstrap algorithm. This is illustrated in the next section.

4 Illustrative Examples

4.1 Some Monte-Carlo experiments

To facilitate the comparison with the results obtained in Daouia et al. (2008), we have chosen in the illustrations the output orientation[3]. Here, the bias corrected regularized estimator is given by $\tilde{\varphi}(x) = \hat{\varphi}_m(x) + \widehat{B}_m(x)$.

Uniform distribution

We first simulate, as in Daouia et al., random samples (X_i, Y_i), $i = 1, \ldots, n$ uniformly distributed on the triangle limited by the frontier $\varphi(x) = x$ with $0 \le x \le 1$. Table 1 displays the results. The estimation is performed for $x = 1$, so that the sample sizes n coincide with the "effective" sample size N_x, the number of observation at the left of $x = 1$. We computed also the estimators with the known true value of ρ, which in this example is $\rho_0 = 2$.

We observe a nice behavior of our estimators, with an increasing accuracy, as expected, when the effective sample size N_x increases. The estimation of ρ and ℓ is not an easy task, but still we have a reasonable behavior, with the simple rule we have chosen for m and a: $m = N_x^{1/3}$ and $a = 10$. The estimator $\tilde{\varphi}$ has a very nice behavior for all values of N_x. It is a regularized estimator with an approximate normal distribution but in addition, it has much better properties than the usual FDH estimator (both in term of bias and mean squared error). It should be noticed, that the estimation of the frontier is stable to the choice of the order-m base estimator because the correction for the bias performs quite well for most of the chosen values of m. This is not true for the estimation of ρ and ℓ, even if we have nice results: here the choice of an optimal m and a remains an open issue, and mainly for the estimation of the tail index ρ. The cost of estimating ρ (which in most econometric applications is supposed to be equal to $p + 1$, i.e. there is a jump of the joint density of (X, Y) at the frontier) appears clearly when comparing the results for the estimation of the frontier when the true value of $\rho = 2$ is known: they are much better.

Finally, by looking to Tables 1 and 3 in Daouia et al. (2008) using also Pickand's estimator of ρ, but with quantile-type frontiers, we see that we obtain much more accurate estimators of both ρ and φ. To summarize this comparison, we have here, in the same scenario, with

[3] We can find in the appendix the change of the notations.

comparable sample sizes, a gain of the MSE when estimating ρ by a factor of the order 1000, 8, 6 and 1.3 for the samples sizes 100, 500, 1000 and 5000, respectively. In addition, for this comparison, we selected in the Tables from Daouia et al., the best order of the quantile whereas here, we selected the order m and a given by our simple rule.

For the estimation of the frontier point, the gain of the MSE has a factor of the order 450, 70, 150 and 60 (respectively). We observe also some qualitative gain when estimating the frontier when ρ is known, but here the gain is of a factor ranging from 2.5 to 5 when N_x goes from 500 to 5000. Again in this comparison, we selected, in the results from the Tables in Daouia et al., the best value of the quantile order.

Also, as a general comparison bewtween the two approaches (using the order-m here and using the order-α quantile as in Daouia et al., 2008), we can say that with the approach here, we gain a lot in terms of the stability of the estimators with respect to the choice of the order of the base estimator. Tables 1 and 3 in Daouia et al. indicate indeed a huge sensitivity to the choice of the quantile order when defining the base estimator (the MSE can change by a factor of several thousands if the wrong order is selected) and this was not the case here where we observe a great stability in the estimation of the frontier (we do not reproduce the results to save place).

Table 1: *Bias (Bias) and Mean Squared Error (MSE) of the estimates over 1000 Monte-Carlo simulations: Uniform case, true values are* $\varphi_0 = 1$, $\rho_0 = 2$ *and* $\ell_0 = 1$

	$N_x = 100$		$N_x = 500$		$N_x = 1000$		$N_x = 5000$	
	Bias	*MSE*	*Bias*	*MSE*	*Bias*	*MSE*	*Bias*	*MSE*
$\hat{\rho}$	-0.423441	0.405546	-0.086590	0.167862	-0.030069	0.103272	0.006700	0.032857
$\hat{\ell}$	0.129996	0.143781	0.134017	0.140254	0.123555	0.109258	0.090057	0.052087
$\hat{\ell}(\rho_0)$	0.436484	0.357537	0.183585	0.063740	0.129845	0.031946	0.071624	0.009915
$\tilde{\varphi}$	-0.070644	0.008321	-0.017592	0.001231	-0.008925	0.000501	-0.002371	0.000087
$\tilde{\varphi}(\rho_0)$	-0.035497	0.003719	-0.011318	0.000603	-0.006952	0.000255	-0.002778	0.000053
$\hat{\varphi}_{FDH}$	-0.090498	0.010401	-0.040257	0.002071	-0.028140	0.000993	-0.012811	0.000206

