Robust Estimation of Conditional Extreme Values

September 2008

Jean-Pierre Florens
(TSE)

(with A. Daouia (TSE) and L. Simar (UCL))

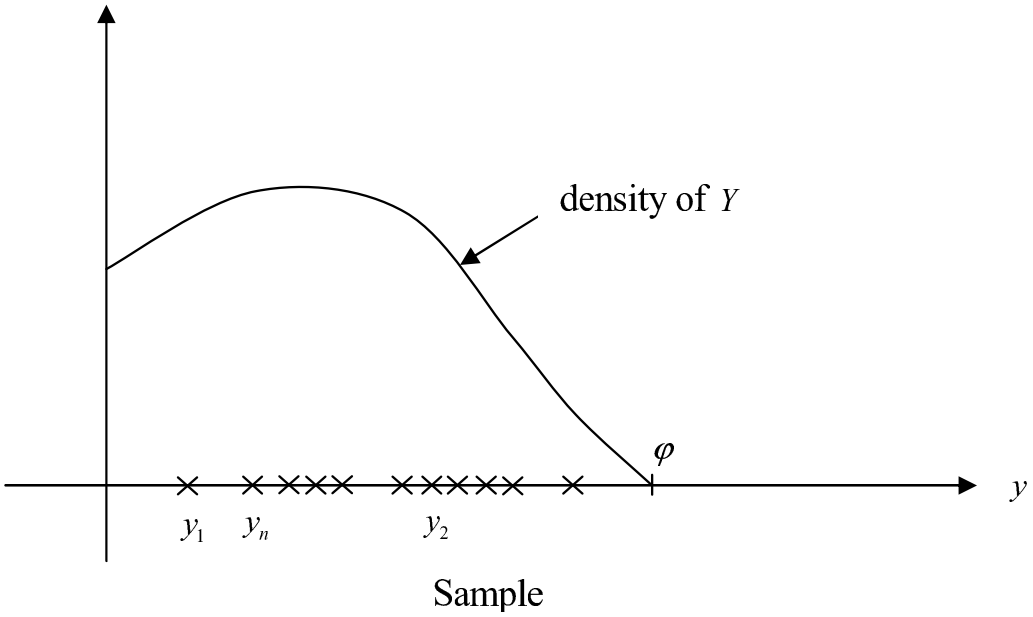
Table of Contents

I - Introduction

- **The elementary model:**

  Y random variable valued in $[0, \varphi]$

  $y_1, ..., y_n$ iid sample of Y Question : Estimation of φ ?



What is the maximum possible value for Y?

Identical theory for the minimum.

Important hypothesis: we assume that φ bounded exists.

- **Numerous applications**

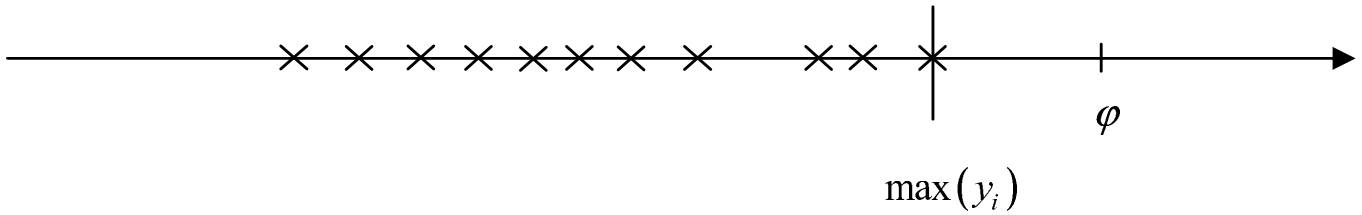
  finance

  insurance

  efficiency in production theory

- **Standard** question in statistics: many books and papers on that field (e.g. de Haan and Ferreira, Extreme Value Theory. Springer 2006).

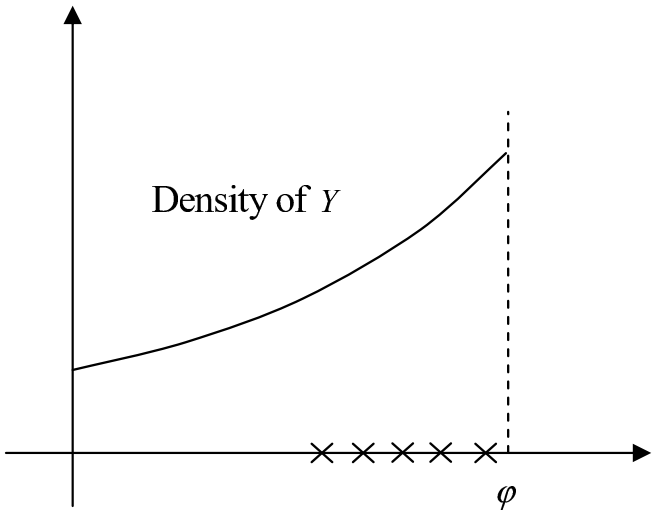
- **Three solutions** to this problem



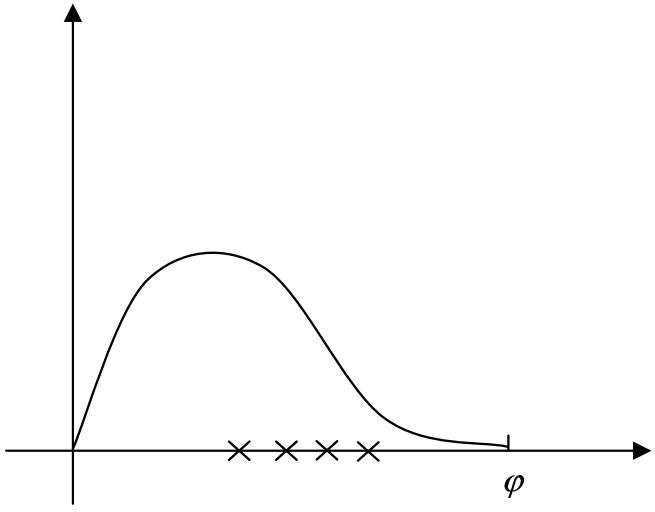
$$\begin{cases} \text{i) } max(y_i) \\ \text{ii)} max(y_i) + \text{ something} \\ \text{iii) a number smaller than } max(y_i) \end{cases}$$

i) : maximum likelihood estimation Well established theory of the property of the maximum.

