Robustness and inference in nonparametric partial-frontier modeling

Abdelaati Daouia[1] and Irène Gijbels[2]

[1] Toulouse School of Economics (GREMAQ), University of Toulouse, France (daouia@cict.fr)

[2] Department of Mathematics and Leuven Statistics Research Center (LStat), Katholieke Universiteit Leuven, Belgium (Irene.Gijbels@wis.kuleuven.be)

Abstract

A major aim in recent nonparametric frontier modeling is to estimate a partial frontier well inside the sample of production units but near the optimal boundary. Two concepts of partial boundaries of the production set have been proposed: an expected maximum output frontier of order $m = 1, 2, \ldots$ and a conditional quantile-type frontier of order $\alpha \in]0, 1]$. In this paper, we answer the important question of how the two families are linked. For each m, we specify the order α for which both partial production frontiers can be compared. We show that even one perturbation in data is sufficient for breakdown of the nonparametric order-m frontiers, whereas the global robustness of the order-α frontiers attains a higher breakdown value. Nevertheless, once the α-frontiers break down, they become less resistant to outliers than the order-m frontiers. Moreover, the m-frontiers have the advantage to be statistically more efficient. Based on these findings, we suggest a methodology for identifying outlying data points. We establish some asymptotic results, contributing to important gaps in the literature. The theoretical findings are illustrated via simulations and real data.

JEL classification: C13; C14; D20

Key words : asymptotics, breakdown values, econometric frontiers, outliers detection.

1 Introduction

In the economics, statistics, management science and related literatures a major aim is to estimate the upper boundary of a sample $\{(X_i, Y_i), i = 1, \ldots, n\}$ of independent copies of a random production unit (X, Y) with support defined by $\{(x, y) \in \mathbb{R}_+^{p+1} | 0 \leq y \leq \varphi(x)\}$. Econometric considerations lead to the natural assumption that the frontier function φ is monotone nondecreasing. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be the probability space on which the vector of inputs $X \in \mathbb{R}_+^p$ and the single output Y are defined. Then following Cazals et al (2002), the optimal value $\varphi(x)$ can be characterized as the right-endpoint of the conditional distribution function $F(y|x) = \mathbb{P}(Y \leq y | X \leq x) = F(x, y)/F_X(x)$, with $F(\cdot, \cdot)$ and $F_X(\cdot)$ being respectively the joint and marginal distribution functions of (X, Y) and X.

The conventional estimate for φ is the Free Disposal Hull (FDH) estimator, *i.e.* the lowest nondecreasing step surface covering all sample points, that is $\hat{\varphi}_n(x) := \sup\{y \geq 0 | \hat{F}_n(y|x) < 1\} = \max_{i|X_i \leq x} Y_i$, where $\hat{F}_n(y|x) = \hat{F}(x, y)/\hat{F}_{X,n}(x)$, with $\hat{F}(x, y) = (1/n) \sum_{i=1}^n \mathbb{1}(X_i \leq x, Y_i \leq y)$ and $\hat{F}_{X,n}(x) = \hat{F}(x, \infty)$. When the frontier function φ is also assumed to be concave, a popular estimator is the Data Envelopment Analysis (DEA) estimator, which is the lowest concave surface covering the FDH frontier. Both FDH and DEA estimators

derive from the pioneering work of Farrell (1957). The DEA frontier has been popularized by Charnes at al (1978), while the FDH has been proposed by Deprins et al (1984). See Simar and Wilson (2008) for a survey on inference techniques using FDH and DEA estimators.

By construction, these envelopment estimators are very sensitive to extremes and/or outliers in the output-direction. This dramatic lack of robustness results in poor estimation of the corresponding economic efficiencies; the efficiency score of a firm is estimated via the distance between the attained produced output and the optimal production level given by the frontier function. Of course, in production activity, outlying outputs Y_i are highly desirable. But in absence of information on whether the observations are measured accurately, it is prudent to seek frontier estimators which are not determined by very few extreme observations. The underlying idea of the two existing methods in the econometric literature is to estimate a partial frontier well inside the cloud of data points but near the upper boundary.

The first concept of a partial boundary of the joint support of (X, Y) has been introduced by Cazals et al (2002). Given an integer $m \geq 1$, they define a notion of expected maximum output function of order m as $\xi_m(x) := \int_0^{\varphi(x)} (1-[F(y|x)]^m)dy$. This partial frontier function converges to the full frontier function $\varphi(x)$ as $m \to \infty$. It is estimated by $\hat{\xi}_{m,n}(x) := \int_0^{\hat{\varphi}_n(x)} (1-[\hat{F}_n(y|x)]^m)dy$. To summarize the properties of this estimator, for fixed sample size n we have $\lim_{m\to\infty} \uparrow \hat{\xi}_{m,n}(x) = \hat{\varphi}_n(x)$, and for fixed order m we have $\sqrt{n}(\hat{\xi}_{m,n}(x) - \xi_m(x)) \to N(0, \sigma^2(x,m))$ as $n \to \infty$, where the asymptotic variance $\sigma^2(x,m)$ is given in (4.2). The second concept of a partial frontier function, suggested by Aragon et al (2005), is defined as the αth quantile function $q_\alpha(x) := \inf\{y \geq 0 | F(y|x) \geq \alpha\}$, with $\alpha \in]0,1]$. This order-α frontier function converges to $\varphi(x)$, as $\alpha \uparrow 1$, and is estimated by $\hat{q}_{\alpha,n}(x) := \inf\{y \geq 0 | \hat{F}_n(y|x) \geq \alpha\}$. For n fixed, this estimator satisfies $\lim_{\alpha\to 1} \uparrow \hat{q}_{\alpha,n}(x) = \hat{\varphi}_n(x)$, and for fixed order α we have $\sqrt{n}(\hat{q}_{\alpha,n}(x) - q_\alpha(x)) \to N(0, \sigma^2(\alpha, x))$ as $n \to \infty$, where $\sigma^2(\alpha, x)$ is given in (4.3), provided that $F(\cdot|x)$ is differentiable at $q_\alpha(x)$ with derivative $f(q_\alpha(x)|x) > 0$.

Unlike usual (FDH, DEA) methods, both alternatives $\{\hat{q}_{\alpha,n}(x)\}$ and $\{\hat{\xi}_{m,n}(x)\}$ are qualitatively robust and bias-robust as shown in Daouia and Ruiz-Gazen (2006). But the order-α quantile frontiers can be more robust to extremes than the order-m frontiers when estimating the true full frontier φ (*i.e.* when $\alpha \uparrow 1$ and $m \uparrow \infty$) since the influence function is no longer bounded for order-m frontiers as m tends to infinity, while it remains bounded for the conditional quantile frontiers as the order α tends to one. This advantage is proved only under the condition that the conditional density function $f(\cdot|x)$ is not null and continuous on its support. No attention was devoted however to the difference between the reliability of the two sequences of estimators $\{\hat{q}_{\alpha,n}(x)\}$ and $\{\hat{\xi}_{m,n}(x)\}$ in the general setting. Moreover, the influence function only offers a local quantification of robustness by measuring the sensitivity of estimators to infinitesimal perturbations, but it is well known that estimators can be

infinitesimally robust and yet still highly sensitive to small, finite perturbations. To measure the global robustness of an estimator, the richest quantitative information is provided by the finite sample breakdown point as shown by Donoho and Huber (1983). It measures the smallest fraction of contamination of an initial sample that can cause an estimator to take values arbitrarily far from its value at the initial sample.

In this paper, we deal with global robustness and some asymptotic aspects of the two sequences $\{\hat{q}_{\alpha,n}\}$ and $\{\hat{\xi}_{m,n}\}$ as estimators of the partial frontiers q_α and ξ_m respectively, for fixed orders α and m, and as estimators of the full frontier φ itself when $\alpha \to 1$ and $m \to \infty$.

In Section 2 we first focus on the replacement breakdown values of the estimators. We show that, as expected, both FDH and DEA frontiers may break down for any contamination (Lemma A.1). The surprising result is that even one outlying observation is sufficient for breakdown of the partial frontier $\hat{\xi}_{m,n}(x)$ for any order m, whereas the partial order-α frontier $\hat{q}_{\alpha,n}(x)$ has the desirable robustness in withstanding the contamination of outlying observations. While the asymptotic breakdown value is 0 for any order-m partial frontier, it is $(1-\alpha)F_X(x)$ for the sequence $\{\hat{q}_{\alpha,n}(x)\}_{n\geq 1}$. But, once the α-frontiers break down, they become less resistant to outliers than the order-m frontiers. A natural question arising is: how to compare the reliability between the two sequences of partial frontiers once the order-α frontier also breaks down? A more general question is: given a fixed order m, which frontier function $\hat{q}_{\alpha,n}(x)$ can be analyzed and compared with $\hat{\xi}_{m,n}(x)$?

The families $\{\xi_m(\cdot), m \geq 1\}$ and $\{q_\alpha(\cdot), \alpha \in]0,1]\}$ have emerged in the econometric literature as two different theoretical concepts of partial production frontiers. See *e.g.* Daraio and Simar (2007) for statistical properties of both concepts of partial boundaries together with several appealing economic features. The estimators $\hat{\xi}_{m,n}$ and $\hat{q}_{\alpha,n}$ (of ξ_m and q_α respectively) cannot be compared since they do not estimate the same quantity except for the limiting case, when $m \uparrow \infty$ and $\alpha \uparrow 1$ (when both estimate the true full frontier φ). We however establish in Proposition 2.2 that the two concepts of partial boundaries are closely linked in the sense that for each order $m \geq 1$, there exists a well-specified order $\alpha = \alpha(m) = (1/2)^{1/m}$ such that the theoretical order-m and order-α frontiers are respectively the mean and median of the same distribution, namely that of the maximum of m independent random variables drawn from the law of Y given X not exceeding some level of inputs. This result also confirms the advantage of $\hat{q}_{\alpha,n}$ over $\hat{\xi}_{m,n}$ in terms of finite sample breakdown point and gross-error sensitivity, but such a robust proposal may sacrifice statistical efficiency.

We show in Section 3 how these results can be exploited to detect outlying data points in the output-direction. In the frontier modeling context, descriptive methods for identifying outliers have been proposed by Wilson (1993,1995). Although very useful, these methods are very computer intensive as the sample size increases and are based on some tuning

parameters whose choice is not justified. See further discussions in Sections 3 and 6.

Section 4 contributes to important gaps in the asymptotic theory for the estimators $\hat{\xi}_{m,n}(x)$ and $\hat{q}_{\alpha,n}(x)$. We establish pointwise and functional asymptotic representations for $\sqrt{n}(\hat{\xi}_{m,n}(x) - \xi_m(x))$ and improve its order of convergence to $O(\sqrt{\log\log n})$. Similar asymptotic properties for $\sqrt{n}(\hat{q}_{\alpha,n}(x) - q_\alpha(x))$ can be found in Daouia (2005) and Daouia et al (2008). However, unlike $\xi_m(x)$, the computation of the asymptotic confidence interval of $q_\alpha(x)$ requires estimation of the quantile density function $f(q_\alpha(x)|x)$, often resulting in estimates of unsatisfactory accuracy for finite samples. To avoid this problem, we derive an alternative asymptotic confidence interval for $q_\alpha(x)$ not requiring knowledge of $f(q_\alpha(x)|x)$. Finally, we show under general conditions that the asymptotic normality of both $\hat{\xi}_{m,n}(x)$ and $\hat{q}_{\alpha,n}(x)$ is still valid when $m = m_n \to \infty$ and $\alpha = \alpha_n \to 1$ as $n \to \infty$. Section 5 illustrates the theoretical findings through simulations and real data and Section 6 concludes.

2 Robustness

The most successful notion of global robustness of an estimator T at a sample $(Z)^n = (Z_1, \ldots, Z_n)$ is provided by the finite sample breakdown point of Donoho and Huber (1983):

$$RB(T,(Z)^n) = \min\left\{\frac{k}{n} : k = 1,\ldots,n, \text{ and satisfies } \sup_{(Z)_k^n} |T\{(Z)_k^n\} - T\{(Z)^n\}| = \infty\right\},$$

where $(Z)_k^n$ denotes the contaminated sample by replacing k points of $(Z)^n$ with arbitrary values. Given $m \geq 1$, $\alpha \in]0,1]$ and $x \in \mathbb{R}_+^p$, the partial boundaries $\hat{\xi}_{m,n}(x)$ and $\hat{q}_{\alpha,n}(x)$ are representable as functionals of the joint empirical distribution function $\hat{F}(\cdot,\cdot)$, or equivalently, of the data set $(X,Y)^n = \{(X_i,Y_i), i = 1,\ldots,n\}$:

$$\left\{\begin{array}{l} \xi_m(x) = S^{m,x}(F) \\ \hat{\xi}_{m,n}(x) = S^{m,x}(\hat{F}) = S^{m,x}((X,Y)^n) \end{array}\right. , \quad \left\{\begin{array}{l} q_\alpha(x) = T^{\alpha,x}(F) \\ \hat{q}_{\alpha,n}(x) = T^{\alpha,x}(\hat{F}) = T^{\alpha,x}((X,Y)^n) \end{array}\right.$$

where the operators $S^{m,x}$ and $T^{\alpha,x}$ associate to a distribution function $G(\cdot,\cdot)$ on $\mathbb{R}_+^p \times \mathbb{R}_+$ such that $G(x,\infty) > 0$, the real values

$$S^{m,x}(G) = \int_0^\infty \left(1 - \left[\frac{G(x,y)}{G(x,\infty)}\right]^m\right) dy \quad \text{and} \quad T^{\alpha,x}(G) = \inf\left\{y \geq 0 | \frac{G(x,y)}{G(x,\infty)} \geq \alpha\right\},$$

with the integrand being identically zero for $y \geq \inf\{y \mid G(x,y)/G(x,\infty) = 1\}$.

As expected we can easily show that even one outlying observation is sufficient for breakdown of the FDH frontier (see Lemma A.1, Appendix), and consequently for breakdown of the DEA frontier. But the surprising result is that the partial order-m boundary breaks down for the same fraction, $1/n$, of contamination as the envelopment FDH and DEA estimators, for any order m.

Theorem 2.1. *Let $x \in \mathbb{R}^p_+$ such that $\hat{F}_{X,n}(x) > 0$. Then, for any order $m \geq 1$,*

$$RB(\hat{\xi}_{m,n}(x), (X,Y)^n) = 1/n.$$

Hence the asymptotic breakdown value is 0 for any order-m partial frontier. In contrast, by an appropriate choice of the order α as a function of n and $\hat{F}_{X,n}(x)$, we can derive a partial quantile-based frontier $\hat{q}_{\alpha,n}(x)$ capable of withstanding arbitrary perturbations of a significant proportion of the data points without disastrous results.

Theorem 2.2. *Let $x \in \mathbb{R}^p_+$ such that $\hat{F}_{X,n}(x) > 0$. Then, for any order $\alpha \in]0,1]$,*

$$RB(\hat{q}_{\alpha,n}(x), (X,Y)^n) = \begin{cases} \left(n(1-\alpha)\hat{F}_{X,n}(x) + 1\right)/n & \text{if} \quad \alpha n \hat{F}_{X,n}(x) = 1, 2, 3, \ldots \\ \left(n\hat{F}_{X,n}(x) - \left[\alpha n \hat{F}_{X,n}(x)\right]\right)/n & \text{otherwise,} \end{cases}$$

where $[\alpha n \hat{F}_{X,n}(x)]$ denotes the integer part of $\alpha n \hat{F}_{X,n}(x)$.

Remark 2.1. The asymptotic breakdown value for the sequence $\{\hat{q}_{\alpha,n}(x)\}_{n\geq 1}$ is then $(1-\alpha)F_X(x)$. When the order α is fixed, this theorem reflects how the corresponding partial frontier $\hat{q}_{\alpha,n}(x)$ suffers from the left-border effect when the vector $x \in \mathbb{R}^p_+$ of inputs-usage is too small. Likewise, increasing the dimension p of input factors x decreases $\hat{F}_{X,n}(x)$, and hence $RB(\hat{q}_{\alpha,n}(x), (X,Y)^n)$ goes down. On the other hand, once we know that $\hat{q}_{\alpha,n}(x) = T^{\alpha,x}((X,Y)^n)$ does not break down for the fraction $(k^*-1)/n$ of contamination, with $k^*/n = RB(\hat{q}_{\alpha,n}(x), (X,Y)^n)$, it is of interest to know how large the bias $|T^{\alpha,x}((X,Y)^n_{k^*-1}) - T^{\alpha,x}((X,Y)^n)|$ can be. For this purpose we compute the upper bound of this bias, here we only focus on contaminated samples $(X,Y)^n_{k^*-1} := (X,Y)^{n,y}_{k^*-1}$ in the direction of Y obtained by replacing k^*-1 points (X_i, Y_i) with outlying extreme-values (X_i, Y_i^*).