Beta densities for the efficiency term

Now, we analyze the results with different behaviors of the density of the efficiencies at the frontier points (density tending to infinity, having a jump or converging to zero at the frontier points). We select the following model $Y = X\,V$ where $X \sim \text{Unif}(0, 1)$ and $V \sim \text{Beta}(\beta, \beta)$ with values of $\beta = 0.5, 1$ and 3. Note that in all the cases, $\mathbb{E}(V) = 0.5$. Again we focus the results for the value $x = 1$, so that $N_x = n$. The results are shown in Tables 2 to 4. In the first case the density tends to infinity at the frontier, and the FDH estimator should be performant. It is indeed the case but our regularized estimator do even slightly better for $N_x = 100$ but much better for larger N_x reaching both less Bias and MSE. Again, the estimation of ρ and ℓ is more difficult but our rule of thumb ($m = N_x^{1/3}$ and $a = 10$) shows nice behavior of the estimators. When β increases (jump at the frontier for $\beta = 1$ and going smoothly to zero when $\beta = 3$), the results for the estimators of the frontier deteriorate a little, as expected but our regularized estimator is always better than the FDH estimator both for bias and MSE. In this latter case, we illustrate the estimation of the frontier for the full range of X in the next section.

Table 2: *Bias (Bias) and Mean Squared Error (MSE) of the estimates over 1000 Monte-Carlo simulations: case of the* $Beta(0.5, 0.5)$*, true values are* $\varphi_0 = 1$*,* $\rho_0 = 1.5$ *and* $\ell_0 = 0.4244$*.*

	$N_x = 100$		$N_x = 500$		$N_x = 1000$		$N_x = 5000$	
	Bias	*MSE*	*Bias*	*MSE*	*Bias*	*MSE*	*Bias*	*MSE*
$\hat{\rho}$	-0.063176	0.168646	0.146422	0.111090	0.123820	0.061474	0.096074	0.021250
$\hat{\ell}$	0.310572	0.124454	0.237085	0.075298	0.201660	0.052083	0.141359	0.024371
$\hat{\ell}(\rho_0)$	0.341740	0.138351	0.188876	0.039628	0.153513	0.025683	0.092683	0.009117
$\tilde{\varphi}$	-0.053538	0.005929	-0.007142	0.000553	-0.004477	0.000217	-0.000053	0.000022
$\tilde{\varphi}(\rho_0)$	-0.048796	0.005060	-0.020620	0.000787	-0.014593	0.000391	-0.005882	0.000062
$\hat{\varphi}_{FDH}$	-0.072121	0.007519	-0.024156	0.000848	-0.016576	0.000391	-0.005347	0.000042

Table 3: *Bias (Bias) and Mean Squared Error (MSE) of the estimates over 1000 Monte-Carlo simulations: case of the* $Beta(1, 1)$*, true values are* $\varphi_0 = 1$*,* $\rho_0 = 2$ *and* $\ell_0 = 0.5$*.*

	$N_x = 100$		$N_x = 500$		$N_x = 1000$		$N_x = 5000$	
	Bias	*MSE*	*Bias*	*MSE*	*Bias*	*MSE*	*Bias*	*MSE*
$\hat{\rho}$	-0.331901	0.444470	0.086348	0.248021	0.109707	0.135392	0.138556	0.065002
$\hat{\ell}$	0.359753	0.193566	0.280033	0.109485	0.253591	0.087861	0.201901	0.053630
$\hat{\ell}(\rho_0)$	0.506769	0.342331	0.273330	0.088737	0.222238	0.055914	0.141044	0.021551
$\tilde{\varphi}$	-0.091613	0.015110	-0.018562	0.002537	-0.009895	0.001006	0.000817	0.000188
$\tilde{\varphi}(\rho_0)$	-0.059530	0.007942	-0.027354	0.001641	-0.019782	0.000782	-0.009198	0.000161
$\hat{\varphi}_{FDH}$	-0.120797	0.018561	-0.055114	0.003931	-0.039773	0.001989	-0.017015	0.000375