Intuition about the speed of convergence of $max(y_i)$ to φ



Fast speed convergence



Slow speed convergence

Depends on the shape of the density at the boundary (intuition: the number of derivatives equal to 0 at φ).

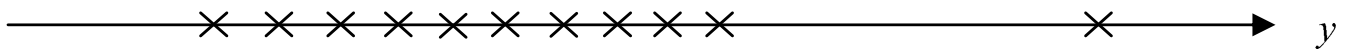
ii) $max(y_i)$ + something depending on the shape of the density closed to the extreme value.

Better small sample properties

Correction of the asymptotic bias

extrapolation "out of the sample".

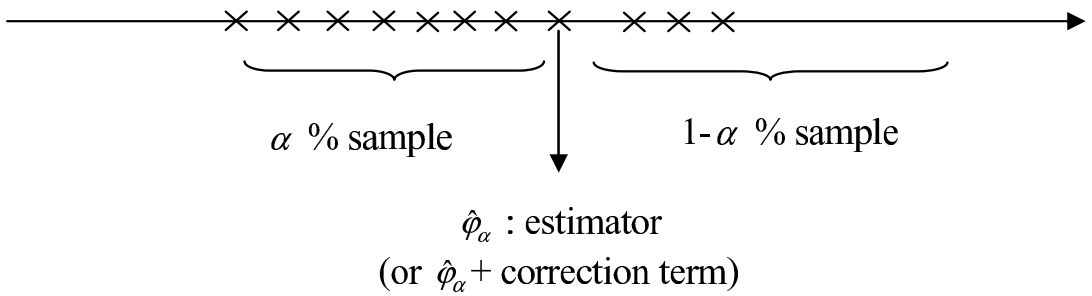
Main question: Extreme sensitivity to outliers!!!



Common practice : elimination of a part of the observations before the estimation of the maximum (errors on the data, heterogenous observations...)

iii) We propose two solutions to address this question:

- Quantile estimation



Idea : select $\alpha \to 1$ where $n \to \infty$ at a correct speed in order to be robust, consistent and to preserve simple asymptotic properties (asymptotic normality instead of complex extreme value distribution).

- m extreme value estimation:
  - drawing of subsamples of size m of the sample
  - computation of the mean of the maxima of these subsamples.

Same analysis but $m \to \infty$ with $n \to \infty$. Convergence to the true extreme, robust to outliers and good asymptotic properties.

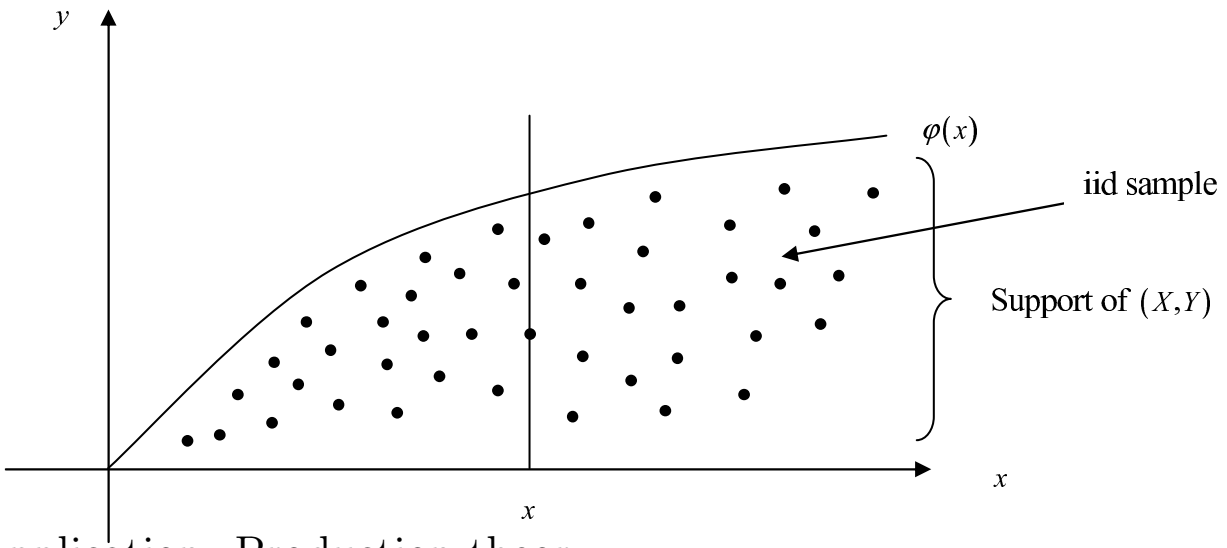
(original paper : Cazals, Florens and Simar (2002)).

- **A more general presentation in the paper: conditional analysis:**

$(Y, \underbrace{X}_{\text{Explanatory variables}})$

Extreme of Y conditional on $X \leq x : \varphi(x)$ function

Frontier of the joint distribution.