Proposition 2.1. *Let $x \in \mathbb{R}^p_+$ such that $\hat{F}_{X,n}(x) > 0$. Then, for any order $\alpha \in]0,1]$,*

$$0 \leq T^{\alpha,x}((X,Y)^{n,y}_{k^*-1}) - T^{\alpha,x}((X,Y)^n) \leq \hat{\varphi}_n(x) - \hat{q}_{\alpha,n}(x)$$

for any contaminated sample $(X,Y)^{n,y}_{k^*-1}$.

Note that when the order α goes to 1, *i.e.* when estimating the full frontier function $\varphi(x)$ itself, the maximal bias tends to zero since $\lim_{\alpha\uparrow 1} \hat{q}_{\alpha,n}(x) = \hat{\varphi}_n(x)$. However, when α is fixed, the maximal bias $\hat{\varphi}_n(x) - \hat{q}_{\alpha,n}(x)$ may become too large as x increases. So, to estimate $\varphi(x)$ by $\hat{q}_{\alpha,n}(x)$, the order α should be chosen appropriately as a function of both x and n.

We next answer the important question of how the two families of order-α and order-m boundaries are linked. We show that these concepts of partial frontiers are closely linked in the sense that $\{q_\alpha(x), \alpha \in]0,1]\}$ defines a "robustified" variant of the family $\{\xi_m(x), m \geq 1\}$

while the latter defines an "efficient" variant of the former. We provide an explicit and exact expression of α as a function of m that allows to select which frontier $\hat{q}_{\alpha,n}$ can be analyzed and compared with $\hat{\xi}_{m,n}$. Indeed, it is easy to see that $\xi_m(x) = \mathbb{E}[\max(Y_{x1}, \ldots, Y_{xm})]$ for any sequence $(Y_{x1}, \ldots, Y_{xm})$ of m independent random variables drawn from the conditional distribution of Y given $X \leq x$. Since the median is known to be more robust than the mean (see *e.g.* Hampel 1968), we can "robustify" the expected value of the maximum, $\xi_m(x)$, by simply replacing the expectation with the median to obtain a median-type expected maximum output frontier.

Proposition 2.2. *Consider the robust-variant of* $\xi_m(x)$ *defined as*

$$\tilde{\xi}_m(x) = \textit{Median}\ [\max(Y_{x1}, \ldots, Y_{xm})].$$

Then for any order $m \geq 1$*, there exists an order* $\alpha(m) = (1/2)^{1/m}$ *such that* $\tilde{\xi}_m(x) = q_{\alpha(m)}(x)$.

Hence for each expected-maximum output m-frontier, there exists a quantile-type frontier of a well-specified order $\alpha = \alpha(m)$ such that their pointwise values $\xi_m(x)$ and $q_\alpha(x)$ are respectively the theoretical mean and median of the same distribution, namely that of the random variable $\max(Y_{x1}, \ldots, Y_{xm})$. When this distribution is symmetric, $\hat{\xi}_{m,n}(x)$ and $\hat{q}_{\alpha(m),n}(x)$ estimate exactly the same quantity.

Remark 2.2. It is difficult to imagine the family $\{\hat{q}_{\alpha,n}(x), \alpha \in]0,1]\}$ being preferred in all contexts: of course $\hat{q}_{\alpha,n}(x)$ is preferred over $\hat{\xi}_{m,n}(x)$ in terms of finite sample breakdown point and gross-error sensitivity, but such a robust proposal may sacrifice in terms of statistical efficiency (measured *e.g.* by means of estimation variance). Moreover, once $\{\hat{q}_{\alpha,n}(x)\}$ breaks down, it becomes less resistant to extreme values than $\{\hat{\xi}_{m,n}(x)\}$. Indeed, putting $N_x = n\hat{F}_{X,n}(x)$ and taking $Y_1^x, \ldots, Y_{N_x}^x$ to be the Y_i's such that $X_i \leq x$, we first have

$$\hat{q}_{\alpha,n}(x) = \begin{cases} Y^x_{(\alpha N_x)} & \text{if} \quad \alpha N_x = 1, 2, 3, \ldots \\ Y^x_{([\alpha N_x]+1)} & \text{otherwise,} \end{cases} \tag{2.1}$$

where $Y^x_{(i)}$ denotes the ith order statistic of the points $Y_1^x, \ldots, Y_{N_x}^x$. Likewise, we have

$$\hat{\varphi}_n(x) - \hat{\xi}_{m,n}(x) = \sum_{i=1}^{N_x - 1} (i/N_x)^m \{Y^x_{(i+1)} - Y^x_{(i)}\}. \tag{2.2}$$

This difference being a sum of weighted spacings, $\hat{\xi}_{m,n}(x)$ is more resistant to FDH points in the sense that it converges slowly to $\hat{\varphi}_n(x)$ as m increases, whereas $\hat{q}_{\alpha(m),n}(x)$, as an order statistic, converges rapidly to $\hat{\varphi}_n(x)$ once it breaks down. It is easy to see that

$$\hat{q}_{\alpha(m),n}(x) = \sum_{i=1}^{N_x} Y^x_{(i)} \mathbb{1}\left\{ \left(\frac{i-1}{N_x} \right)^m < \frac{1}{2} \leq \left(\frac{i}{N_x} \right)^m \right\}, \tag{2.3}$$

and so $\hat{q}_{\alpha(m),n}(x)$ coincides with $\hat{\varphi}_n(x)$ for all $m > \log(2)/\log(N_x/(N_x-1))$, which is not the case for

$$\hat{\xi}_{m,n}(x) = \sum_{i=1}^{N_x} Y_{(i)}^x \left\{ \left(\frac{i}{N_x} \right)^m - \left(\frac{i-1}{N_x} \right)^m \right\}. \tag{2.4}$$

These sensitivity and resistance characteristics of $\hat{\xi}_{m,n}(x)$ and $\hat{q}_{\alpha(m),n}(x)$, as well as their statistical efficiency, are illustrated in Subsection 5.1 with simulated and real data sets. To conclude, the $\alpha(m)$-frontier $\tilde{\xi}_{m,n}$ is sometimes preferred over the m-frontier $\hat{\xi}_{m,n}$ and sometimes not according to the values of m. So a sensible practice is not to restrict the frontier analysis to one procedure, but to check whether both concepts of partial boundaries point toward similar conclusions. See the practical guidelines in Subsection 5.4.

3 Detection of anomalous data

The word "anomalous" is used here for detecting isolated data points in the direction of Y. From now on, we write $\tilde{\xi}_{m,n} := \hat{q}_{\alpha(m),n}$.

Local distance : Let (x_a, y_a) be an isolated outlier, that is, $(x_a, y_a) = (x_a, \hat{\varphi}_n(x_a))$ is an FDH observation clearly outlying the cloud of data points. We know that both partial boundaries $\hat{\xi}_{m,n}(x_a)$ and $\tilde{\xi}_{m,n}(x_a) \nearrow \hat{\varphi}_n(x_a)$ as $m \to \infty$. We distinguish between two different behaviors of $\hat{\xi}_{m,n}(x_a)$ and $\tilde{\xi}_{m,n}(x_a)$ as the order m increases :

i. While $\hat{\xi}_{m,n}(x_a)$ breaks down (*i.e.* $\hat{\xi}_{m,n}(x_a)$ becomes attracted by the outlying value $y_a = \hat{\varphi}_n(x_a)$) for any order $m \geq 1$ in view of Theorem 2.1, the quantile-type value $\tilde{\xi}_{m,n}(x_a)$, being determined solely by the frequency $\alpha(m)$, remains unaffected even when m increases (quantiles are known to be robust in this sense). In this situation, the distance between the robust value $\tilde{\xi}_{m,n}(x_a)$ and the influencable value $\hat{\xi}_{m,n}(x_a)$ shall increase rapidly as m increases;

ii. However, when m achieves a sufficiently large threshold m_a, the partial boundary $\tilde{\xi}_{m,n}(x_a)$ also breaks down in view of Theorem 2.2 and converges rapidly, as an order statistic (see (2.1) and (2.3)), to the outlying maximum value $\hat{\varphi}_n(x_a)$. Even more strongly, it is easy to see that $\tilde{\xi}_{m,n}$ coincides overall with the FDH frontier $\hat{\varphi}_n$ for any $m \geq \log(1/2)/\log((n-1)/n)$. In contrast, $\hat{\xi}_{m,n}(x_a)$ being a linear combination of order statistics (see (2.2) and (2.4)), converges more slowly to the largest order statistic $\hat{\varphi}_n(x_a)$. Hence, although its sensitivity to the magnitude of the outlying value $\hat{\varphi}_n(x_a)$ for any $m \geq 1$, $\hat{\xi}_{m,n}(x_a)$ becomes more resistant than $\tilde{\xi}_{m,n}(x_a)$ as m exceeds m_a. Thus, the distance between $\tilde{\xi}_{m,n}(x_a)$ and $\hat{\xi}_{m,n}(x_a)$ shall decrease slowly as $m > m_a$ increases.

To summarize, if $\hat{\varphi}_n(x_a)$ is really outlying, the curve $m \mapsto |\tilde{\xi}_{m,n}(x_a) - \hat{\xi}_{m,n}(x_a)|$ shall have roughly a "Λ" structure, that is, a sharp positive slope (indicating that $\hat{\xi}_{m,n}(x_a)$ breaks down while $\tilde{\xi}_{m,n}(x_a)$ remains still unaffected as m increases) followed by a smooth decreasing slope (indicating that $\tilde{\xi}_{m,n}(x_a)$ becomes non-robust for m large enough whereas $\hat{\xi}_{m,n}(x_a)$ is more resistant). Here the "Λ" effect appears at $m_a - 1$ such that the value of $|\tilde{\xi}_{m,n}(x_a) - \hat{\xi}_{m,n}(x_a)|$ at $m = (m_a - 1)$ is sufficiently large compared with its initial value at $m = 1$.

However, if $\hat{\varphi}_n(x_a)$ is only extreme (not really isolated), the graph of $m \mapsto |\tilde{\xi}_{m,n}(x_a) - \hat{\xi}_{m,n}(x_a)|$ will have a slight "$\wedge$" curvature, that is, a non-decreasing slope followed by a non-increasing slope such that the maximal value of the distance $|\tilde{\xi}_{m,n}(x_a) - \hat{\xi}_{m,n}(x_a)|$ is very close to its initial value at $m = 1$.

So, in general, if the graph of the distance function $m \mapsto |\tilde{\xi}_{m,n}(x_i) - \hat{\xi}_{m,n}(x_i)|$ shows clearly a sharp "Λ" curvature for a given observed value x_i, this indicates a potential outlier in the data set. The suspicious outlying point can be then easily recovered: it corresponds to the FDH point (x_k, y_k) for which $y_k = \hat{\varphi}_n(x_i)$. This is the basic idea of our procedure.

Global distance : Consider now the maximal "distance" between the partial boundaries $\tilde{\xi}_{m,n}$ and $\hat{\xi}_{m,n}$, defined as $d(m) = \max_{1 \le i \le n} |\tilde{\xi}_{m,n}(x_i) - \hat{\xi}_{m,n}(x_i)|$. Assume that (x_a, y_a) is the unique outlier in the sample. If this point is far enough from the cloud of data points, then the local distance $|\tilde{\xi}_{m,n}(x_a) - \hat{\xi}_{m,n}(x_a)|$ coincides for all $m \ge 1$ with the global distance $d(m)$. In this case, as described above, the shape of the entire curve $m \mapsto d(m)$ (and not only a part of this graph) should be a sharp "Λ".

If, instead, the sample contains two isolated outliers (x_a, y_a) and (x_b, y_b) with $x_a < x_b$, it is easy to see from Theorem 2.2 that $\tilde{\xi}_{m,n}(x_a)$ breaks down before $\tilde{\xi}_{m,n}(x_b)$. Let m_a and m_b be respectively the values of m at which $\tilde{\xi}_{m,n}(x_a)$ and $\tilde{\xi}_{m,n}(x_b)$ break down. Then $m_a < m_b$. On the other hand, due to the conditioning on $X \le x$, both $\hat{\xi}_{m,n}$ and $\tilde{\xi}_{m,n}$ are more resistant to outliers at x_b than at x_a (left-border effect). It follows that :

i. for $m < m_a$, both $\tilde{\xi}_{m,n}(x_a)$ and $\tilde{\xi}_{m,n}(x_b)$ are unaffected by the two outliers, while $\hat{\xi}_{m,n}(x_a)$ is more attracted by these outliers than $\hat{\xi}_{m,n}(x_b)$ due to the left-border effect. This implies that $|\tilde{\xi}_{m,n}(x_a) - \hat{\xi}_{m,n}(x_a)| \ge |\tilde{\xi}_{m,n}(x_b) - \hat{\xi}_{m,n}(x_b)|$ as m increases, whence $d(m) = |\tilde{\xi}_{m,n}(x_a) - \hat{\xi}_{m,n}(x_a)|$ as $m \uparrow m_a$. Therefore the graph of $d(m)$ should have a sharp positive slope as $m \uparrow m_a$;

ii. once m exceeds m_a, the local distance $|\tilde{\xi}_{m,n}(x_a) - \hat{\xi}_{m,n}(x_a)|$ decreases smoothly to zero (breakdown of $\tilde{\xi}_{m,n}(x_a)$), while $|\tilde{\xi}_{m,n}(x_b) - \hat{\xi}_{m,n}(x_b)|$ still increases rapidly as $m \uparrow m_b$. Let $m_{a,b}$ be the value of m at which $|\tilde{\xi}_{m,n}(x_b) - \hat{\xi}_{m,n}(x_b)|$ exceeds $|\tilde{\xi}_{m,n}(x_a) - \hat{\xi}_{m,n}(x_a)|$. Then $m_a \le m_{a,b} < m_b$. If $m_a = m_{a,b}$, then $d(m) = |\tilde{\xi}_{m,n}(x_b) - \hat{\xi}_{m,n}(x_b)|$ for $m \ge m_a$. Whence $d(m)$ increases rapidly as $m \uparrow m_b$ and decreases smoothly as $m \ge m_b$. In

contrast, if $m_a < m_{a,b} < m_b$, then $d(m) = |\tilde{\xi}_{m,n}(x_a) - \hat{\xi}_{m,n}(x_a)|$ decreases smoothly for $m \in [m_a, m_{a,b})$ whereas $d(m) = |\tilde{\xi}_{m,n}(x_b) - \hat{\xi}_{m,n}(x_b)|$ increases rapidly for $m \in [m_{a,b}, m_b)$ and decreases smoothly for $m \geq m_b$.

In summary, in presence of two outliers far from the cloud of data points, the shape of the entire graph $m \mapsto d(m)$ should be either one sharp "Λ" or two successive "Λ" effects showing an "M" structure. It should be also clear that if (x_a, y_a) is only a suspicious extreme (not really isolated), then the strong "Λ" effect corresponding to the outlier (x_b, y_b) could be preceded by a slight "$\wedge$" oscillation due to the presence of the extreme observation (x_a, y_a).

In general, in presence of k outliers, the graph $m \mapsto d(m)$ shows at least one sharp "Λ" effect and at most k "Λ" effects. However, in absence of outliers, the graph shows only slight "$\wedge$" oscillations as m increases and shall have a decreasing trend. To avoid any ambiguity of appreciation between sharp "Λ" effects and slight "$\wedge$" oscillations, we also make use of the concave envelopment of $m \mapsto d(m)$ (*i.e.* the lowest concave curve enveloping the graph).

The methodology : For a given order m, let $x(m)$ denote the observed input x_j for which $d(m) = |\tilde{\xi}_{m,n}(x_j) - \hat{\xi}_{m,n}(x_j)|$. Then the basic tool will be a picture plotting the graph of $d(m)$ and its concave envelopment for increasing equidistant values of m. Remember that $d(m) \searrow 0$ as $m \to \infty$. So, if the graph of $d(m)$ ends with an increasing slope, it should be redone by adding larger values of m until it ends with a decreasing slope. Note also that, if the graph of $d(m)$ is plotted by using $(2J+1)$ or $(2J+2)$ values of m (with $J = 1, 2, \ldots$), then it has at most J sharp "Λ" effects or slight "$\wedge$" oscillations. The different possible behaviors of the graph of $d(m)$ and its concave envelopment can be summarized as follows:

(a) If the shape of the entire graph of $m \mapsto d(m)$ is a sharp "Λ", then the order m_* at which the graph is maximal should indicate that the FDH point (x_k, y_k), with $y_k = \hat{\varphi}_n(x(m_*))$, is an isolated outlier. The concave envelopment curve should have also a sharp "Λ" effect. Likewise, if the entire graph of $d(m)$ shows a sharp "Λ" effect followed by a second one, that is, a structure "M", then each local maximum m_* allows to detect an outlier. In this case, the concave envelopment curve should have roughly a structure "$\cap$".