Table 4: *Bias (Bias) and Mean Squared Error (MSE) of the estimates over 1000 Monte-Carlo simulations: case of the* $Beta(3, 3)$*, true values are* $\varphi_0 = 1$*,* $\rho_0 = 4$ *and* $\ell_0 = 2.5$*.*

	$N_x = 100$		$N_x = 500$		$N_x = 1000$		$N_x = 5000$	
	Bias	*MSE*	*Bias*	*MSE*	*Bias*	*MSE*	*Bias*	*MSE*
$\hat{\rho}$	-2.053562	4.921744	-0.825120	4.708721	-0.630922	2.096024	-0.419817	0.647445
$\hat{\ell}$	-1.068665	1.595360	-1.057668	1.413049	-1.042189	1.310455	-0.967238	1.047182
$\hat{\ell}(\rho_0)$	0.348089	5.674396	-0.585639	0.849339	-0.670422	0.725935	-0.742065	0.634133
$\tilde{\varphi}$	-0.199484	0.051431	-0.063119	0.030965	-0.041521	0.013676	-0.018208	0.003018
$\tilde{\varphi}(\rho_0)$	-0.019372	0.008809	0.005711	0.002002	0.008604	0.001276	0.011473	0.000432
$\hat{\varphi}_{FDH}$	-0.237955	0.061255	-0.155909	0.026241	-0.131248	0.018605	-0.086933	0.008159

4.2 Estimation of the frontier function

One simulated sample

We first illustrate the behavior of the frontier estimate in the case of a beta density for the efficiencies, with the model described in the preceding subsection. We show the case where the density is converging smoothly to zero at the frontier ($\beta = 3$). For estimating the frontier function over the full range of X, it seems reasonable to assume that the function $\rho_x = \rho$ is constant over the range of X (which is true in the simulated scenario). We estimate this value by a trimmed mean of the local values $\hat{\rho}_x$ computed over a fixed grid of 100 values of x from 0.25 till 1 (the trimming is used to eliminate local numerical instability when computing (3.1)). In this step of estimating ρ_x, we got better results by taking $m = N_x^{1/3}$ but $a = 5$. We obtained the value 3.5126 where the true value is 4. Then we compute the values $\hat{\ell}_x$ and $\tilde{\varphi}(x)$ on the same grid of values for x. Here we have chosen, as in the Monte-Carlo experiments, the value of $m = N_x^{1/3}$ and $a = 10$. The 95% confidence intervals for each value of x were obtained by using the normal approximation, centered on $\tilde{\varphi}(x)$ and variance estimated by a bootstrap algorithm.

The results are quite sensible and we see in Figure 1 that our estimate is better than the FDH estimate (closer to the true frontier). On the left panel we see clearly that the pointwise confidence intervals cover the true frontier. In particular it appears in this sample that the FDH estimator is even outside the 95% confidence intervals for all x. We can also appreciate the robustness of the frontier estimate (relative to the FDH estimator) looking to the right panel, when we add one outlier in the data set (keeping our original estimate of ρ). Of course, in practice, we could easily detect this outlier (even for dimension $p > 1$, because it is far outside the confidence interval at this point). Once this is observed, and as always when detecting potentail outliers, this point could be removed from the sample only after a careful analysis.

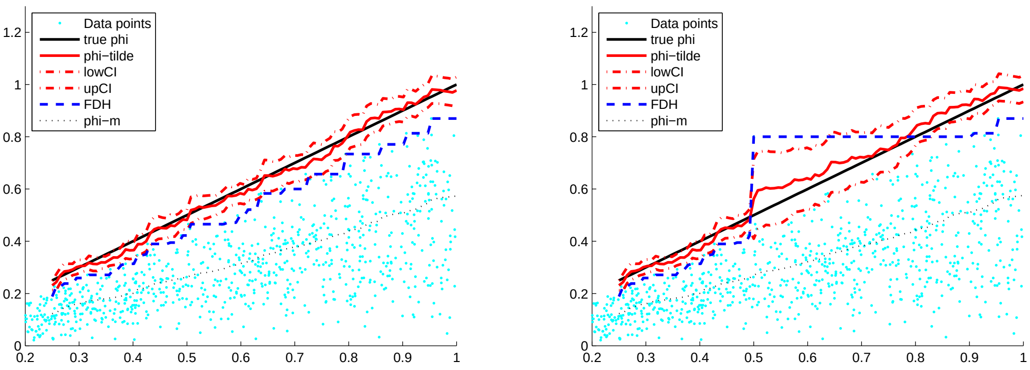

Figure 1: *Linear frontier:* $Y = XV$ *with* $V \sim \mathrm{Beta}(3,3)$ *and* $n = 1000$. *Right panel, with one outlier at* $X_{n+1} = 0.5, Y_{n+1} = 0.8$. *The base (biased) estimator 'phi-m' is* $\hat{\varphi}_m(x)$ *and our regularized estimator 'phi-tilde' is* $\tilde{\varphi}(x)$.