Important application: Production theory

X: outputs $\in \mathbb{R}^{\rho}$

Y: input

- the objective is to estimate non parametrically a function $\varphi(x) : \mathbb{R}^{p} \to \mathbb{R}$

Theory of regularized estimation: $\hat{\varphi}_{\alpha}(x)$ or $\hat{\varphi}_{m}(x)$

(α and m are regularization parameters)

The true function φ is assumed to have some regularity property characterized by a regularity index function $\rho(x)$.

II - Definitions

– **Probabilistic definitions:** (X, Y) random element of $\mathbb{R}^p \times \mathbb{R}$ $Y \geq 0$ $X \geq 0$ joint c.d.f

$$\begin{aligned} F(x,y) &= \text{Proba } (X \leq x, Y \leq y) \\ F_X(x) &= \text{marginal c.d.f. } = \text{Proba}(X \leq x) \\ F(y|x) &= \text{Proba } (Y \leq y | X \leq x) \\ &= \frac{F(x,y)}{F_X(x)} \end{aligned}$$

- Conditional extreme value function (frontier)

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

- α quantile function:

$$\varphi_\alpha(x) = inf\{y \geq 0 / F(y|x) \geq \alpha\}$$

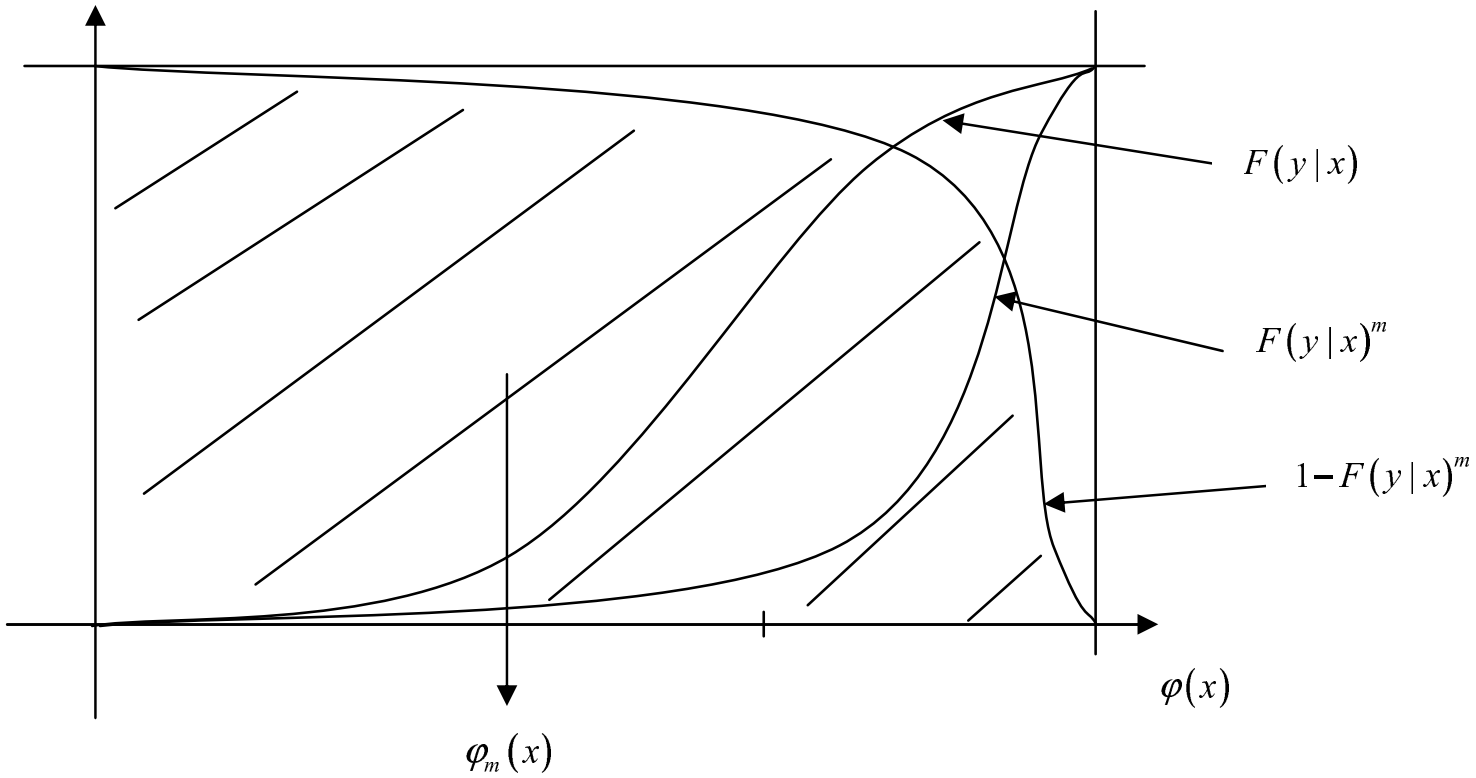
$$\alpha \to 1 \quad \varphi_\alpha(x) \to \varphi_1(x) = \varphi(x)$$

– m-frontier:

$$\varphi_m(x) = \int_0^\infty (1 - (F(y|x))^m) dy$$

$\tilde{Y}_1, ..., \tilde{Y}_m$ m sample conditional on $X \leq x$ c.d.f. of $max(\tilde{Y}_1, ..., \tilde{Y}_m) = F(.|x)^m$

Mean of max = integral of the survivor function $1 - F(.|x)^m$



$m \to \infty\ \varphi_m(x) \to \varphi_\infty(x) = \varphi(x)$

– **Estimation:** $(x_i, y_i) i = 1, ..., n$ iid sample

$$\hat{F}_n(x, y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(x_i \leq x, y_i \leq y)$$

$$\hat{F}_{Xn}(x), \hat{F}_n(y|x)$$

Plug in estimators:

$$\begin{aligned} \hat{\varphi}(x) &= sup\{y \geq 0 | \hat{F}_n(y|x) < 1\} \\ \hat{\varphi}_\alpha(x) &= sup\{y \geq 0 | \hat{F}_n(y|x) \geq \alpha\} \\ \hat{\varphi}_m(x) &= \int_0^\infty (1 - \hat{F}_n(y|x)^m) dy \end{aligned}$$

$\hat{\varphi}_m$ may be equivalently computed by resampling in the conditional empirical distribution.