(b) If the graph of $d(m)$ begins with a sharp positive slope as m increases, it could have a global structure of at most J successive "Λ" effects. The local maxima corresponding to these sharp effects will allow to detect isolated outliers. Here also, the concave envelopment curve should have a structure "Λ" or "$\cap$".

(c) If, in contrast, the graph of $d(m)$ begins with a smooth positive slope followed by a decreasing trend showing a global maximum value very close from the initial value $d(1)$, this indicates the presence of only suspicious extreme observations (not really isolated). In this case, the concave envelopment curve should not have a clear structure "Λ" or "$\cap$".

(d) If, in contrary, the graph of $d(m)$ decreases overall, this indicates clearly the absence of both outliers and suspicious extremes. Here also, a structure "Λ" or "$\cap$" of the concave envelopment curve should not appear.

(e) If, instead, the graph of $d(m)$ begins with a decreasing slope followed by an increasing one, then we distinguish between two situations: either **(e1)** the (short) decreasing slope is too smooth compared with the (longer) increasing one showing roughly a curvature "$\checkmark$" for the first values of m, or **(e2)** the decreasing deviation is, at least, as important as the increasing one. In situation (e1), the concave envelopment curve should have a structure "Λ" or "$\cap$" whose maxima allow to detect isolated outliers as described in (a). In situation (e2), the concave envelopment curve should behave as the graph of $d(m)$ in (c) or (d) leading thus to the same conclusions.

In conclusion, the above description tells us that a "Λ" or "$\cap$" structure of the concave envelopment curve is necessary and sufficient for detecting outliers. It is also important to note that looking only at the graph of $d(m)$ may result in some confusion between the desirable "Λ" effects (isolated outliers) and possibly contestable "$\wedge$" oscillations (suspicious extremes). To overcome such a subjectivity of appreciation, it is best to overlay in the same picture the graph of $d(m)$ and its concave envelopment. Then only sharp "Λ" deviations of the graph of $d(m)$ whose maximal points belong to the concave envelopment curve should be retained to identify potential outliers. This smoothing strategy however allows to detect only few outliers per picture. An outlier can "mask" other outliers situated near the first one and who are less isolated. To avoid such a masking effect pointed out earlier by Wilson (1993,1995), the analysis should be redone without the identified outliers until the concave envelopment curve shows no more "Λ" or "$\cap$" effects. Then, a careful analysis is to plot again the last graph of $d(m)$ and its concave envelopment by using a refined sequence of "small" equidistant values of m in order to detect potential masked outliers at the left-border of the sample. Indeed, when the increasing values of m are large, our procedure cannot detect outliers having too small values of x_i since, in this case, $\tilde{\xi}_{m,n}(x_i) = \hat{\xi}_{m,n}(x_i) = \hat{\varphi}_n(x_i)$. All this results in the following simple practical algorithm (illustrated in Subsection 5.3):

[1] Plot the graph of $d(m)$ and its concave envelopment for $m = 1, [\frac{n}{10}], [\frac{2n}{10}], \ldots, [\frac{9n}{10}], n$.

[2] If the concave envelopment curve shows a "Λ" or "$\cap$" effect, then the order m_* at which this curve attains its maximum indicates that the FDH point (x_k, y_k), with $y_k = \hat{\varphi}_n(x(m_*))$, is a potential outlier. This suspicious point can be really identified as an isolated outlier only if the maximal value $d(m_*)$ is clearly distant above from the initial value $d(1)$. To avoid the masking effect, proceed again to step **[1]** without the identified outliers.

[3] If the concave envelopment curve shows neither a "Λ" nor a "∩" effect, let $m_1 > 1$ be the first value of m in the chosen sequence in step **[1]** at which the graph of $d(m)$ shows a decreasing deviation. Then,

[3a] if $m_1/10 \leq 1$, there are no isolated outliers in the sample of interest.

[3b] if $m_1/10 > 1$, proceed to **[1]** by using $m = 1, [\frac{m_1}{10}], [\frac{2m_1}{10}], \ldots, [\frac{9m_1}{10}], m_1$.

Multivariate extensions : Let us now extend the ideas to the full multivariate setup where a set of inputs $X \in \mathbb{R}^p_+$ is used to produce a set of outputs $Y \in \mathbb{R}^q_+$. Let Ψ denote the joint support of the random vector (X, Y) that we assume to be free disposal, *i.e.*, $(x, y) \in \Psi$ implies $(x', y') \in \Psi$ as soon as $x' \geq x$ and $y' \leq y$ (the inequalities here have to be understood componentwise). Let $Y^{(j)}$, $(y^{(j)})$ denote the jth component of Y, (of y). Since a natural ordering of Euclidean spaces of dimension greater than one does not exist, we overcome the difficulty by utilizing the conditional distribution of the dimensionless transformation $\mathcal{Y}^y := \min_{j=1,\ldots,q} Y^{(j)}/y^{(j)}$ given $X \leq x$ instead of the multivariate distribution of $Y \in \mathbb{R}^q_+$ conditioned by $X \leq x$. The distribution function of this univariate transformation is given by

$$\mathbb{P}(\mathcal{Y}^y \leq \lambda | X \leq x) = 1 - \mathbb{P}(Y > \lambda y | X \leq x) = 1 - S_{Y|X}(\lambda y|x) \quad \text{for all} \quad \lambda \geq 0,$$

where $S_{Y|X}(\cdot|x)$ denotes the conditional survival function of Y given $X \leq x$. Its endpoint

$$\lambda(x, y) := \sup\{\lambda \geq 0 | S_{Y|X}(\lambda y|x) > 0\}$$

coincides with the conventional Farrell efficiency score, $\sup\{\lambda \geq 0 | (x, \lambda y) \in \Psi\}$, for the unit $(x, y) \in \Psi$, and the set $Y^\partial(x) := \{\lambda(x, y)y \,|\, y : (x, y) \in \Psi\}$ represents the set of maximal outputs a unit operating at the level x can produce. The point $y^\partial(x) := \lambda(x, y)y$ is the radial projection of (x, y) on the support frontier $Y^\partial := \{(x, \lambda(x, y)y) \,|\, (x, y) \in \Psi\}$ in the output-orientation (orthogonal to the vector x). In the particular case of $q = 1$, we have the equalities $\lambda(x, y) \equiv \varphi(x)/y$ and $Y^\partial(x) \equiv \{\varphi(x)\}$.

Parallely to the concepts of partial frontier functions $q_\alpha(x)$ and $\xi_m(x)$ related to the conditional distribution of Y given $X \leq x$ in the case of one output, we define the quantile function of order α and the expected maximum output function of order m for the dimensionless distribution of $\mathcal{Y}^y$ given $X \leq x$, respectively, as

$$Q_\alpha(x, y) := \inf\{\lambda \geq 0 | 1 - S_{Y|X}(\lambda y|x) \geq \alpha\}, \quad \mathcal{X}_m(x, y) := \int_0^{\lambda(x,y)} \{1 - [1 - S_{Y|X}(\lambda y|x)]^m\} d\lambda.$$

As a matter of fact, $\mathcal{X}_m(x, y)$ coincides with the order-m output efficiency score for the unit (x, y), introduced by Cazals *et al.* (2002), while $Q_\alpha(x, y)$ coincides with the αth

quantile output efficiency score favored by Daouia and Simar (2007). The sets $Y_\alpha^\partial := \{(x, Q_\alpha(x,y)y) \,|\, (x,y) \in \Psi\}$ and $Y_m^\partial := \{(x, \mathcal{X}_m(x,y)y) \,|\, (x,y) \in \Psi\}$ represent, respectively, the efficient order-α and order-m partial surfaces in the output direction. In the particular case of one output, $Q_\alpha(x,y) = q_\alpha(x)/y$ and $\mathcal{X}_m(x,y) = \xi_m(x)/y$. In this case, the sets Y_α^∂ and Y_m^∂ coincide with the graphs of the frontier functions $q_\alpha(\cdot)$ and $\xi_m(\cdot)$, respectively. See, *e.g.*, Daraio and Simar (2007) for a detailed description of both partial efficiency measures and for their economic meaning.

The sample estimators $\hat{Q}_{\alpha,n}(x,y)$ and $\hat{\mathcal{X}}_{m,n}(x,y)$ of $Q_\alpha(x,y)$ and $\mathcal{X}_m(x,y)$, respectively, are obtained by replacing $S_{Y|X}(\lambda y|x)$ with its empirical version $\hat{S}_{Y|X}(\lambda y|x) = \sum_{i=1}^n \mathbb{1}(X_i \leq x, Y_i > \lambda y)/\sum_{i=1}^n \mathbb{1}(X_i \leq x)$. They can be easily computed in the same way as the quantities $\hat{q}_{\alpha,n}(x)$ and $\hat{\xi}_{m,n}(x)$, respectively, by simply replacing in (2.1) and (2.4) the Y_i's such that $X_i \leq x$ with the dimensionless observations $\mathcal{Y}_i^y$ such that $X_i \leq x$. Moreover, it is not hard to show that all robustness and sensitivity properties established in the univariate case for the classes $\{q_\alpha(x), \hat{q}_{\alpha,n}(x)\}$ and $\{\xi_m(x), \hat{\xi}_{m,n}(x)\}$ hold true for the transformations $\{Q_\alpha(x,y), \hat{Q}_{\alpha,n}(x,y)\}$ and $\{\mathcal{X}_m(x,y), \hat{\mathcal{X}}_{m,n}(x,y)\}$. In particular, the practical algorithm described above in the three steps [1]-[3] for detecting potential outliers remains still valid in the full multivariate case up to two natural adaptations :

i. the maximal distance $d(m)$ between the curves of $\hat{q}_{\alpha(m),n}(\cdot)$ and $\hat{\xi}_{m,n}(\cdot)$ in the case of $q = 1$ extends naturally to the distance

$$d(m) = \max_{1 \leq i \leq n} ||\hat{Q}_{\alpha(m),n}(X_i, Y_i)Y_i - \hat{\mathcal{X}}_{m,n}(X_i, Y_i)Y_i||$$

between the empirical partial surfaces $\hat{Y}_{\alpha(m),n}^\partial = \{(X_i, \hat{Q}_{\alpha(m),n}(X_i,Y_i)Y_i) \,|\, i = 1, \ldots, n\}$ and $\hat{Y}_{m,n}^\partial = \{(X_i, \hat{\mathcal{X}}_{m,n}(X_i,Y_i)Y_i) \,|\, i = 1, \ldots, n\}$ in the general case of $q \geq 1$, where $||\cdot||$ denotes the Euclidean norm on $\mathbb{R}^q$;

ii. the outlying FDH point (X_k, Y_k) to be identified in step [2], with $Y_k = \hat{\varphi}_n(x(m_*))$ in the case of one output, is determined in the case of multi-outputs by

$$Y_k = \hat{\lambda}_n\left(x(m_*), y(m_*)\right) y(m_*),$$

where $\hat{\lambda}_n(x,y) = \sup\{\lambda \geq 0 | \hat{S}_{Y|X}(\lambda y|x) > 0\} = \max_{i|X_i \leq x} \min_{j=1,\ldots,q} Y_i^{(j)}/y^{(j)}$ is the FDH estimator of $\lambda(x,y)$, and where $(x(m_*), y(m_*))$ is the observation (X_j, Y_j) for which $d(m_*) = ||\hat{Q}_{\alpha(m),n}(X_j,Y_j)Y_j - \hat{\mathcal{X}}_{m,n}(X_j,Y_j)Y_j||$.

Subsection 5.3 illustrates the procedure with simulated and real data.

4 Asymptotic properties

We first derive the following pointwise and uniform asymptotic representations for $\hat{\xi}_{m,n}(x)$.

Proposition 4.1. *(i) For all $m \geq 1$ and any $x \in \mathbb{R}^p_+$ such that $F_X(x) > 0$, we have*

$$\sqrt{n}(\hat{\xi}_{m,n}(x) - \xi_m(x)) = \sqrt{n}\Phi_{m,n}(x) + o_p(1) \quad as \quad n \to \infty \tag{4.1}$$

where $\Phi_{m,n}(x) = \frac{m}{F_X(x)}\hat{F}_{X,n}(x) \int_0^{\varphi(x)} F^{m-1}(y|x)[F(y|x) - \hat{F}_n(y|x)]dy$.

(ii) Suppose the upper boundary of the support of Y is finite. Then, for all $m \geq 1$ and any $\mathcal{X} \subset \mathbb{R}^p_+$ such that $\inf_{x\in\mathcal{X}} F_X(x) > 0$, (4.1) holds uniformly in $x \in \mathcal{X}$, i.e. $\{\sqrt{n}(\hat{\xi}_{m,n}(x) - \xi_m(x));\ x \in \mathcal{X}\} = \{\sqrt{n}\Phi_{m,n}(x);\ x \in \mathcal{X}\} + o_p(1)$.

As an immediate consequence of Proposition 4.1(i), $\sqrt{n}(\hat{\xi}_{m,n}(x)-\xi_m(x))$ is asymptotically normal with mean 0 and variance

$$\begin{aligned} \sigma^2(x,m) &= \mathbb{E}\left\{\frac{m}{F_X(x)}\mathbb{1}(X \leq x)\int_0^{\varphi(x)} F^{m-1}(y|x)\,[F(y|x) - \mathbb{1}(Y \leq y)]\,dy\right\}^2 \\ &= \frac{2m^2}{F_X(x)}\int_0^{\varphi(x)}\int_0^{\varphi(x)} F^m(y|x)F^{m-1}(u|x)[1 - F(u|x)]\,\mathbb{1}(y \leq u)dydu. \end{aligned} \tag{4.2}$$

Even more strongly, it follows from Proposition 4.1(ii) (see also the proof) that the process $\{\sqrt{n}(\hat{\xi}_{m,n}(x) - \xi_m(x)),\ x \in \mathcal{X}\}$ converges in distribution in the space $L^\infty(\mathcal{X})$ of bounded functions on $\mathcal{X}$ to the centered Gaussian process $\{\mathbb{G}_m(x);\ x \in \mathcal{X}\}$ as $n \to \infty$, where

$$\mathbb{G}_m(x) = \frac{m}{F_X(x)}\int_0^{\varphi(x)} F^{m-1}(y|x)\,[\mathbb{F}(x,\infty)F(y|x) - \mathbb{F}(x,y)]\,dy$$

with $\mathbb{F}$ being a $(p+1)$ dimensional F-Brownian bridge. Similar results can be found in Cazals et al (2002, Theorem 3.1 and Appendix B). Their techniques of proof rely on the differentiability of the operator $S^{m,x}$ in the Fréchet sense with respect to the sup-norm. In statistical applications however, Fréchet differentiability may not hold, whereas Hadamard differentiability does, the latter being a less restrictive concept of differentiability than the former. The results in Proposition 4.1 are derived by applying the functional delta method in conjunction with the (less restrictive) Hadamard differentiability.

Note that the functional convergence of the process $\{\sqrt{n}(\hat{\xi}_{m,n}(x) - \xi_m(x)),\ x \in \mathcal{X}\}$ in Proposition 4.1(ii) provides the consistency and asymptotic distribution of parametric approximations of the order-m frontiers, as shown in Florens and Simar (2005). Their elegant approach tries to capture the shape of the cloud points near its boundary by combining parametric and nonparametric approaches.

Next we show that $\sqrt{n}(\hat{\xi}_{m,n}(x) - \xi_m(x))$ also obeys a law of the iterated logarithm, which improves the order of convergence to $O(\sqrt{\log\log n})$ and even gives the proportionality constant.