One real data example

We use the same real data example as in Cazals et al. (2002) and Daouia et al. (2008) on the frontier analysis of 9521 French post offices observed in 1994, with X as the quantity of labor and Y as the volume of delivered mail. In this illustration, we only consider the $n = 4000$ observed post offices with the smallest levels x_i.

We first start by assuming, as in most econometric frontier studies, that the joint density of (X, Y) has a jump on the frontier, so $\rho_x = p+1 = 2$. The cloud of points and the resulting estimates are provided in Figure 2. The FDH estimator is clearly determined by only a few very extreme points. If we delete 4 extreme points from the sample (represented by circles in the figures), we obtain the pictures of the right panel: the FDH estimator changes drastically, whereas the regularized estimator is very robust to the presence of these 4 extreme points. Again the confidence intervals were obtained by a bootstrap algorithm.

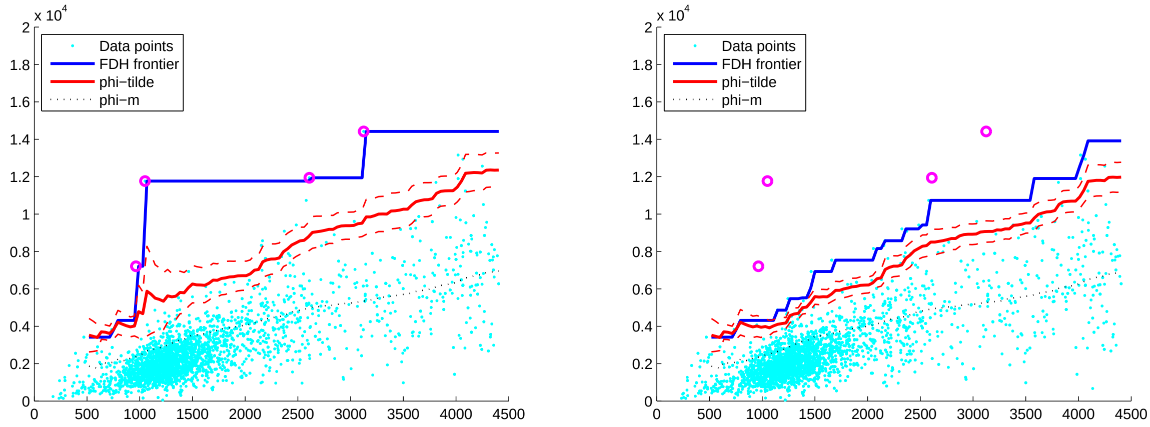

Figure 2: *Resulting estimator $\tilde{\varphi}(x)$ for the French post offices with $\rho = 2$. In the left panel, the 4 extreme data points (circles) are used in the estimation of the two frontier functions.*

If we prefer to estimate ρ_x, we proceed as above by assuming that $\rho_x = \rho$ is an unknown constant and we average the values of $\hat{\rho}_x$ obtained over a grid of values of x (again with a trimmed mean). In our case here, larger values of m were needed for computing $\hat{\rho}_x$, to avoid numerical instabilities in (3.1): we choose $m = 10N_x^{1/3}$ keeping $a = 10$. In this first step estimation of ρ, we only used the sample without the 4 outliers detected above. This provided the estimator of $\hat{\rho} = 3.3465$, indicating that the density of the efficiencies is tending to zero at the frontier, but not its first derivative; a reasonable result when looking to the cloud of data points in Figure 2.

Then we proceed as usual for estimating the frontier with the full sample, as if the true value of the tail index would be 3.3465, keeping the basic rule of thumb $m = N_x^{1/3}$ and $a = 10$, as in the Monte-Carlo exercices above. The results are displayed on the left panel of Figure 3, we see that this higher value of $\hat{\rho}$ (compared to $\rho = 2$ in Figure 2) push our estimator to the North, as expected, because the correction for the bias is larger.