III - Regularity Scale for $F(.|x)$

– The general $\mathcal{C}_\rho^*$ scale:

$$\rho(x) : \mathbb{R}^\rho \to \mathbb{R} \text{ index function}$$

$F(y|x) \in \mathcal{C}_\rho^*$ if

$$\forall x \quad \lim_{u \downarrow 0} \frac{1 - F(\varphi(x) - uz|x)}{1 - F(\varphi(x) - u|x)} = z^{\rho(x)}$$

$$\Leftrightarrow \forall x \; F_X(x)(1 - F(y|x)) = L_x(\varphi(x) - y)(\varphi(x) - y)^{\rho(x)}$$

where $L_x(\varphi(x) - y) \in \mathcal{C}_0^*$

– A particular subclass: $\mathcal{C}_\rho$

$$\forall x \quad L_x(\varphi(x) - y) = \ell(x)$$

$$y \text{ closed to } \varphi(x)$$

$$\begin{aligned} \Leftrightarrow \qquad F(y|x) \quad &= 1 - \frac{\ell(x)}{F_X(x)}(\varphi(x) - y)^{\rho(x)} \\ f(y|X \le x) \quad &= \frac{\ell(x)\rho(x)}{F_X(x)} \quad (\varphi(x) - y)^{\rho(x)-1} \end{aligned}$$

The index function measures the smoothness at the boundary.

- Examples
  - **Uniform**: (X, Y) are uniformly distributed on $\{(x, y) \mid 0 \leq y \leq x \leq 1\}$:

$$\rho_x = 2 \text{ and } \ell_x = 1$$

  - **Cobb Douglass**: $Y = X^{1/2} \exp(-U)$ with $U \sim \text{Expo}(\lambda = 3)$ and $X \sim U(0, 1)$:

$$\rho_x = 2 \text{ and } L_x(z) = \frac{F_x(x)}{[\varphi(x)]^3} \left[3\varphi(x) - \frac{2}{z} \right] \forall z > 0 \quad \forall x \in]0, 1]$$

  - See below why $\rho_x = 2$ in these examples, ...

IV - Asymptotic theory

– **Non regularized conditional extreme value estimation:**

- If $F(y|x) \in \mathcal{C}_\rho^* \quad \exists b_n$ such that

$$b_n^{-1}(\varphi(x) - \hat{\varphi}(x)) \xrightarrow{\mathcal{L}} \text{Weibull } (1, \rho(x))$$
$$b_n = \varphi(x) - \varphi_{1-\frac{1}{nF_X(x)}}(x) \text{ spacing}$$

- A more explicit result:

If $F(y|x) \in \mathcal{C}_\rho$ where$\rho(x) > p$

$\varphi(.)$ differentiable and strictly monotone increasing

Then:

i) $f(x,y) = \ell_x\{\varphi(x) - y\}^{\beta_x} + 0((\varphi(x) - y)^{\beta_x})$

ii) $\beta_x = \rho(x) - (p+1) > -1$

$$(n\ell x)^{\frac{1}{\beta_x + p + 1}}(\varphi(x) - \hat{\varphi}(x)) \stackrel{\mathcal{L}}{\Longrightarrow} \text{ Weibull } (1, \beta_x + p + 1)$$

Conditioning increases the smoothness at the boundary.

– **Behavior of regularized conditional extreme values estimation** with **fixed** regularization parameters (α and m).

Two stochastic process convergence (w.r.t. x):

$$\left.\begin{array}{rcl} \sqrt{n}(\hat{\varphi}_\alpha(.) - \varphi_\alpha(.)) & \Rightarrow & \mathcal{N}(0,\Gamma) \\ \sqrt{n}(\hat{\varphi}_m(.) - \varphi_m(.)) & \Rightarrow & \mathcal{N}(0,\Omega) \end{array}\right\} \text{ gaussian processes}$$

m estimation: Cazals, Florens and Simar (2002)

α estimation: Daouia, Florens and Simar (2008)

– **Behavior of regularized conditional extreme values estimation** with **fast varying** regularization parameter.

If $\alpha \to 1$ or $m \to \infty$ "very fast":

$\hat{\varphi}_\alpha(x)$ and $\hat{\varphi}_m(x)$ have the same behavior as $\hat{\varphi}(x)$ (see Aragon, Daouia, Thomas (2005) and Cazals, Florens and Simar (2002)

– Main interesting results: **behavior of the regularized conditional extreme values** estimation with **slow** varying regularization parameter.