Theorem 4.1. *For all $m \geq 1$ and any $x \in \mathbb{R}_+^p$ such that $F_X(x) > 0$, we have almost surely for either choice of sign*

$$\limsup_{n\to\infty} \pm \frac{\sqrt{n}(\hat{\xi}_{m,n}(x) - \xi_m(x))}{(2\log\log n)^{1/2}} = \sigma(x,m).$$

By the asymptotic normality we have $\lim_{n\to\infty} \mathbb{P}\{\xi_m(x) \in [\hat{\xi}_{m,n}(x) \pm 2\sigma(x,m)/\sqrt{n}]\} = 2\Phi(2) - 1 \approx 95\%$, where Φ denotes the standard normal distribution function. An intriguing implication of the law of the iterated logarithm (see, *e.g.*, Serfling 1980) is that we can be sure that $\xi_m(x)$ is outside the asymptotic confidence interval $[\hat{\xi}_{m,n}(x) \pm 2\sigma(x,m)/\sqrt{n}]$ infinitely often, but this is of little practical consequence. Monte-Carlo experiments are provided in Subsection 5.2 to illustrate the performance of the asymptotic confidence interval $Q_n := [\hat{\xi}_{m,n}(x) \pm z\hat{\sigma}(x,m)/\sqrt{n}]$ which satisfies $\lim_{n\to\infty} \mathbb{P}[\xi_m(x) \in Q_n] = 2\Phi(z) - 1$ for any $z > 0$, where $\hat{\sigma}^2(x,m)$ is a strongly consistent estimator of $\sigma^2(x,m)$:

$$\hat{\sigma}^2(x,m) = \frac{m^2}{\hat{F}_{X,n}(x)} \int_0^{\hat{\varphi}_n(x)} \int_0^{\hat{\varphi}_n(x)} [\hat{F}(y|x)\hat{F}(u|x)]^{m-1} \left\{ \hat{F}(y \wedge u|x) - \hat{F}(y|x)\hat{F}(u|x) \right\} dydu.$$

Note that similar results to Proposition 4.1 and Theorem 4.1 have been proved for $\sqrt{n}(\hat{q}_{\alpha,n}(x) - q_\alpha(x))$ in Daouia (2005) and Daouia et al (2008). Note also that, as pointed in Section 1, we have $\sqrt{n}\,(\hat{q}_{\alpha,n}(x) - q_\alpha(x)) \xrightarrow{d} \mathcal{N}\,(0, \sigma^2(\alpha,x))$ as $n \to \infty$, where

$$\sigma^2(\alpha,x) = \alpha(1-\alpha)/f^2(q_\alpha(x)|x)F_X(x). \tag{4.3}$$

Then the interval $I_n = [\hat{q}_{\alpha,n}(x) \pm z\sigma(\alpha,x)/\sqrt{n}]$ satisfies $\lim_{n\to\infty} \mathbb{P}[q_\alpha(x) \in I_n] = 2\Phi(z) - 1$, for any $z > 0$. Putting $z = \Phi^{-1}(1 - a/2)$ to be the $(1-a/2)$th quantile of Φ, we obtain $(1-a)$ as the confidence coefficient. However, the computation of the asymptotic confidence interval I_n requires estimation of the quantile density function $f(q_\alpha(x)|x)$, which often results in estimates of unsatisfactory accuracy for finite samples. In the following theorem, we derive an alternative confidence interval for $q_\alpha(x)$ which is asymptotically equivalent to I_n, but does not need $f(q_\alpha(x)|x)$ to be known or estimated.

Theorem 4.2. *Let $0 < \alpha_1 < \alpha_2 < 1$ and assume that $F(\cdot|x)$ is continuously differentiable on the interval $[a,b] := [q_{\alpha_1}(x) - \varepsilon, q_{\alpha_2}(x) + \varepsilon]$ for some $\varepsilon > 0$, with strictly positive derivative $f(\cdot|x)$. For any $\alpha \in]\alpha_1, \alpha_2[$ and any $z > 0$, let $C_n =]\hat{q}_{\alpha_{n1},n}(x), \hat{q}_{\alpha_{n2},n}(x)[$ where $\alpha_{n1} = \alpha - z[\alpha(1-\alpha)/n\hat{F}_{X,n}(x)]^{1/2}$ and $\alpha_{n2} = \alpha + z[\alpha(1-\alpha)/n\hat{F}_{X,n}(x)]^{1/2}$. Then*

$$\lim_{n\to\infty} \mathbb{P}[q_\alpha(x) \in C_n] = 2\Phi(z) - 1 \quad \textit{and} \quad \sqrt{n}|\textit{length}(C_n) - \textit{length}(I_n)| \xrightarrow{p} 0 \quad \textit{as} \quad n \to \infty.$$

In case the true partial frontiers $q_\alpha(\cdot)$ and $\xi_m(\cdot)$ coincide, one can compare the performances of the asymptotic confidence intervals C_n and Q_n. See Subsection 5.2.

It should be clear that the estimation of the partial frontiers $q_\alpha(\cdot)$ and $\xi_m(\cdot)$ instead of the full frontier $\varphi(\cdot)$ itself is mainly motivated by the construction of robust frontier estimators which are well inside the sample $\{(X_i, Y_i), i = 1, \ldots, n\}$ but near its upper boundary. It is then natural to investigate whether the asymptotic normality of the estimators $\hat{q}_{\alpha,n}(x)$ and $\hat{\xi}_{m,n}(x)$ is still valid when $\alpha = \alpha_n \to 1$ and $m = m_n \to \infty$ as $n \to \infty$. First, note that

$$\lim_{\alpha\uparrow 1} \hat{q}_{\alpha,n}(x) = \lim_{m\uparrow\infty} \hat{\xi}_{m,n}(x) = \hat{\varphi}_n(x).$$

Note also that the necessary and sufficient condition under which the FDH estimator $\hat{\varphi}_n(x)$ converges to a non-degenerate distribution is given by

$$1 - F(y|x) = \ell_x\left(\{\varphi(x) - y\}^{-1}\right)\{\varphi(x) - y\}^{\rho_x} \quad \text{as} \quad y \uparrow \varphi(x) \tag{4.4}$$

(Daouia et al (2010), Theorem 2.1), where $\rho_x > 0$ is a constant and ℓ_x is a slowly varying function, *i.e.*, $\lim_{t\uparrow\infty} \ell_x(tz)/\ell_x(t) = 1$ for all $z > 0$. In the particular case where $\ell_x(\{\varphi(x) - y\}^{-1}) = \ell(x)$ is a strictly positive function in x, it is shown in Daouia et al (2010, Corollary 2.1) that

$$\{n\ell(x)\}^{1/\rho_x}\big(\varphi(x) - \hat{\varphi}_n(x)\big) \xrightarrow{d} \text{Weibull}(1, \rho_x) \quad \text{as} \quad n \to \infty.$$

For the estimator $\hat{q}_{\alpha_n,n}(x)$ to keep the same limit Weibull distribution as $\hat{\varphi}_n(x)$, it suffices to choose $\alpha_n \to 1$ rapidly so that $n^{1+1/\rho_x}(1 - \alpha_n) \to 0$ (see Daouia et al (2010), Theorem 2.2). This result has been also proved by Aragon et al (2005) in the restrictive case where the joint density of (X, Y) has a sudden jump at the frontier, which corresponds to $\rho_x = p+1$ in (4.4). Likewise, in this restrictive setting, Cazals et al (2002) recover the same asymptotic Weibull distribution of $\hat{\varphi}_n(x)$ for the estimator $\hat{\xi}_{m_n,n}(x)$ provided that $m_n = O(n \log n)$.

Instead of the Weibull extreme-value distribution, we provide in the next proposition sufficient conditions under which $\hat{q}_{\alpha_n,n}(x)$ and $\hat{\xi}_{m_n,n}(x)$ are rather asymptotically normal.

Proposition 4.2. *(i) Suppose* (4.4) *holds with* $\ell_x(\{\varphi(x) - y\}^{-1}) = \ell(x) > 0$ *and* $F(\cdot|x)$ *is differentiable in a left neighborhood of* $\varphi(x)$ *with a strictly positive derivative* $f(\cdot|x)$. *If* $n(1 - \alpha_n) \to \infty$ *as* $n \to \infty$, *then* $\sqrt{n}\{\sigma(\alpha_n, x)\}^{-1}\left(\hat{q}_{\alpha_n,n}(x) - q_{\alpha_n}(x)\right) \xrightarrow{d} \mathcal{N}(0, 1)$.

(ii) If $m_n \to \infty$ *and* $\frac{m_n(m_n-1)}{\sigma(x,m_n)} = O\left(\frac{\sqrt{n}}{\log\log n}\right)$ *as* $n \to \infty$, *then* $\sqrt{n}\{\sigma(x, m_n)\}^{-1}(\hat{\xi}_{m_n,n}(x) - \xi_{m_n}(x)) \xrightarrow{d} \mathcal{N}(0, 1)$.

Thus, the convergence in distribution of both $\frac{\sqrt{n}}{\sigma(\alpha,x)}(\hat{q}_{\alpha,n}(x) - q_\alpha(x))$ and $\frac{\sqrt{n}}{\sigma(x,m)}(\hat{\xi}_{m,n}(x) - \xi_m(x))$ to $\mathcal{N}(0, 1)$, for fixed orders α and m, is still valid when the partial frontiers $q_\alpha(x)$ and $\xi_m(x)$ approach the true full frontier $\varphi(x)$.

5 Numerical Illustration

We present simulation studies to illustrate the robustness and statistical efficiency of the empirical partial boundaries $\hat{\xi}_{m,n}$ and $\tilde{\xi}_{m,n} := \hat{q}_{\alpha(m),n}$ and to compare the asymptotic confidence intervals C_n and Q_n. We also provide illustrations with a real data set.

5.1 Comparing $\hat{\xi}_{m,n}$ and $\tilde{\xi}_{m,n}$

Simulated example: Consider the Cobb-Douglas model $Y = X^{1/2}\exp(-U)$, where X is uniform on $[0,1]$ and U is exponential with mean $1/3$. This model was studied by Gijbels et al (1999) among others. Here $\varphi(x) = x^{1/2}$ and $F(y|x) = 3x^{-1}y^2 - 2x^{-3/2}y^3$ for $0 < x \leq 1$ and $0 \leq y \leq \varphi(x)$. As can be seen from Figure 1, in this example the theoretical partial frontiers ξ_m (solid lines) and $\tilde{\xi}_m = q_{\alpha(m)}$ (dotted lines) are very close.

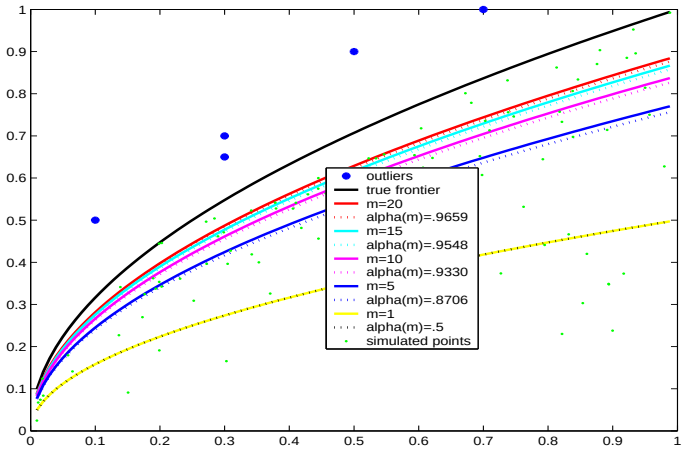

Figure 1: *The true frontiers ξ_m and $\tilde{\xi}_m$ for several values of m (Cobb-Douglas model).*

We also represent in Figure 1 a simulated sample of size 100 (green points) and we add five outliers (blue points) to this sample. For the resulting sample $(X,Y)^n$ of size $n = 105$, we compute the finite sample breakdown points $RB(\tilde{\xi}_{m,n}(x)) := RB(\hat{q}_{\alpha(m),n}(x), (X,Y)^n)$ for several values of m and x and provide in Table 1 the values $n \times RB(\tilde{\xi}_{m,n}(x))$.

Since the data set contains five outlying points in the output-direction, the estimator $\tilde{\xi}_{m,n}(x)$ can break down whenever $RB(\tilde{\xi}_{m,n}(x)) \leq 5/n$. This is clearly seen from Figure 2 where the frontiers $\tilde{\xi}_{m,n}$ and $\hat{\xi}_{m,n}$ are plotted in absence of outliers (bottom: $n = 100, 200, 300$) and in presence of the 5 outliers (top: $n = 105, 205, 305$). Moreover, as pointed in Remark 2.2, once $\tilde{\xi}_{m,n}(x)$ breaks down, it becomes less resistant to the influential outliers than $\hat{\xi}_{m,n}(x)$ as m increases. This is exactly what happens for $\tilde{\xi}_{25,105}$ and $\tilde{\xi}_{25,205}$ at $x = 0.3$, where these order-$\alpha(25)$ frontiers are clearly more influenced than $\hat{\xi}_{25,105}$ and $\hat{\xi}_{25,205}$ (here $m = 25$). In contrast, before breaking down at the point $x = 0.3$, we see that $\tilde{\xi}_{10,105}$

and $\tilde{\xi}_{10,205}$ (here $m = 10$) are rather more robust than $\hat{\xi}_{10,105}$ and $\hat{\xi}_{10,205}$, respectively.

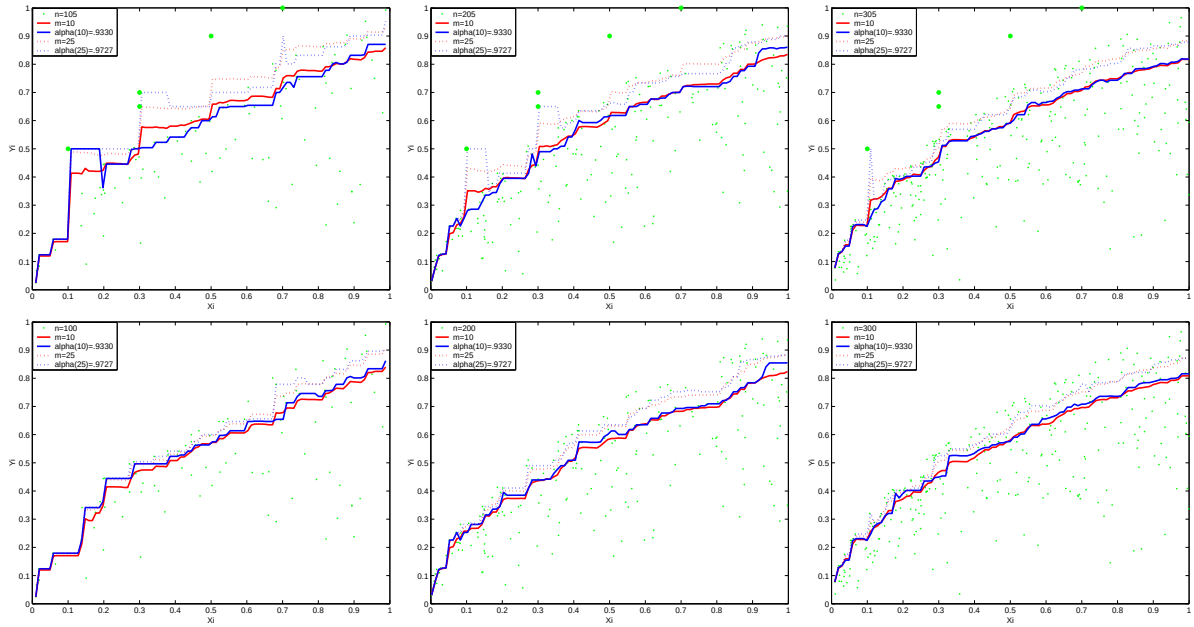

Figure 2: *In each picture,* $\tilde{\xi}_{10,n}$ *and* $\tilde{\xi}_{25,n}$ *(respectively:* $\hat{\xi}_{10,n}$ *and* $\hat{\xi}_{25,n}$*) in solid and dotted blue (respectively: red) lines. From left to right and from top to bottom:* $n = 105, 205, 305, 100, 200, 300$.

On the other hand, for too small values of x (*e.g.* $x = 0.1$), we see that both $\hat{\xi}_{m,n}(x)$ and $\tilde{\xi}_{m,n}(x)$ coincide with the non-robust FDH estimator, or at least, are drastically attracted by $\hat{\varphi}_n(x)$. As pointed out in Remark 2.1, this left-border defect is due to the conditioning on $X \leq x$ in the construction of these two estimators. However, when the number $n\hat{F}_{X,n}(x)$ of observations (X_i, Y_i) with $X_i \leq x$ increases, we see clearly that both $\hat{\xi}_{m,n}$ and $\tilde{\xi}_{m,n}$ become more robust to the outlying points.