The right panel of the figure, where the 4 extreme data points are excluded from the sample, indicates how the frontier estimate is robust to the outliers (as compared to FDH). We observe again that after the first outlier, most of the FDH frontier is outside the 95% confidence intervals.

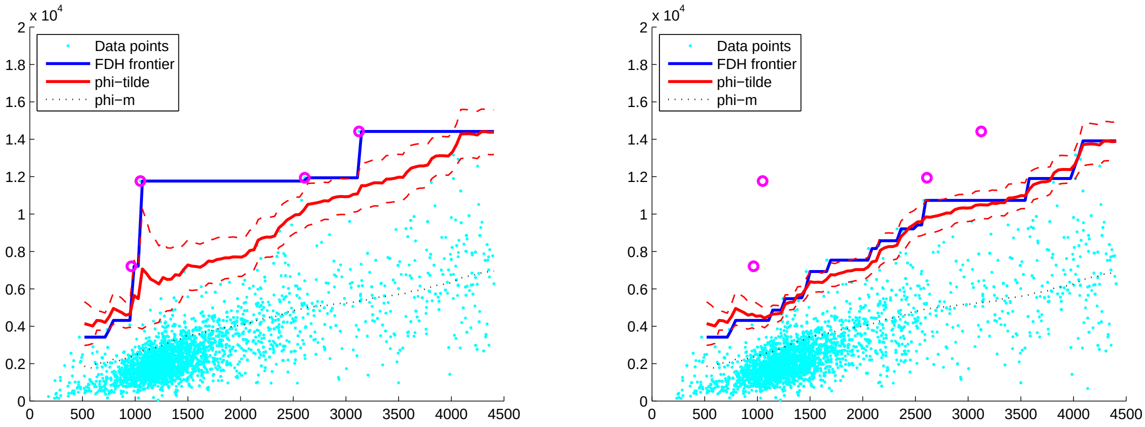

Figure 3: *Resulting estimator* $\tilde{\varphi}(x)$ *for the French post offices with* $\hat{\rho} = 3.3465$. *In the left panel, the 4 extreme data points (circles) are used in the estimation of the two frontier functions.*

5 Conclusions

We have derived in this paper the theory of an estimator of a frontier function having an asymptotic normal distribution. The basic tools is the order-m partial frontier where we let the order m to converge to infinity when $n \to \infty$ but at a slow rate. Indeed if the rate is too fast, the order-m frontier will converge to quickly to the full frontier and the corresponding estimator will converge to the FDH estimator, having a Weibull limiting distribution. The final estimator is then corrected for its inherent bias. We thus can view our estimator as a regularized frontier estimator which, in addition, is more robust to extreme values and outliers than the usual nonparametric frontier estimators.

In addition, if the tail index ρ_y and the behavior of the conditional distribution of X given that $Y \geq y$ near the frontier points is not known (ℓ_y), we provide an easy way to estimate them consistently.

The performances of our estimators are evaluated in finite samples trough some Monte-Carlo experiments, showing very nice regular behavior of the estimators in particular for the estimator of the frontier. We illustrate also how to provide, in an easy way, confidence intervals for the frontier function in a simulated data set where the FDH estimator gives very poor results. We also illustrate our procedure with a real data set from the French Post Offices.

Important research issues are still open and deserve for future work. This includes a way for selecting an optimal m, which is particularly important for deriving the estimator of the tail index ρ_y. But this is known as a difficult problem in extreme values. Once ρ_y is well estimated (or assumed to be known) the estimate of the frontier itself is much more robust to the choice of the order m. An other trail of research would be to define estimators of ρ_y and ℓ_y when they are considered as smoothed function of y.

Appendix: The Output Oriented Case

In this section we only give the useful notations and formulas for the output oriented case. Here the attainable production set is defined as $\Psi = \{(x, y) \in \mathbb{R}_+^p \times \mathbb{R}_+ \mid x \text{ can produce } y\}$ and the production frontier is represented by the graph of the production function $\varphi(x) = \sup\{y \,|\, (x, y) \in \Psi\}$.. The distribution function of (X, Y) can be denoted $F(x, y)$ and $F(\cdot|x) = F(x, \cdot)/F_X(x)$ will be used to denote the conditional distribution function of Y given $X \leq x$, with $F_X(x) = F(x, \infty) > 0$. It has been proven in Cazals et al. (2002) that under the free disposability assumption, the production function can equivalently be defined by

$$\varphi(x) = \sup\{y \geq 0 | F(y|x) < 1\} \tag{A.1}$$

The order-m partial frontier is now defined as

$$\varphi_m(x) = \mathbb{E}\big[\max(Y_1, \ldots, Y_m) | X \leq x\big], \tag{A.2}$$

where $(Y_1, \ldots, Y_m)$ are m i.i.d. random variables generated by the conditional distribution of Y given $X \leq x$. It is shown in Cazals et al. that $\varphi_m(x) = \int_0^\infty \big(1 - [F(u|x)]^m\big)\, du = \varphi(x) - \int_0^{\varphi(x)} [F(u|x)]^m\, du$, so that $\varphi_m(x) \to \varphi(x)$ as $m \to \infty$.