- α quantile

  If $F(y|x) \in \mathcal{C}_{\rho(x)}$ and if $k_n(x)$ satisfies

  – $k_n(x) \to \infty$

  – $\frac{n}{k_n(x)} \to \infty$

Then, $\forall x$:

$$k_n^{\frac{1}{2}} \left(\frac{n}{k_n(x)} \right)^{\frac{1}{\rho(x)}} \left[\hat{\varphi}_{1-\frac{k_n-1}{n\hat{F}_X(x)}}(x) - \varphi(x) + \left(\frac{k_n(x)}{n\ell(x)} \right)^{\frac{1}{\rho(x)}} \right]$$

$$\stackrel{\mathcal{L}}{\Rightarrow} N(0, \tau^2(x))$$

Sequence $\alpha_n(x) = 1 - \frac{k_n-1}{n\hat{F}_X(x)}$

$\alpha_n(x) \to 1$ such that $n(1-\alpha_n) \to \infty$

Bias corrected α quantile estimation:

$$\tilde{\varphi}_{\alpha_n}(x) = \hat{\varphi}_{1-\frac{k_n-1}{n\hat{F}_X(x)}}(x) + \frac{1}{2^{\frac{1}{\rho(x)}-1}}\left(\hat{\varphi}_{1-\frac{k_n(x)-1}{n\hat{F}_X(x)}} - \hat{\varphi}_{1-\frac{2k_n(x)-1}{n\hat{F}_n(x)}}(x)\right)$$

Asymptotically normal without bias

Very powerfull for computation of confidence intervals and tests.

- *m*-estimation

  If $F(.|x) \in \mathcal{C}_\rho$ and if $m_n(x)$ satisfies

  i) $m_n(x) \to \infty$

  ii) $\frac{n}{m_n^2(x)} \to \infty$ then, $\forall x$,

$$\frac{\sqrt{n}}{[m_n(x)]^{\frac{\rho(x)}{\rho(x)+1}}} \left(\hat{\varphi}_{m_n(x)}(x) - \varphi(x) + a(x) \left[\frac{1}{l_n(x)} \right]^{\frac{1}{m_n(x)+1}} \right)$$

$$\stackrel{\mathcal{L}}{\Rightarrow} N(0, \sigma^2(x))$$

  $a(x)$ and $\sigma^2(x)$ depend on $\ell(x), \rho(x)$ and $F_X(x)$.

– **Two questions:**
  – optimal choice of $k_n(x)$ and $m_x(x)$ if $\ell(x)$ and $\rho(x)$ are given.
  – estimation of $\ell(x)$ and $\rho(x)$.

V - A decision theoretic framework

for the choice of the regularization parameter. Rates of convergence

- The choice of α or m is view as a decision based on the minimisation of a risk function
- Usual choice of risk function in non parametric statistic: the mean square error which leads to balance the bias square and the variance. Not adapted to extreme value estimation.
- Proposed risk function: biais + robustness
- Bias $\varphi(x) - \varphi_\alpha(x)$ α quantile estimation $\varphi(x) - \varphi_m(x)$ m estimation
- Robustness $\varphi_\alpha(x) = T_\alpha(F(.|x)) = [F(.|x)]^{-1}(1-\alpha)$

$\hat{F}(.|x)$ the empirical conditional c.d.f

$\hat{G}_\delta(.|x)$ empirical conditional c.d.f where the maximum of the y_i such that $x_i \leq x$ is perturbated by the addition of δ.

$T'_{\alpha,F(.|x)}()$ Frechet derivative of T_α at F.

Robutness:

$$\left\{ \frac{\partial}{\partial \delta} \left[T'_{\alpha,F(.|x)}(\hat{G}_\delta(.|x) - \hat{F}(.|x)) \right] \right\}_{\delta=0}$$

measurement of the sensitivity of the estimation to an infinitisimal perturbation of the maximum.

Identical for the m-estimation.

$$\varphi_m(x) = R_m(F(.|x)) = \int_0^\infty [1 - F(y|x)^m] dy$$

$$\frac{\partial}{\partial \delta} E\left[R'_{m,F(.|x)}(\hat{G}_\delta(.|x) - \hat{F}(.|x)) \right]_{|\delta=0}$$

– The optimal α or m are theoretic (depending on the unknown $F(.|x)$) but may be estimated (plug in method).

– **Main questions:** what are the rate of convergence of α and m selected by the risk minimization? Do this rate satisfied the asymptotic normality condition? What the deduced rates of convergence?

  – Solution for the α quantile:

$$k_n(x) \sim \ln[nF_X(x)] \sim \ln(n_x)$$

n_x number of observations such that $x_i \leq x$.

Rate: $\hat{\varphi}_{\alpha_n}(x)$ converges to $\varphi(x)$ at the rate $n^{\frac{1}{\rho_x}}(\ln n)^{\frac{1}{2}-\frac{1}{\rho_x}}$

- Solution for the m estimation:

$$m_n(x) \sim C(x) n^{\frac{\rho(x)+1}{\rho(x)+2}}$$

Rate $n^{\frac{2}{\rho(x)+2}}$

Regularized estimation have a better rate as the conditional maximum if $\rho(x) > 2$. (rata of the maximum: $n^{\frac{1}{\rho(x)}}$)

Recall $\rho(x) =$ $\underbrace{\beta(x)}_{\text{regularity of the joint distribution}} + \underbrace{p}_{\text{dimension } of X} + 1$

Estimation of the Regularity index ρ_x

- **Pickands estimate**
  - Pickands type estimator (needs **large sample sizes** n):

$$\hat{\rho}_x = \log 2 \left(\log \frac{\hat{\varphi}_{1-\frac{2k-1}{n\hat{F}_X(x)}}(x) - \hat{\varphi}_{1-\frac{4k-1}{n\hat{F}_X(x)}}(x)}{\hat{\varphi}_{1-\frac{k-1}{n\hat{F}_X(x)}}(x) - \hat{\varphi}_{1-\frac{2k-1}{n\hat{F}_X(x)}}(x)} \right)^{-1}.$$

  - Property: Under regularity conditions and if $F(\cdot \mid x) \in \mathcal{C}$,

$$\sqrt{k_n}(\hat{\rho}_x - \rho_x) \xrightarrow{\mathcal{L}} N(0, \sigma^2(\rho_x)),$$

    where $\sigma^2(\rho_x)$ is given and $k_n \to \infty$ at an appropriate rate (see DFS (2008b) for details.