We also simulated 1000 samples of size $n = 1000$ to analyze the bias and the mean squared error (MSE) of $\hat{\xi}_{m,n}$ and $\tilde{\xi}_{m,n}$ as estimators of $\xi_m \simeq \tilde{\xi}_m$. According to the numerical results reported in Table 2 (l-h.s), we can see that $\hat{\xi}_{m,n}$ is slightly more efficient than $\tilde{\xi}_{m,n}$ in terms of MSE, whereas the latter estimator is better than the former in terms of bias. When the data are contaminated by adding five outliers (indicated by "$*$" in Figure 1), we see in Table 2 (r-h.s) the improvement of $\tilde{\xi}_{m,n}$ over $\hat{\xi}_{m,n}$ in terms of MSE. Moreover, $\tilde{\xi}_{m,n}$ still outperforms $\hat{\xi}_{m,n}$ in terms of bias. Therefore, we can say that $\tilde{\xi}_{m,n}$ is globally more robust to the outlying points than $\hat{\xi}_{m,n}$ in this particular example. This can be explained by the fact that, even when the $\alpha(m)$-frontier $\tilde{\xi}_{m,n}$ breaks down at a value x, it is influenced only locally on a right neighborhood of x, whereas the m-frontier $\hat{\xi}_{m,n}$ remains attracted overall between the 5 outliers as illustrated in Figure 2. Remember in comparing the $\alpha(m)$- and

m-frontiers that, in absence of outliers, $\tilde{\xi}_{m,n}$ is almost overall larger than or equal to $\hat{\xi}_{m,n}$, which is no more the case when adding the five outliers.

A real data set: To further illustrate the sensitivity and resistance properties of the empirical partial frontiers $\hat{\xi}_{m,n}$ and $\tilde{\xi}_{m,n}$, we use the real data example of Cazals et al (2002) and Aragon et al (2005) 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 compared $\hat{\xi}_{m,n}$ and $\tilde{\xi}_{m,n}$ for different orders $m \in \{100, 200, 1000, 4000\}$. The cloud of points and the resulting estimates are provided in Figure 3 (top): for m large enough ($m \in \{100, 200\}$), the quantile-based frontier $\tilde{\xi}_{m,n}$ is clearly more resistant to the extreme points than the expected maximal output frontier $\hat{\xi}_{m,n}$. But for m too large (*i.e.* $m \in \{1000, 4000\}$), both partial boundaries $\tilde{\xi}_{m,n}$ and $\hat{\xi}_{m,n}$ are drastically influenced by the few ostensible FDH points. Nevertheless, while $\tilde{\xi}_{4000,n}$ coincides overall with the FDH frontier, $\hat{\xi}_{4000,n}$ has the advantage to be still resistant to this envelopment frontier. These results are expected in our theory.

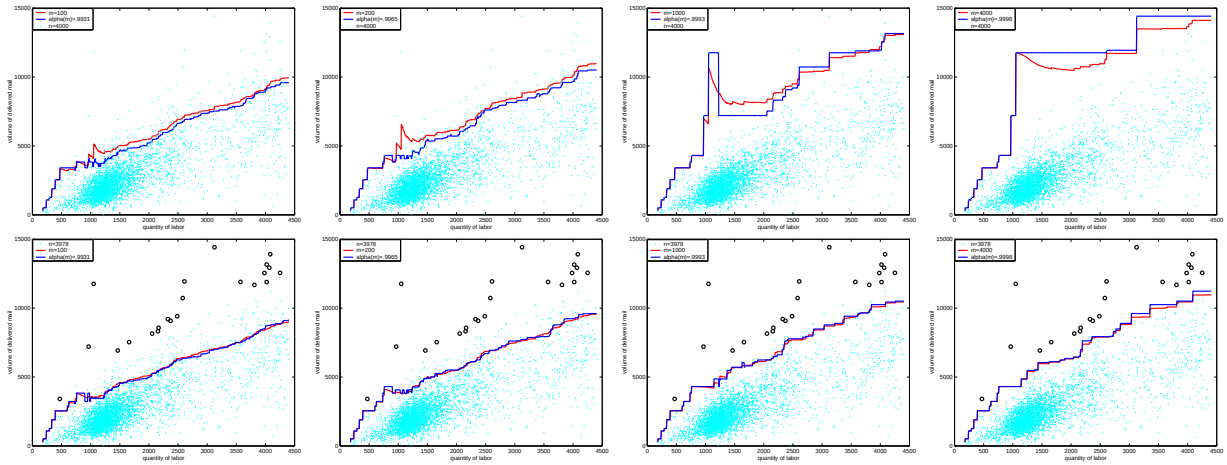

Figure 3: *Top (full sample):* $\hat{\xi}_{m,n}$ *and* $\tilde{\xi}_{m,n}$ *for* $m = 100, 200, 1000, 4000$. *Bottom: as above without the anomalous data indicated by circles.*

5.2 Comparing C_n and Q_n

In the Cobb-Douglas model described above, as pointed in Daouia and Ruiz-Gazen (2006), $\xi_m(\cdot)$ coincides with $q_\alpha(\cdot)$ if and only if $\alpha = \frac{1}{2}(1 - \cos[3\arccos(\frac{1}{2} - B_m) - 4\pi])$, with $B_m = \sum_{j=0}^{m} \binom{m}{j} 3^j (-2)^{m-j}/(3m - j + 1)$. For example, we obtain $\alpha = .8557$ for $m = 5$ and $\alpha = .9242$ for $m = 10$. In this case, the partial frontier $\xi_m \equiv q_\alpha$ can be estimated by $\hat{\xi}_{m,n}$ as well as $\hat{q}_{\alpha,n}$, and one can compare the confidence intervals Q_n and C_n. The true partial frontier and its 95% confidence intervals are displayed in Figure 4 with $(m = 5, \alpha = .8557)$ on the l-h.s and $(m = 10, \alpha = .9242)$ on the r-h.s. Here we consider two simulated samples of size $n = 100$ (top) and $n = 1000$ (bottom).

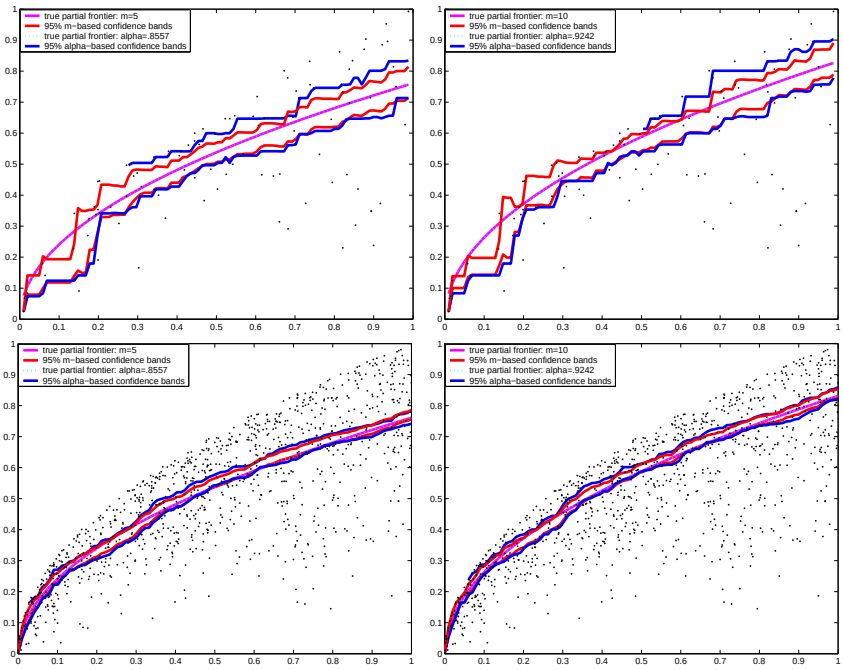

Figure 4: 95% *confidence intervals* Q_n *(red) and* C_n *(blue) of* $\xi_m = q_\alpha$ *(Simulated example).*

By construction, the upper bound $\hat{q}_{\alpha_{n2},n}(x)$ of C_n does not exist (see the upper blue solid lines) for small inputs-usage x and high values of α which result in orders $\alpha_{n2} > 1$. This is the major drawback of the confidence interval C_n. On the other hand, even if the confidence bands of Q_n (in red lines) are overall well-defined, they do not contain $\xi_m(x)$ for small levels x and high orders m. Apart from these left-border defects, we observe that C_n and Q_n have very similar lower bounds, but Q_{100} performs globally better than C_{100} in terms of upper bounds. This is the price to be paid in order to avoid the estimation of the conditional quantile density $f(q_\alpha(x)|x)$ involved in the asymptotic variance $\sigma^2(\alpha, x)$ of $\hat{q}_{\alpha,n}(x)$. For $n = 1000$, the two confidence intervals provide more similar results. Note also that Q_n is computationally prohibitive when the sample size is of the order of several thousands. On the contrary, C_n is very easy and very fast to implement.

Table 3 provides the average lengths and the achieved coverages of the 95% asymptotic confidence intervals C_n and Q_n computed over 1000 random replications, for sample sizes $n = 100$ and $n = 1000$. The table only displays results for values of x where the upper bound of C_n exists, *e.g.*, for x ranging over $\{0.4, 0.5, \cdots, 1\}$ when $m = 5$ and for $x \in \{0.65, 0.7, \cdots, 0.95\}$ when $m = 10$. For $n = 100$ both C_n and Q_n provide reasonably good confidence intervals, but Q_n performs clearly better than C_n in terms of average lengths. In contrast, C_n outperforms Q_n in terms of achieved coverages. For $n = 1000$ the confidence interval Q_n performs as C_n in terms of achieved coverages, however it still outperforms C_n

in terms of average lengths. We repeated this exercise for other values of m and MC trials and obtained the same conclusion: no winner in all contexts. On the other hand, when comparing the MSE and bias of $\hat{\xi}_{m,n}$ and $\hat{q}_{\alpha,n}$ as estimators of $\xi_m(x) = q_\alpha(x)$, we find here also that $\hat{\xi}_{m,n}$ is more efficient than $\hat{q}_{\alpha,n}$ in terms of MSE and that $\hat{q}_{\alpha,n}$ performs better in terms of bias. We do not reproduce the tables in order to save place.

5.3 Detection of anomalous data

A univariate simulated example: We consider the cloud of $n = 105$ points represented in Figure 1. The procedure based on the analysis of the curve of $m \mapsto d(m)$ and its concave envelopment will detect only the five points '*' as isolated outliers. The first picture in Figure 5 (l-h.s) gives these two curves for $m = 1, 10, 20, \ldots, 90, 100$: here the graph of $d(m)$, in solid blue line, shows clearly two sharp "Λ" effects and the concave envelopment curve, in dotted red line, has roughly a structure "$\cap$". The "Λ" effects attain their maximal values at $m_* = 10$ and $m_* = 40$. We also see that each maximal point $(m_*, d(m_*))$ belongs to the concave envelopment curve, which indicates that the FDH points (x_k, y_k), for which $y_k = \hat{\varphi}_n(x(m_*))$, can be really identified as isolated points in the direction of Y. A simple computation code (using matlab) allows to detect two outlying FDH points: the result is $(x_k, y_k) = (0.1, 0.5)$ for $m_* = 10$ and $(x_k, y_k) = (0.5, 0.9)$ for $m_* = 40$.

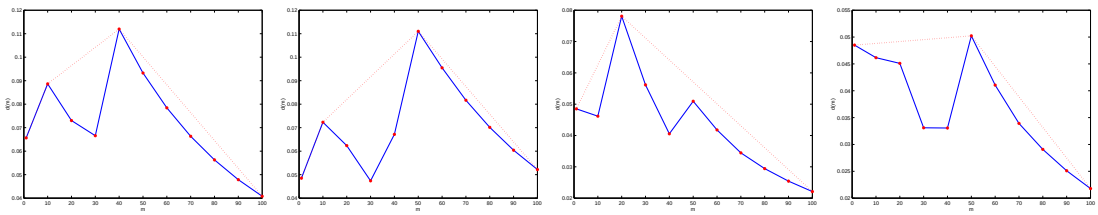

Figure 5: *Simulated example. The graph of $d(m)$ in solid blue line and its concave envelopment in dotted red line.*

To avoid the masking effect, we redid the same work on the same data set without the detected two outliers. The second picture of Figure 5 (from left to right) provides the resulting curve of $d(m)$ and its concave envelopment: here also looking to the sharp "Λ" effects of the graph of $d(m)$ which appear at $m_* = 10$ and $m_* = 50$, we identify respectively the additional outlying points $(0.3, 0.7)$ and $(0.7, 1)$.

When these two outliers are also deleted from the sample, we obtain the curves in the third picture of Figure 5: clearly the graph of $d(m)$ begins with a too smooth decreasing slope followed by a sharp "Λ" effect which attains its maximum at $m_* = 20$, and ends with a slight "$\wedge$" oscillation. Here, the shape of the entire concave envelopment curve shows an indisputable sharp "Λ" effect which attains its maximum at $m_* = 20$, indicating that the

FDH point $(x(20), \hat{\varphi}_n(x(20))) \equiv (0.3, 0.65)$ is an outlier. The slight "$\wedge$" oscillation of the graph of $d(m)$, attaining its maximum at $m = 50$, indicates only the presence of a suspicious extreme observation (not really isolated) since its maximal value $d(50)$ is very close to $d(1)$.

Now, when the five outliers $(0.1, 0.5)$, $(0.5, 0.9)$, $(0.3, 0.7)$, $(0.7, 1)$ and $(0.3, 0.65)$ are deleted from the sample, we get the curves in the last picture of Figure 5 (r-h.s): the concave envelopment curve shows neither a sharp "Λ" effect nor a "$\cap$" structure, which indicates the absence of really isolated outliers. Here, the graph of $d(m)$ begins with a decreasing deviation followed by a slight "$\wedge$" oscillation whose maximal value $d(50)$ is very close to the initial value $d(1)$ and so cannot be used for detecting outliers. Therefore only the five points indicated by '*' in Figure 1 are detected by our semi-automatic procedure, which is quite remarkable although "no optimal procedure nor miracle procedure can be defined to detect outliers" as stated by Simar (2003).

Application to postal data: We test our procedure on the French post offices data set which contains several outlying points in the output-orientation. Proceeding to step **[1]** and

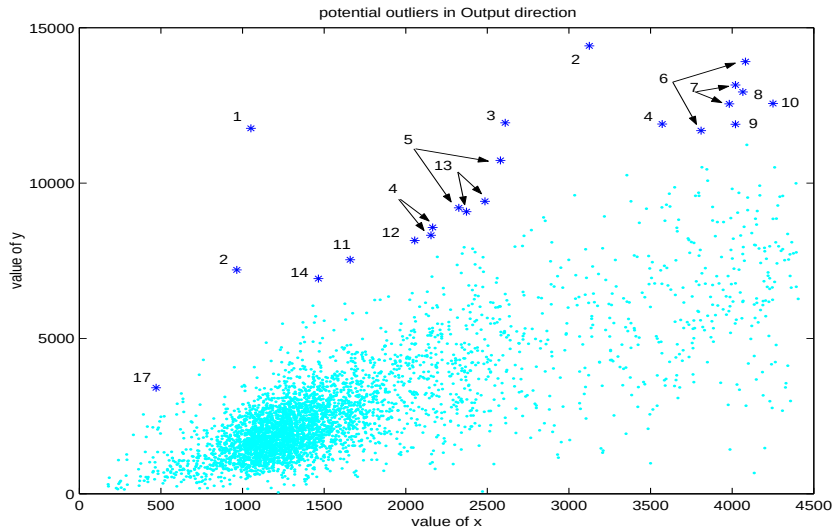

Figure 6: Potential outlying post offices detected by the semi-automatic procedure.

then step **[2]** of our algorithm for $m = 1, [\frac{n}{10}], [\frac{2n}{10}], \ldots, [\frac{9n}{10}], n$ (here $n = 4000$), we obtain successively the pictures in Figure 7 from left to right and from top to bottom. Except for the last picture, the concave envelopment curves in dotted lines show roughly "Λ" or "$\cap$" effects allowing to identify at most three outliers per picture indicated in Figure 6 by the number of the corresponding picture (from #1 to #14). Looking at the last picture, #15, we see a too smooth increasing slope (an approximately horizontal deviation) of the concave curve followed by a sharp decreasing slope, which makes the "Λ" effect clearly more contestable than the one appearing in the preceding picture. So we cannot proceed to step **[2]**.