Nonparametric estimators of these frontiers are obtained by plugging the empirical version of the unknown distribution $F(\cdot|x)$ in the definition above. So we obtain

$$\hat{\varphi}(x) = \sup\{y \geq 0 | \widehat{F}(y|x) < 1\} = \max_{\{i: X_i \leq x\}} Y_i \tag{A.3}$$

$$\hat{\varphi}_m(x) = \hat{\varphi}(x) - \int_0^{\hat{\varphi}(x)} [\widehat{F}(u|x)]^m\, du, \tag{A.4}$$

where $\widehat{F}(y|x) = \widehat{F}(x, y)/\widehat{F}_X(x)$ with $\widehat{F}(x, y) = 1/n \sum_{i=1}^n 1\!\!I(X_i \leq x, Y_i \leq y)$ and $\widehat{F}_X(x) = 1/n \sum_{i=1}^n 1\!\!I(X_i \leq x)$. For any given x and a fixed value of m, we have as $n \to \infty$,

$$\frac{\sqrt{n}}{\sigma(m, x)} \big(\hat{\varphi}_m(x) - \varphi_m(x)\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1), \tag{A.5}$$

where the variance can be written, as in (2.2), as

$$\sigma^2(m, x) = \frac{2m^2}{F_X(x)} \int_0^{\varphi(y)} \int_0^{\varphi(y)} F^m(y|x) F^{m-1}(u|y)(1 - F(u|x))\, 1\!\!I(y \leq u)\, dy\, du. \tag{A.6}$$

The regularity condition can be written here as

$$F_X(x)(1 - F(y|x)) = \ell_x(\varphi(x) - y)^{\rho_x} + o(\varphi(x) - y)^{\rho_x}, \quad \text{as } y \uparrow \varphi(x), \tag{A.7}$$

where $\ell_x > 0$, $\rho_x > p$ and $\varphi(x)$ is differentiable in x with strictly positive first partial derivatives. Then, from the equation (2.5) in Daouia et al. (2008), we obtain the useful relation, as $m \to \infty$,

$$\varphi(x) - \varphi_m(x) = \Gamma\left(1 + \frac{1}{\rho_x}\right)\left(\frac{1}{m\,\ell_x}\right)^{1/\rho_x} + o(m^{-1/\rho_x}). \tag{A.8}$$

Then, the asymptotic theory given in Sections 2 and 3 can easily be adapted.

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

[3] Charnes, A., Cooper W.W. and E. Rhodes (1978), Measuring the inefficiency of decision making units, *European Journal of Operational Research* 2 (6), 429-444.

[4] Daouia, A., J.P. Florens and L. Simar (2008), Frontier estimation and Extreme value theory, Discussion paper 0806, Institut de Statistique, UCL.

[5] Daraio, C. and L. Simar (2007), *Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and Applications*, Springer, New-York.

[6] 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.

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

[8] Daouia, A. and I. Gijbels(2009), Robustness and inference in nonparametric partial-frontier modeling, manuscript.

[9] Kneip, A., B. Park and L. Simar (1998), A Note on the Convergence of Nonparametric DEA Estimators for Production Efficiency Scores, *Econometric Theory*, 14, 783-793.

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

[11] Park, B. Simar, L. and Ch. Weiner (2000), The FDH Estimator for Productivity Efficiency Scores : Asymptotic Properties, *Econometric Theory*, Vol 16, 855–877.

[12] Shephard, R.W. (1970). *Theory of Cost and Production Function*. Princeton University Press, Princeton, New-Jersey.

[13] Simar, L. and P.W. Wilson (2008), Statistical Inference in Nonparametric Frontier Models: recent Developments and Perspectives, in *The Measurement of Productive Efficiency*, 2nd Edition, Harold Fried, C.A.Knox Lovell and Shelton Schmidt, editors, Oxford University Press.