- **Bias corrected estimator of the frontier**

As above by **plugging** $\hat{\rho}_x$ in the formula: the expression of the asymptotic variance has to be changed (see DFS, 2008b, for details).

Examples

- **Uniform**

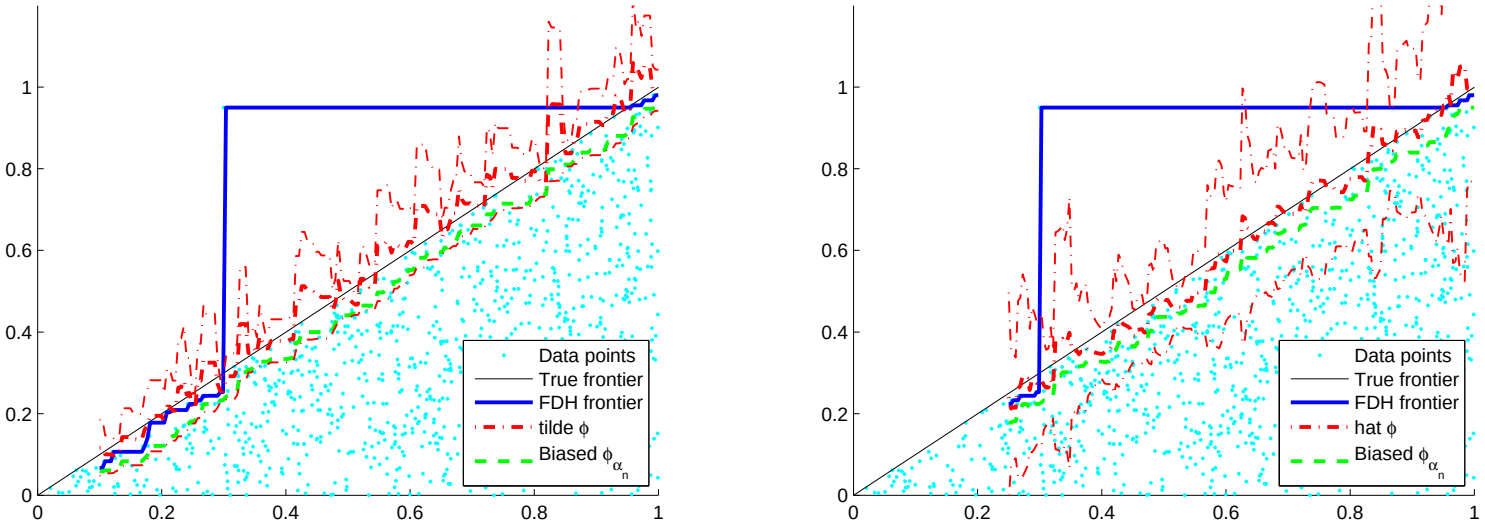


Figure 1: *Uniform case with one outlier,* $n = 1000$*: left panel* $\rho_x = 2$ *is known and right panel,* $\hat{\rho}_x$ *is used.*