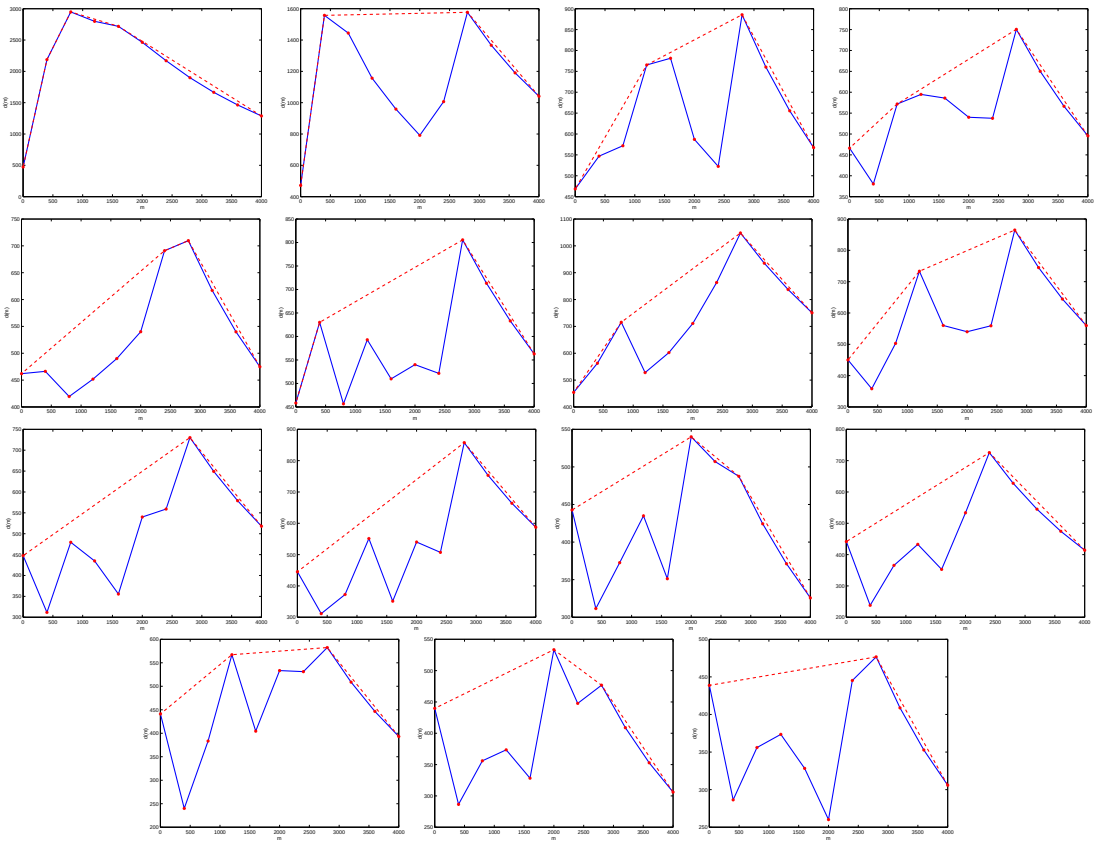

Figure 7: *The resulting pictures for* $m = 1, 400, 800, \ldots, 4000$ *(French post offices).*

Instead, we proceed to step **[3b]**, *i.e.*, the last picture is redone by using a sequence of smaller values of m in order to detect potential masked outliers at the left-border of the sample. Looking again to picture #15, the first value of m at which the graph of $d(m)$ (solid blue line) shows a decreasing deviation is $m_1 = 400$. The resulting new graph of $d(m)$ and its concave envelopment, for $m = 1, [\frac{m_1}{10}], [\frac{2m_1}{10}], \ldots, [\frac{9m_1}{10}], m_1$, is in the first picture of Figure 8: here we only see a slight "$\wedge$" oscillation of the concave curve whose maximal value is close to the initial value $d(1)$. The first value of m at which the graph of $d(m)$ shows a decreasing deviation being $m_1 = 80$, the last picture is redone by using the refined sequence $m = 1, 8, 16, \ldots, 80$. This gives the second picture of Figure 8, which allows to identify only one potential outlier indicated by #17 in Figure 6. When this point is deleted from the sample, we obtain the third picture of Figure 8 which shows no more "Λ" or "$\cap$" effects of the concave envelopment curve and so, there are no more outlying post offices. In summary, our semi-automatic procedure detects 22 potential outliers. Some of these points (*e.g.* #1,#2,#3) are clearly outlying due to measurement errors, but other isolated observations (*e.g.* #4,#9,#10,#17) might contain useful information on the process under analysis and so, they deserve to be carefully examined.

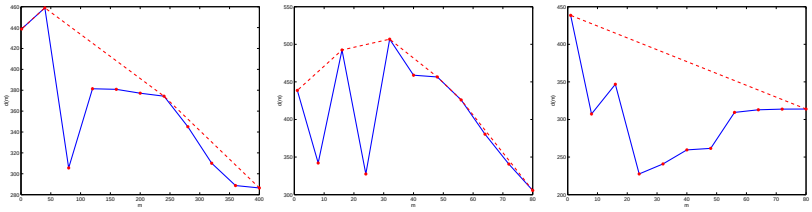

Figure 8: *As above with* $m = 1, 40, \ldots, 400$ *(l-h.s) and* $m = 1, 8, \ldots, 80$ *(Middle and r-h.s).*

A multivariate simulated example: Here a multi-input and multi-output ($p = q = 2$) data set is simulated as in Park *et al.* (2000). In this setup, the function describing the efficient frontier is given by

$$y^{(2)} = 1.0845(x^{(1)})^{0.3}(x^{(2)})^{0.4} - y^{(1)},$$

where $y^{(j)}$, $(x^{(j)})$, stands for the jth component of y, (of x), for $j = 1, 2$. We draw $X_i^{(j)}$ independent uniforms on $(1, 2)$ and $\tilde{Y}_i^{(j)}$ independent uniforms on $(0.2, 5)$. Then the generated random rays in the output space are characterized by the slopes $K_i = \tilde{Y}_i^{(2)}/\tilde{Y}_i^{(1)}$. The generated random points on the frontier are defined by

$$Y_{i,eff}^{(1)} = \frac{1.0845(X_i^{(1)})^{0.3}(X_i^{(2)})^{0.4}}{K_i + 1}, \quad Y_{i,eff}^{(2)} = 1.0845(X_i^{(1)})^{0.3}(X_i^{(2)})^{0.4} - Y_{i,eff}^{(1)}.$$

The efficiencies are generated by $\exp(-U_i)$ where U_i are drawn from an exponential with mean $1/3$. Finally, we define $Y_i = Y_{i,eff} * \exp(-U_i)$. We simulate 100 observations according to this scenario and we add five outliers $\#1, \ldots, \#5$, as in Daouia and Simar (2007), respectively at the following values of X: (1.25,1.5), (1.25, 1.75), (1.5,1.5), (1.75, 1.25) and (1.5, 1.25); the corresponding values for the slopes in the Y space are (0.25, 0.75, 1, 3, 5).

Our working algorithm results in the successive graphs of $d(m)$ and their concave envelopment curves displayed on Figure 9, with $m = 1, 10, 20, \ldots, 100$. In the first picture (from left to right and from top to bottom), the graph of $d(m)$ (solid blue line) and its concave envelopment curve (dashed red line) have a sharp structure "$\cap$" which attains its maximal values at $m_* = 10$ and $m_* = 30$ and allows to identify only the outlying observation #5. A similar "$\cap$" structure is obtained in the second picture after removing the first detected outlier : here also the attained maximal values at $m_* = 10$ and $m_* = 30$ allow to identify the same outlier #1. When this additional outlier is deleted from the sample, the new graphs in the third picture show a sharp "Λ" effect which appears at $m_* = 10$ and results in the identification of the outlier #2. The fourth picture provides the resulting graphs after removing this outlier : looking here to the structure "$\cap$" which attains its maximal values at $m_* \in \{20, 30, 60\}$, we identify the same outlying point #3. The new graphs obtained

without this outlier are shown in the fifth picture, where a sharp "Λ" effect of the concave envelopment curve, appearing at $m_* = 20$, allows to identify the outlier #4. For the same data set without this outlier, we get in the last picture a decreasing concave envelopment curve, indicating the absence of any suspicious observation among the remaining simulated 100 points. Thus, only the introduced five outliers are detected by our procedure. We repeated the same exercise with other simulated data sets with the same kind of results.

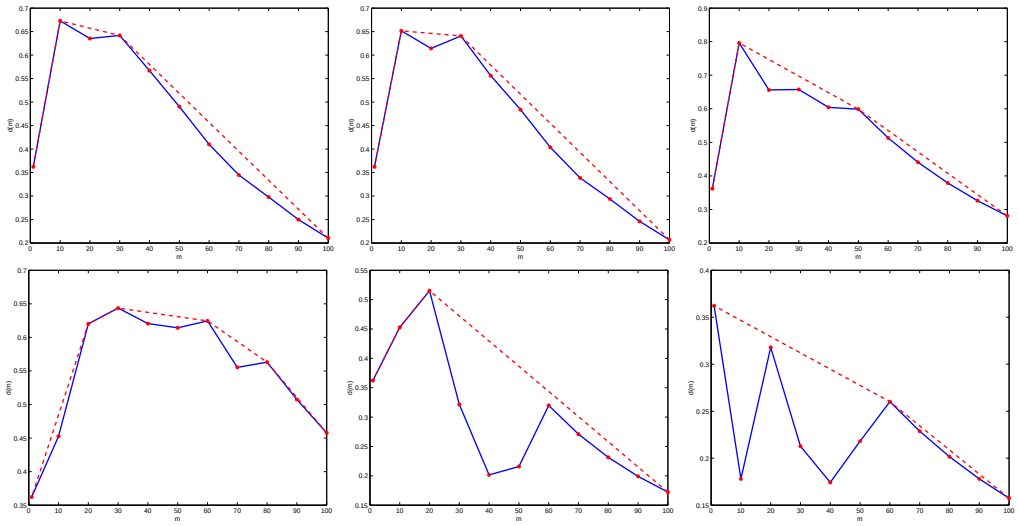

Figure 9: *The resulting pictures for* $m = 1, 10, 20, \ldots, 100$ *(multivariate simulated data).*

Application to PFT (Program Follow Through) data: We examine here the popular data set reported by Charnes, Cooper and Rhodes (1981) on an experimental education program administrated in 70 US schools, with $p = 5$ inputs and $q = 3$ outputs. The observations #59 and #44 are detected by the procedure of Wilson (1993) as potential outliers. The results obtained by Simar (2003) confirm this and point out two additional suspicious observations #54 and #1 that deserve at least careful attention. Our methodology confirms that only the units #59 and #44 are really isolated from the sample in the output-orientation. Moreover, it turns out that the unit #52 is more suspicious than the extreme observations #54 and #1. Proceeding to step **[1]** and then to step **[2]** of our algorithm for $m = 1, 7, 14, 21, \ldots, 70$, we find successively the eight pictures displayed on Figure 10.

In the first picture, the concave envelopment curve (dashed line) indicates a clear structure "Λ", and the graph of $d(m)$ (solid line) shows two sharp "Λ"effects which attain their maximal values at $m_* = 49$ [with $d(m_*) - d(1) = 32.9483$] and at $m_* = 21$ [with $d(m_*) - d(1) = 17.2833$]. Both local maximum points $(m_*, d(m_*))$ belong to the concave envelopment curve, but they allow to identify the same unit #59 as a potential outlier. When this suspicious unit is deleted from the sample, we obtain in the second picture a "Λ" (or

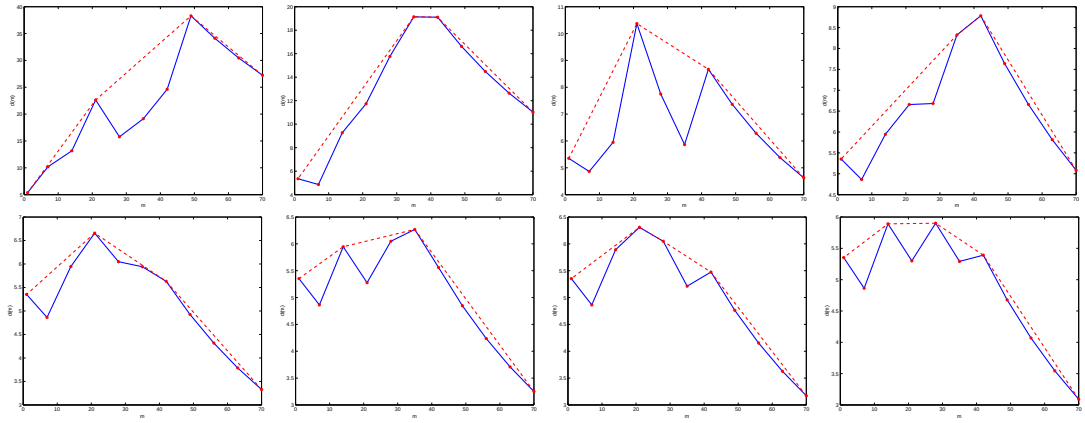

Figure 10: *The resulting pictures for* $m = 1, 7, 14, 21, \ldots, 70$ *(PFT data).*

roughly a "∩") structure of the graph of $d(m)$ and its concave envelopment : here also the two maximal values attained at $m_* = 35$ [with $d(m_*) - d(1) = 13.7869$] and at $m_* = 42$ [with $d(m_*) - d(1) = 13.7506$] result in one detected unit #44. When this extreme unit is deleted from the sample, we get the third picture which allows to identify the unit #52 at $m_* = 21$ [with $d(m_*) - d(1) = 5.0187$] and the unit #54 at $m_* = 42$ [with $d(m_*) - d(1) = 3.3166$]. When deleting these two additional suspicious points from the data set, we obtain in the fourth picture a structure "Λ" of the graph of $d(m)$ and its concave envelopment, which attains its maximum at $m_* = 42$ [with $d(m_*) - d(1) = 3.4307$] and allows to identify the unit #1 as a potential outlier. Likewise, our semi-automatic procedure detects,

- in picture 5, the unit #21 at $m_* = 21$ [with $d(m_*) - d(1) = 1.3047$];
- in picture 6, the unit #10 at $m_* = 35$ [with $d(m_*) - d(1) = 0.9163$] and the unit #27 at $m_* = 14$ [with $d(m_*) - d(1) = 0.5941$];
- in picture 7, the unit #12 at $m_* = 21$ [with $d(m_*) - d(1) = 0.9554$] and the unit #50 at $m_* = 28$ [with $d(m_*) - d(1) = 0.6958$];
- in picture 8, the unit #20 at $m_* = 14$ [with $d(m_*) - d(1) = 0.5382$] and the unit #16 at $m_* = 28$ [with $d(m_*) - d(1) = 0.5486$].

As a matter of fact, due to the high 8-dimensional space with a small sample ($n = 70$), we can identify more extreme points as potential outliers. However, we recall that before deleting any suspicious observation from the sample, our methodology requires to first check whether the suspicious point is really isolated in the output-orientation by comparing the maximal value $d(m_*)$ of the corresponding "Λ" effect with the initial value $d(1)$. As it can be seen from Figure 11 (l-h.s), which represents the difference $d(m_*) - d(1)$ for each suspicious

point, only the two units #59 and #44 can be really identified as potential outliers since, for each one of them, $d(m_*)$ is clearly distant above from $d(1)$. The three suspicious units #52, #54 and #1 cannot be viewed as isolated outliers, but they are certainly extreme/influential observations. In contrast, the remaining units (#21, #10, #27, #12, #50, #20, #16,...) are not even suspicious since the difference $d(m_*) - d(1)$ is clearly negligible for all of them.

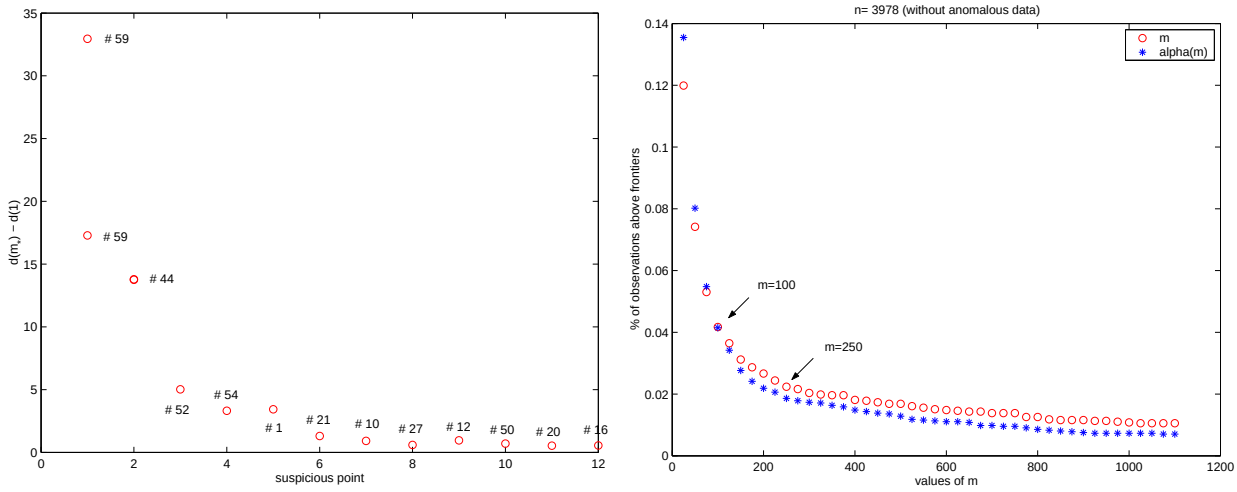

Figure 11: (l-h.s) The difference $d(m_*) - d(1)$ for the suspicious observations. (r-h.s) Evolution of the % of sample points outside the partial frontiers $\hat{\xi}_{m,3978}$ and $\tilde{\xi}_{m,3978}$.