Examples

- **Cobb Douglas**

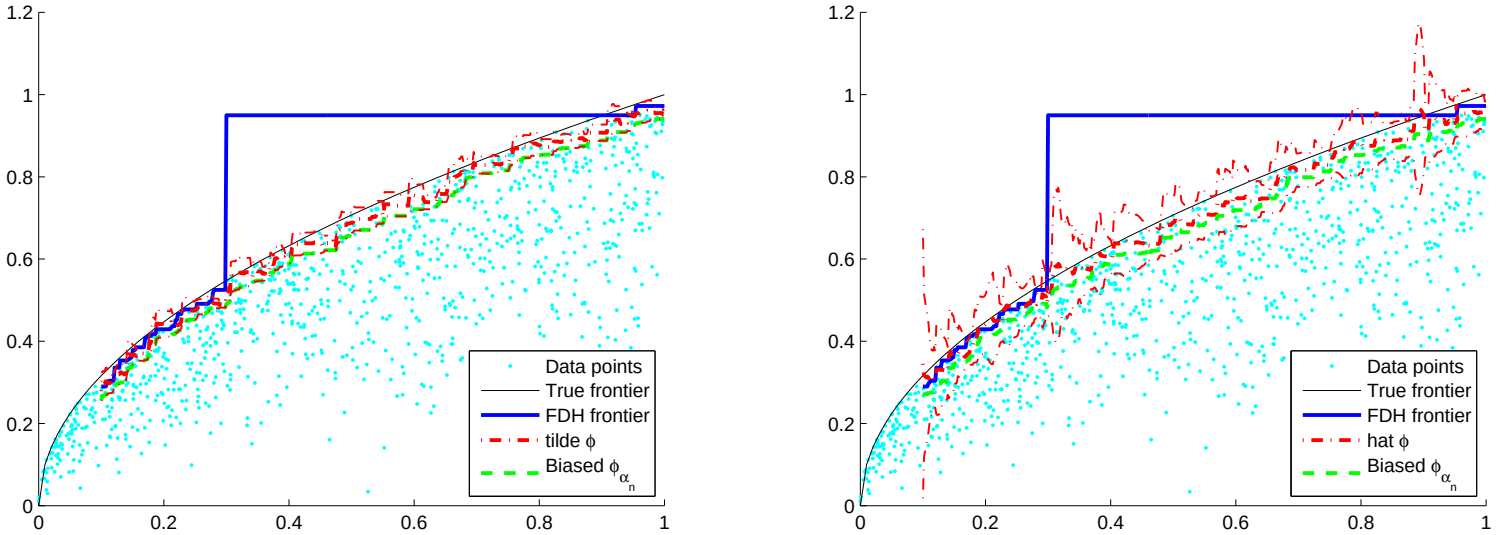


Figure 2: *Cobb Douglas case with one outlier,* $n = 1000$*: left panel* $\rho_x = 2$ *is known and right panel,* $\hat{\rho}_x$ *is used.*

Examples

- **Uniform and Cobb Douglas, ρ_x unknwon and $n = 5000$**

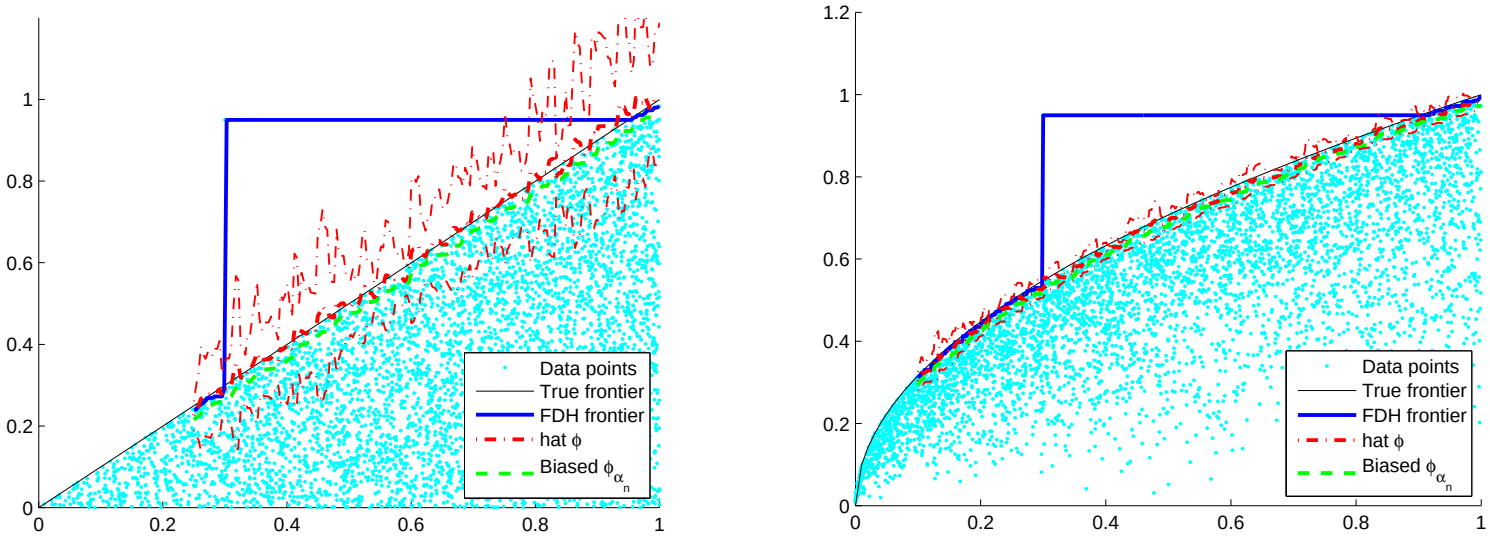


Figure 3: *ρ_x unknown, $n = 5000$: left panel uniform case is known and right panel Cobb Douglas case.*

Examples

- **French Post Data:** $\rho_x = 2$ **known**

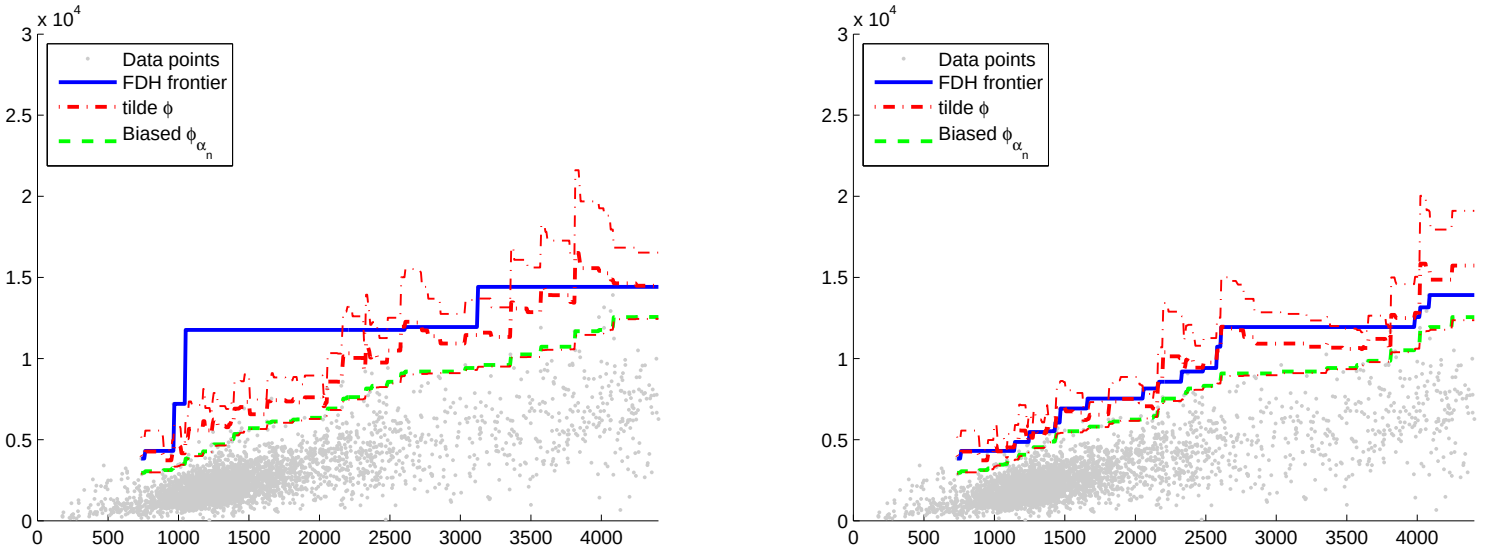


Figure 4: *Post data:* $\rho_x = 2$, *left panel all the data, right panel 2 potential outliers deleted.*

Examples

- **French Post Data: ρ_x unknown**

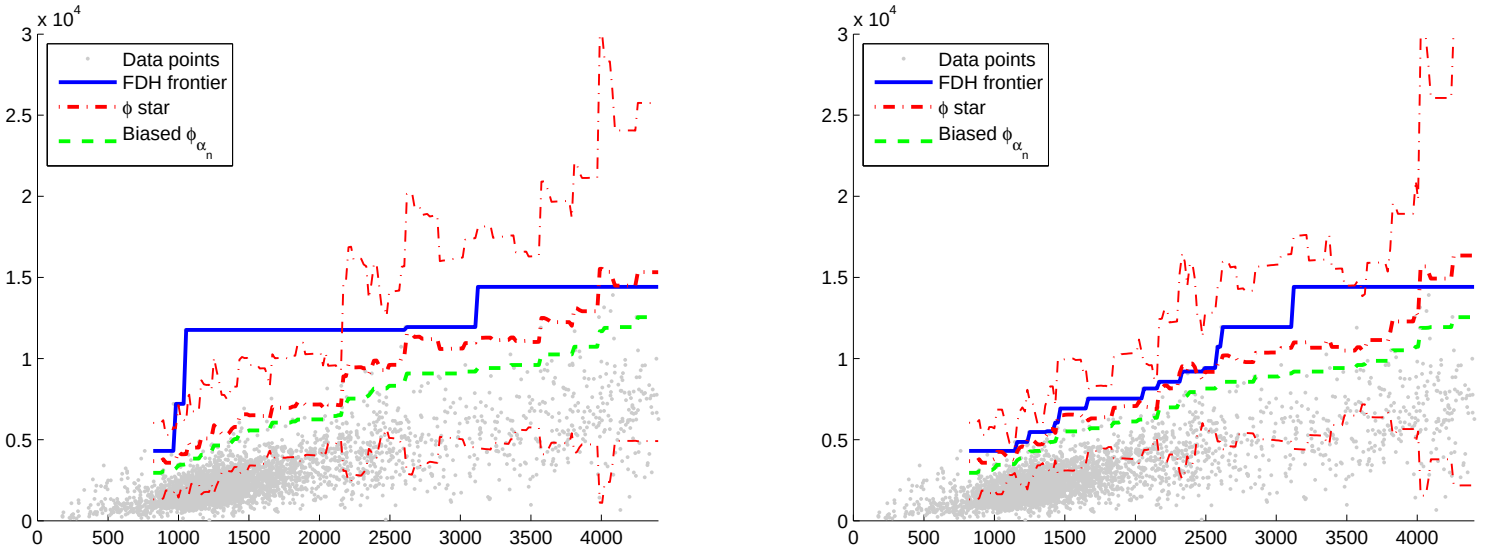


Figure 5: *Post data: ρ_x unknown, left panel all the data, right panel 2 potential outliers deleted.*

Conclusions

- Theory of robust estimation of conditional extreme value.
- Many applications in production efficiency theory
  - Postal delivery or sorting offices
  - Hospitals
  - University ranking
- Main theoretical questions
  - Improvement of the regularity index estimation use of the regularity of $\varphi(x)$ and $\rho(x)$ as functions of x

Main References

- Aragon, Y., A. Daouia and C. Thomas-Agnan (2005), Nonparametric Frontier Estimation: A Conditional Quantile-based Approach, *Econometric Theory*, 21, 358–389.
- Cazals, C. J.P. Florens and L. Simar (2002), Non Parametric Frontier Estimation: A Robust Approachs, *Journal of Econometrics*, 106, 1–25.
- Daouia, A., J.P. Florens and L. Simar (2008a), Functional Convergence of Quantile-type Frontiers with Application to Parametric Approximations, *Journal of Statistical Planning and Inference*, 138, 708–725.
- Daouia, A., J.P. Florens and L. Simar (2008b), Frontier estimation and Extreme value theory, Discussion paper 0806, Institut de Statistique, UCL.
- Daouia, A., J.P. Florens and L. Simar (2008c), Regularization of boundary estimators with application to frontier models, in progress.

- 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.
- 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.
- Hwang, J.H., B.U. Ponk and W. Ryu (2002), Limit theorems for boundary functions estimators, *Statistics et Probability Letters*, 59, 353-360.
- Park, B. L. Simar and Ch. Weiner (2000), The FDH Estimator for Productivity Efficiency Scores : Asymptotic Properties, *Econometric Theory*, Vol 16, 855–877.
- van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics, 3, Cambridge University Press, Cambridge.