5.4 Practical guidelines

In view of Proposition 2.2 and Theorems 2.1-2.2, we know that a significant difference between the expected-maximum estimate $\hat{\xi}_{m,n}$ and the median-maximum estimate $\tilde{\xi}_{m,n}$ indicates the presence of influential extreme observations above the order-m frontier that could be outlying. This suggests the following two steps in order to perform the frontier estimation:

Step 1: Apply the semi-automatic prescription (as illustrated above) in order to detect any potential outliers. Then consider the sample without the identified anomalous data points.

For the median- and mean-maximum estimators $\tilde{\xi}_{m,n}$ and $\hat{\xi}_{m,n}$ to provide similar conclusions, an intuitive idea is to seek the order m for which the percentage of sample points above each partial frontier is approximately the same. This leads to Step 2.

Step 2: Overlay in a same picture the evolution of the percentage of observations outside each partial frontier with respect to m. Remember that the sample still contains extreme points (not really outlying) that influence $\tilde{\xi}_{m,n}$ more or less than $\hat{\xi}_{m,n}$ following the values of m. Therefore the two decreasing percentage curves shall "cross" since $\tilde{\xi}_{m,n}$ is less sensitive (and so envelopes less points) than $\hat{\xi}_{m,n}$ to the magnitude of extreme outputs even when m increases, but once m attains a sufficiently large threshold, $\hat{\xi}_{m,n}$ becomes more resistant

(and so envelopes less points) than $\tilde{\xi}_{m,n}$. The value of m at which the two percentage curves cross corresponds to the most similar large order-m and order-$\alpha(m)$ frontiers that capture the shape of the cloud points near its optimal boundary.

The extreme observations left outside the resulting similar frontiers $\tilde{\xi}_{m,n}$ and $\hat{\xi}_{m,n}$ might be useful to emulate: the managers of any decision-making unit (DMU) operating at (x, y) and situated below these partial frontiers could study the relevant peers (X_i, Y_i) above $\hat{\xi}_{m,n}$ or $\tilde{\xi}_{m,n}$ among those dominating (x, y) (*i.e.* with $X_i \leq x$ and $Y_i \geq y$) in order to learn how to reduce inputs and/or increase outputs. Refined "relevant practices" that might be useful to emulate could be identified as follows: the partial frontiers $\tilde{\xi}_{m,n}$ and $\hat{\xi}_{m,n}$ being less sensitive to the choice of the order m as $m \to \infty$, the decrease of the percentage of points outside each frontier becomes approximately stable as $m \to \infty$. In particular, the first value of m from which the two percentage curves are approximately horizontal/stable, corresponds to the frontiers $\hat{\xi}_{m,n}$ and $\tilde{\xi}_{m,n}$ that are sensible to the magnitude of the most extreme observations whose outputs are highly desirable, but in the same time, they are resistant to these extremes in the sense that they do not envelope them. Such extreme practices could be emulated by the managers of dominated DMUs to improve their own operations.

Application to postal data: In order to capture in a robust way the shape and curvature of the sample boundary, we compared in Figure 3 (top) both partial boundaries $\hat{\xi}_{m,n}$ and $\tilde{\xi}_{m,n}$ for a sequence of large values of $m \in \{100, 200, 1000, 4000\}$. We observe a distance between the two frontiers, for each order m, due to the presence of outliers above the m-frontiers. Following our practical guidelines, a sensible practice is to remove, in a first step, the identified 22 potential outliers from the sample. Figure 3 (bottom) shows how $\hat{\xi}_{m,n}$ and $\tilde{\xi}_{m,n}$ become very close for the resulting sample of size $n = 3978$. Then, in a second step, we overlay in a same picture the evolution of the percentage of sample points outside each partial frontier with respect to m. As can be seen from Figure 11 (r-h.s), the two decreasing percentage curves cross at $m \approx 100$ and become approximately linearly stable from $m \approx 250$. The partial frontiers $\hat{\xi}_{m,n}$ and $\tilde{\xi}_{m,n}$ for $m \in \{100, 250\}$ are graphed in Figure 12 together with their 95% confidence intervals Q_n and C_n, respectively. $\hat{\xi}_{100,n}$ and $\tilde{\xi}_{100,n}$ are the largest order-m and order-$\alpha(m)$ frontiers which provide the most similar estimates. The extreme post offices left outside these frontiers, whose outputs are highly desirable, might be useful to emulate. The partial frontiers $\hat{\xi}_{250,n}$ and $\tilde{\xi}_{250,n}$ also provide a refined identification of relevant post offices to be emulated and satisfactory estimates of the shape of the sample boundary.

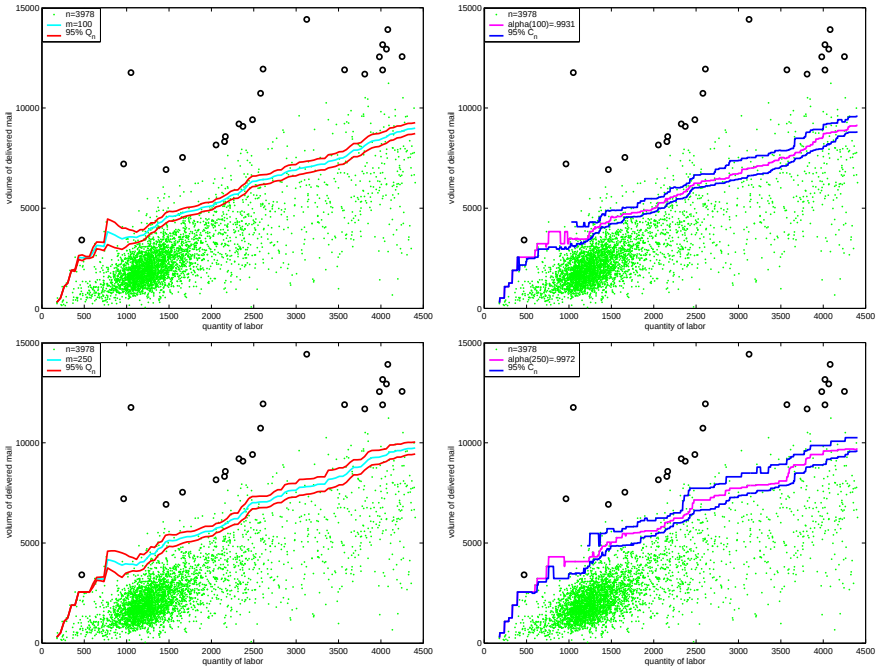

Figure 12: 95% *confidence intervals* Q_n *(left) and* C_n *(right). Here* $n = 3978$ *(without anomalous data). From top to bottom:* $m = 100$ *and* $m = 250$ *(French post offices).*

6 Conclusions

We show that the two classes of partial frontiers, $\{q_\alpha\}$ and $\{\xi_m\}$, are closely related when $\alpha = \alpha(m) = (1/2)^{1/m}$, in the same sense as the mean and median of a same distribution do. This answers in particular the important question of how to choose the order α as a function of m for a possible comparison between order-α and order-m frontiers.

Non of the two classes can be claimed to be preferable in all contexts. A sensible practice is to check whether both partial frontier analyses point toward similar conclusions. Obtaining different results from $\hat{\xi}_{m,n}$ and $\hat{q}_{\alpha(m),n}$, for sufficiently large values of m, indicates the presence of suspicious extreme data points that could be outlying or perturbed by noise. Before performing any frontier estimation, a useful empirical strategy is to first detect and remove the anomalous data and then to determine, in a second step, the order m at which the percentage of sample points outside the order-m and order-$\alpha(m)$ frontiers is approximately the same. This value of m corresponds to the largest frontiers $\hat{\xi}_{m,n}$ and $\hat{q}_{\alpha(m),n}$ having the most similar behaviors. These extreme partial frontiers provide satisfactory estimates of the shape of the sample boundary and identify relevant peers that might be useful to emulate.

The theoretical comparison between the reliability of $\{\hat{q}_{\alpha,n}\}$ and $\{\hat{\xi}_{m,n}\}$ is exploited to derive an appealing identification methodology, very easy and fast to implement and providing very good results. The use of partial frontiers for detecting influential observations is

not new. A basic tool can be found in Simar (2003) and Daraio and Simar (2007) consisting of a picture showing the evolution of the "proportion" of sample points outside either the order-m or order-α frontier as a function of the order and of another tuning parameter. Our prescription is rather based on the evolution of the "maximal distance" between the related order-m and order-$\alpha(m)$ frontiers as a function of m. Adapting this tool to the input-orientation is straightforward. Our robustness study provides also a theoretical justification for the descriptive technique of Simar (2003) and Daraio and Simar (2007).

We derive, among others, an asymptotic confidence interval C_n for $q_\alpha(x)$ not requiring estimating the conditional quantile density function. We provide sufficient conditions for ensuring the asymptotic normality of both $\hat{\xi}_{m,n}(x)$ and $\hat{q}_{\alpha,n}(x)$ for the limiting cases $m = m_n \to \infty$ and $\alpha = \alpha_n \to 1$ as $n \to \infty$. Instead of the assumption involving the asymptotic variance $\sigma^2(x, m_n)$ (see Proposition 4.2(ii)), a main challenge is to get the asymptotic normality of $\hat{\xi}_{m_n,n}(x)$ under the more conventional condition (4.4). This problem is worth investigating in future. When estimating the same partial frontier $q_\alpha(x) = \xi_m(x)$, the empirical study reveals interesting findings regarding the performances of the estimators $\hat{\xi}_{m,n}$ and $\hat{q}_{\alpha,n}$ and the performances of the confidence intervals Q_n and C_n.

Appendix: lemmas and proofs

A Robustness

Lemma A.1. *Let $x \in \mathbb{R}^p_+$ such that $\hat{F}_{X,n}(x) > 0$. Then $RB(\hat{\varphi}_n(x), (X,Y)^n) = 1/n$.*

Proof Since $\hat{\varphi}_n(x) = T^{1,x}((X,Y)^n) := \max_{i|X_i \le x} Y_i$, there exists $j \in \{1, \ldots, n\}$ such that $X_j \le x$ and $Y_j = \hat{\varphi}_n(x)$. Let Y^* be any arbitrary point such that $Y^* > Y_j$. Then, if we replace the FDH point (X_j, Y_j) in the sample $((X_1,Y_1), \ldots, (X_j,Y_j), \ldots, (X_n,Y_n))$ by (X_j, Y^*), we get the contaminated FDH estimator $T^{1,x}((X_1,Y_1), \ldots, (X_j,Y^*), \ldots, (X_n,Y_n)) = Y^*$. Hence

$$\sup_{(Z)_1^n} |T^{1,x}\{(Z)_1^n\} - T^{1,x}\{(Z)^n\}| \ge |T^{1,x}\{(X_1,Y_1), \ldots, (X_j,Y^*), \ldots, (X_n,Y_n)\} - T^{1,x}\{(Z)^n\}|$$

for all $Y^* > Y_j$. Therefore a breakdown occurs as $Y^* \to \infty$. $\square$

Proof of Theorem 2.1 Let $N_x = n\hat{F}_{X,n}(x)$ be the number of observations (X_i, Y_i) with $X_i \le x$ and let $Y_1^x, \ldots, Y_{N_x}^x$ be the Y_i's such that $X_i \le x$. For $i = 1, \ldots, N_x$, denote by $Y_{(i)}^x$ the ith order statistic of the points $Y_1^x, \ldots, Y_{N_x}^x$. We have $\hat{\varphi}_n(x) = Y_{(N_x)}^x$ and so

$$\hat{\xi}_{m,n}(x) = S^{m,x}((X,Y)^n) := Y_{(N_x)}^x - \int_0^{Y_{(N_x)}^x} [\hat{F}_n(y|x)]^m dy.$$

If $N_x = 1$, $\hat{\xi}_{m,n}(x) = \hat{\varphi}_n(x)$ and so $RB(\hat{\xi}_{m,n}(x), (X,Y)^n) = 1/n$ by Lemma A.1. Otherwise,

$$S^{m,x}((X,Y)^n) = Y^x_{(N_x)} - \sum_{i=1}^{N_x-1} [i/N_x]^m (Y^x_{(i+1)} - Y^x_{(i)}).$$

Consider the same contaminated sample $(X,Y)^n_1 = ((X_1,Y_1),\ldots,(X_j,Y^*),\ldots,(X_n,Y_n))$ used in the proof of Lemma A.1, obtained by replacing the FDH observation (X_j,Y_j) by (X_j,Y^*), where Y^* is an arbitrary point such that $Y^* > Y^x_{(N_x)}$. Then, if $N_x = 2$, we have $S^{m,x}((X,Y)^n_1) = (1-(1/2)^m)Y^* + (1/2)^m Y^x_{(1)}$, and thus a breakdown occurs as $Y^* \to \infty$. Likewise, if $N_x > 2$, we have $S^{m,x}((X,Y)^n_1) = (1 - [\frac{N_x-1}{N_x}]^m)Y^* + [\frac{N_x-1}{N_x}]^m Y^x_{(N_x-1)} - \sum_{i=1}^{N_x-2}[i/N_x]^m(Y^x_{(i+1)} - Y^x_{(i)})$ and thus a breakdown occurs as $Y^* \to \infty$. $\square$

Proof of Theorem 2.2 The quantile $\hat{q}_{\alpha,n}(x) = T^{\alpha,x}((X,Y)^n)$ of the sample $(X,Y)^n$ is given by (2.1). Denote by $\mathbb{N}^*$ the set of all positive integers. In what follows the index j is such that $T^{\alpha,x}((X,Y)^n) = Y^x_{(j)}$, *i.e.*, $j = \alpha N_x$ if $\alpha N_x \in \mathbb{N}^*$ and $j = [\alpha N_x] + 1$ otherwise.

(i) First let us show that $k = N_x - j + 1$ points are sufficient for breakdown of $\hat{q}_{\alpha,n}(x)$: If we replace, among the observations (X_i,Y_i) with $X_i \le x$, the k largest outputs $Y^x_{(j)},\ldots,Y^x_{(N_x)}$ by an arbitrary point $Y^* > Y^x_{(N_x)}$ without replacing their corresponding inputs X_i, then the X_i's of the obtained contaminated sample $(X,Y)^n_k$ such that $X_i \le x$ are the same as those of the initial sample $(X,Y)^n$ and their corresponding ordered Y_i's are $Y^x_{(1)} \le \ldots \le Y^x_{(j-1)} \le Y^* \le \ldots \le Y^*$, where Y^* occurs k times. Hence the αth quantile of $(X,Y)^n_k$, defined as the jth order statistic, is $T^{\alpha,x}((X,Y)^n_k) = Y^*$. Therefore a breakdown occurs as $Y^* \to \infty$.

(ii) Let us now show that $k-1 = N_x - j$ points are not sufficient for breakdown of $\hat{q}_{\alpha,n}(x)$: Let $(X,Y)^n_{k-1} = ((X_1^*,Y_1^*),\ldots,(X_n^*,Y_n^*))$ be a contaminated sample by replacing $k-1$ points of $(X,Y)^n$ with arbitrary values in $\mathbb{R}^p_+ \times \mathbb{R}_+$. Let ℓ_x be the number of replaced points among the observations (X_i,Y_i) with $X_i \le x$. It is clear that $\max\{0,(k-1)-(n-N_x)\} \le \ell_x \le k-1$. Let N_x^* be the number of points (X_i^*,Y_i^*) such that $X_i^* \le x$. Then it is easy to see that

$$N_x^* \le N_x + (k-1) - \ell_x. \tag{A.1}$$

Let $Y_1^{*x},\ldots,Y_{N_x^*}^{*x}$ be the points Y_i^* such that $X_i^* \le x$, and for $i = 1,\ldots,N_x^*$, denote by $Y^{*x}_{(i)}$ the ith order statistic such that $Y^{*x}_{(1)} \le \ldots \le Y^{*x}_{(N_x^*)}$. Then

$$T^{\alpha,x}((X,Y)^n_{k-1}) = \begin{cases} Y^{*x}_{(\alpha N_x^*)} & \text{if} \quad \alpha N_x^* \in \mathbb{N}^* \\ Y^{*x}_{([\alpha N_x^*]+1)} & \text{otherwise.} \end{cases}$$

Because $k-1 = N_x - j$ and $k-1 \ge \ell_x$, we have $N_x - \ell_x \ge j$. Since $\alpha N_x \le j$, we obtain $N_x(1-\alpha) \ge \ell_x$ and so $(2N_x - \alpha N_x - \ell_x)\alpha \le N_x - \ell_x$. Using $\alpha N_x \le j$, we get $(N_x + (k-1) - \ell_x)\alpha = (2N_x - j - \ell_x)\alpha \le N_x - \ell_x$. It follows from (A.1) that

$$\alpha N_x^* \le N_x - \ell_x. \tag{A.2}$$

It is then clear that $Y^{*x}_{(\alpha N^*_x)} \leq Y^{*x}_{(N_x-\ell_x)}$ if $\alpha N^*_x \in \mathbb{N}^*$, otherwise it follows from (A.2) that $[\alpha N^*_x]+1 \leq N_x - \ell_x$, whence $Y^{*x}_{([\alpha N^*_x]+1)} \leq Y^{*x}_{(N_x-\ell_x)}$. Thus

$$0 \leq T^{\alpha,x}((X,Y)^n_{k-1}) \leq Y^{*x}_{(N_x-\ell_x)}. \tag{A.3}$$

Since we only replace ℓ_x points among the N_x observations (X_i, Y_i) with inputs $X_i \leq x$, the remaining $N_x - \ell_x$ non-replaced observations (X_i, Y_i) have outputs $Y^x_i \leq Y^x_{(N_x)}$. Since these non-contaminated $N_x - \ell_x$ outputs Y^x_i are contained in the set $\{Y^{*x}_1, \ldots, Y^{*x}_{N^*_x}\}$, we have $Y^{*x}_{(N_x-\ell_x)} \leq Y^x_{(N_x)}$. Therefore $T^{\alpha,x}((X,Y)^n_{k-1}) \leq Y^x_{(N_x)}$ in view of (A.3). Thus $|T^{\alpha,x}((X,Y)^n_{k-1}) - T^{\alpha,x}((X,Y)^n)| \leq \hat{\varphi}_n(x)$ for any $(X,Y)^n_{k-1}$. □

Proof of Proposition 2.1 Let $(X,Y)^{n,y}_{k^*-1} = ((X_1, Y^*_1), \ldots, (X_n, Y^*_n))$ be an arbitrary contaminated sample. Using the notations of the proof of Theorem 2.2, we have here $N^*_x = N_x$, $Y^{*x}_{(1)} \leq \cdots \leq Y^{*x}_{(N_x-\ell_x)}$ are the $N_x - \ell_x$ non-contaminated Y^x_i's and $Y^{*x}_{(N_x-\ell_x+1)} \leq \cdots \leq Y^{*x}_{(N_x)}$ are the resulting ℓ_x outliers in the direction of Y. Since the points $Y^{*x}_{(1)} \leq \cdots \leq Y^{*x}_{(j)}$ belong to the set of non-contaminated Y^x_i's, the point $Y^{*x}_{(j)}$ is then larger than or equal to j points among these non-contaminated Y^x_i's. Therefore $Y^x_{(j)} \leq Y^{*x}_{(j)}$. On the other hand, we have $T^{\alpha,x}((X,Y)^n) = Y^x_{(j)}$ and $T^{\alpha,x}((X,Y)^{n,y}_{k^*-1}) = Y^{*x}_{(j)}$ since $\alpha N^*_x = \alpha N_x$. Thus $T^{\alpha,x}((X,Y)^n) \leq T^{\alpha,x}((X,Y)^{n,y}_{k^*-1})$. The second inequality $T^{\alpha,x}((X,Y)^{n,y}_{k^*-1}) \leq Y^x_{(N_x)} = \hat{\varphi}_n(x)$ is established in the proof of Theorem 2.2. □

Proof of Proposition 2.2 The result is immediate since by definition of the median we have $\tilde{\xi}_m(x) = \inf\{y \geq 0 | F^m(y|x) \geq 1/2\} = q_{(1/2)^{1/m}}(x)$. □

B Asymptotics

Fix $m \geq 1$ and $x \in \mathbb{R}^p_+$ such that $F_X(x) > 0$. Define the domain $\mathbb{D}_x$ to be the set of distribution functions $G(\cdot,\cdot)$ on $\mathbb{R}^p_+ \times \mathbb{R}_+$ such that

$$G(x,\infty) > 0 \quad \text{and} \quad G^{-1}(1|x) \leq \varphi(x) \tag{B.1}$$

where $G^{-1}(1|x) := \inf\{y \geq 0 |\ G(y|x) = 1\}$ stands for the upper boundary of the support of the conditional distribution function $G(\cdot|x) = G(x,\cdot)/G(x,\infty)$. For any $G \in \mathbb{D}_x$ define

$$\overset{m,x}{\phi}(G) = \int_0^\infty [1 - G^m(y|x)]dy$$

where the integrand is identically zero for $y \geq G^{-1}(1|x)$. It follows from (B.1) that $\overset{m,x}{\phi}(G) = \int_0^{\varphi(x)} [1 - G^m(y|x)]dy$ for all $G \in \mathbb{D}_x$. In particular, we have $\overset{m,x}{\phi}(F) = \xi_m(x)$ and $\overset{m,x}{\phi}(\hat{F}) = \int_0^{\hat{\varphi}_n(x)} (1 - [\hat{F}_n(y|x)]^m)dy = \hat{\xi}_{m,n}(x) \overset{a.s.}{=} \int_0^{\varphi(x)} (1 - [\hat{F}_n(y|x)]^m)dy$ since $\hat{\varphi}_n(x) \leq \varphi(x)$ with probability 1. The following lemma will be useful for the proof of Proposition 4.1(i).

Lemma B.1. *The map* $\overset{m,x}{\phi}: \mathbb{D}_x \subset L^\infty(\bar{\mathbb{R}}^{p+1}) \longrightarrow [0, \varphi(x)]$ *is Hadamard-differentiable at* F *with derivative*

$$(\overset{m,x}{\phi})'_F : h \in L^\infty(\bar{\mathbb{R}}^{p+1}) \longmapsto (\overset{m,x}{\phi})'_F(h) = \frac{m}{F_X(x)} \int_0^{\varphi(x)} F^{m-1}(y|x)[h(x,\infty)F(y|x) - h(x,y)]dy.$$

Proof Let $h \in L^\infty(\bar{\mathbb{R}}^{p+1})$ and $h_t \to h$ uniformly in $L^\infty(\bar{\mathbb{R}}^{p+1})$, where $F + th_t \in \mathbb{D}_x$ for all small $t > 0$. Write $\xi_{mt}(x) := \overset{m,x}{\phi}(F + th_t)$. Following the definition of the Hadamard differentiability (see van der Vaart (1998), p.296), we shall show that $(\xi_{mt}(x) - \xi_m(x))/t$ converges to $(\overset{m,x}{\phi})'_F(h)$ as $t \downarrow 0$. We have

$$\xi_{mt}(x) - \xi_m(x) = \int_0^{\varphi(x)} \left([F(y|x)]^m - \left[\frac{F(x,y) + th_t(x,y)}{F_X(x) + th_t(x,\infty)} \right]^m \right) dy.$$

By Taylor's formula, for any $y \in [0, \varphi(x)]$, there exists a point $\zeta_{t,x}(y)$ interior to the interval joining $F(y|x)$ and $(F(x,y) + th_t(x,y))/(F_X(x) + th_t(x,\infty))$ such that

$$[F(y|x)]^m - \left[\frac{F(x,y) + th_t(x,y)}{F_X(x) + th_t(x,\infty)} \right]^m = mt\zeta_{t,x}^{m-1}(y) \left(\frac{h_t(x,\infty)F(y|x) - h_t(x,y)}{F_X(x) + th_t(x,\infty)} \right).$$

Whence

$$\frac{\xi_{mt}(x) - \xi_m(x)}{t} = \frac{m}{F_X(x) + th_t(x,\infty)} \int_0^{\varphi(x)} \zeta_{t,x}^{m-1}(y)[h_t(x,\infty)F(y|x) - h_t(x,y)]dy. \quad \text{(B.2)}$$

It follows from the definition of $\zeta_{t,x}(y)$ and the uniform convergence $h_t \to h$ in $L^\infty(\bar{\mathbb{R}}^{p+1})$ that $\zeta_{t,x}^{m-1}(y)[h_t(x,\infty)F(y|x) - h_t(x,y)]$ converges to $F^{m-1}(y|x)[h(x,\infty)F(y|x) - h(x,y)]$ uniformly in y as $t \downarrow 0$. Therefore, we obtain $\lim_{t\downarrow 0}(\xi_{mt}(x) - \xi_m(x))/t = (\overset{m,x}{\phi})'_F(h)$. □

Proof of Proposition 4.1(i) It is well known that the empirical process $\sqrt{n}(\hat{F} - F)$ converges in distribution in $L^\infty(\overline{\mathbb{R}}^{p+1})$ to $\mathbb{F}$, a $p+1$ dimensional F-Brownian bridge (see van der Vaart and Wellner 1996, p.82). $\mathbb{F}$ is a Gaussian process with zero mean and covariance function $E(\mathbb{F}(t_1)\mathbb{F}(t_2)) = F(t_1 \wedge t_2) - F(t_1)F(t_2)$, for all $t_1, t_2 \in \overline{\mathbb{R}}^{p+1}$. Then, by applying the functional delta method (see *e.g.* van der Vaart 1998, Theorem 20.8, p.297) in conjunction with Lemma B.1, we obtain $\sqrt{n}(\overset{m,x}{\phi}(\hat{F}) - \overset{m,x}{\phi}(F)) = (\overset{m,x}{\phi})'_F(\sqrt{n}(\hat{F} - F)) + o_p(1)$. □

Let us now consider $\sqrt{n}(\hat{\xi}_{m,n}(x) - \xi_m(x))$ as a process indexed by $x \in \mathcal{X}$, an arbitrarily fixed set such that $\inf_{x\in\mathcal{X}} F_X(x) > 0$. Here $m \geq 1$ is still fixed. Define the domain $\mathbb{D}_\mathcal{X}$ to be the set of distribution functions G on $\mathbb{R}_+^{p+1}$ such that $G \in \mathbb{D}_x$ for all $x \in \mathcal{X}$. Let ν be the finite upper boundary of the support of Y and define, for any $G \in \mathbb{D}_\mathcal{X}$, the map $\overset{m}{\phi}(G) : x \mapsto \overset{m,x}{\phi}(G)$ as a map $\mathcal{X} \longrightarrow [0, \nu]$. Finally, define the functional $\overset{m}{\phi}: G \mapsto \overset{m}{\phi}(G)$ as a map $\mathbb{D}_\mathcal{X} \subset L^\infty(\bar{\mathbb{R}}^{p+1}) \to L^\infty(\mathcal{X})$. We have $\overset{m}{\phi}(\hat{F}) := \{\overset{m,x}{\phi}(\hat{F}); x \in \mathcal{X}\} = \{\hat{\xi}_{m,n}(x); x \in \mathcal{X}\} \overset{a.s.}{=}$

$\{\int_0^{\varphi(x)}(1-[\hat{F}_n(y|x)]^m)dy; x \in \mathcal{X}\}$ since $P[\hat{\varphi}_n(x) \leq \varphi(x), \forall x \in \mathcal{X}] = 1$.The following lemma will be useful for the proof of Proposition 4.1(ii).

Lemma B.2. *$\overset{m}{\phi}$ is Hadamard-differentiable at $F \in \mathbb{D}_{\mathcal{X}}$ with derivative $(\overset{m}{\phi})'_F(h) : x \in \mathcal{X} \mapsto (\overset{m,x}{\phi})'_F(h)$, for any $h \in L^\infty(\bar{\mathbb{R}}^{p+1})$.*

Proof It suffices to make the proof of Lemma B.1 uniform in $x \in \mathcal{X}$. We use the same notation: let $h_t \to h$ in $L^\infty(\bar{\mathbb{R}}^{p+1})$, where $F+th_t$ is contained in $D_{\mathcal{X}}$ for all small t. Abbreviate $\overset{m,x}{\phi}(F+th_t)$ to $\xi_{mt}(x)$. By the uniform convergence of h_t and the definition of $\zeta_{t,x}(y)$, we have $\inf_{x\in\mathcal{X}}|F_X(x)+th_t(x,\infty)| \to \inf_{x\in\mathcal{X}} F_X(x)$ and $\sup_{x\in\mathcal{X},y\in\bar{\mathbb{R}}}|\zeta_{t,x}^{m-1}(y) - F^{m-1}(y|x)| \to 0$ as $t \downarrow 0$. By using $\sup_{x\in\mathcal{X},y\in\bar{\mathbb{R}}}|\zeta_{t,x}(y)| \leq 1$ and $\sup_{x\in\mathcal{X}}\varphi(x) \leq \nu$, it can be easily seen that $\sup_{x\in\mathcal{X}}|(\xi_{mt}(x)-\xi_m(x))/t - (\overset{m,x}{\phi})'_F(h)| \to 0$ as $t \downarrow 0$, which ends the proof. □

Proof of Proposition 4.1(ii) By applying the functional delta method in conjunction with Lemma B.2, it is immediate that $\sqrt{n}(\overset{m}{\phi}(\hat{F}) - \overset{m}{\phi}(F))$ converges in distribution in $L^\infty(\mathcal{X})$ to the linear transformation $\mathbb{G}_m = (\overset{m}{\phi})'_F(\mathbb{F})$ of the Gaussian process $\mathbb{F}$. Furthermore, the linear operator $(\overset{m}{\phi})'_F(\cdot)$ is defined and continuous on the whole space $L^\infty(\bar{\mathbb{R}}^{p+1})$ since

$$||(\overset{m}{\phi})'_F(h)||_{L^\infty(\mathcal{X})} = \sup_{x\in\mathcal{X}}|(\overset{m,x}{\phi})'_F(h)| \leq \frac{2m\nu}{\inf_{x\in\mathcal{X}} F_X(x)}||h||_{L^\infty(\bar{\mathbb{R}}^{p+1})}$$

for any $h \in L^\infty(\bar{\mathbb{R}}^{p+1})$. Therefore $\sqrt{n}(\overset{m}{\phi}(\hat{F}) - \overset{m}{\phi}(F)) = (\overset{m}{\phi})'_F(\sqrt{n}(\hat{F}-F)) + o_p(1)$ by Theorem 20.8 in van der Vaart (1998, p.297). □

Proof of Theorem 4.1 Write $R_{m,n}(x) := \sqrt{n}(\hat{\xi}_{m,n}(x) - \xi_m(x)) - \sqrt{n}\Phi_{m,n}(x)$. By Taylor's formula, for any $y \in [0, \varphi(x)]$, there exists a point $\eta_{x,n}(y)$ interior to the interval joining $F(y|x)$ and $\hat{F}_n(y|x)$ such that $[\hat{F}_n(y|x)]^m - F^m(y|x) = mF^{m-1}(y|x)[\hat{F}_n(y|x) - F(y|x)] + (m/2)(m-1)[\eta_{x,n}(y)]^{m-2}[\hat{F}_n(y|x) - F(y|x)]^2$. By using the fact that $\hat{\xi}_{m,n}(x) - \xi_m(x) \overset{a.s.}{=} \int_0^{\varphi(x)}(F^m(y|x) - [\hat{F}_n(y|x)]^m)dy$, we get

$$\begin{aligned}(\hat{\xi}_{m,n}(x) - \xi_m(x)) - m\int_0^{\varphi(x)} F^{m-1}(y|x)[F(y|x) - \hat{F}_n(y|x)]dy \qquad \text{(B.3)}\\ \overset{a.s.}{=} -(m/2)(m-1)\int_0^{\varphi(x)}[\eta_{x,n}(y)]^{m-2}[\hat{F}_n(y|x) - F(y|x)]^2 dy.\end{aligned}$$

On the other hand, we have by the law of the iterated logarithm (LIL) for empirical processes

$$\sup_x|\hat{F}_{X,n}(x) - F_X(x)| = O\left(\frac{\log\log n}{n}\right)^{1/2}, \sup_{(x,y)}|\hat{F}(x,y) - F(x,y)| = O\left(\frac{\log\log n}{n}\right)^{1/2} \quad \text{(B.4)}$$

with probability 1. It follows that $\sup_y|\hat{F}_n(y|x) - F(y|x)| = O\left((\log\log n/n)^{1/2}\right)$ with probability 1, whence $\sup_y\{\sqrt{n}[\hat{F}_n(y|x) - F(y|x)]^2\} \xrightarrow{a.s.} 0$ as $n \to \infty$. Finally, since $0 \leq \eta_{x,n}(y) \leq 1$

for all y, we arrive at $\sqrt{n}\{(\hat{\xi}_{m,n}(x)-\xi_m(x))-m\int_0^{\varphi(x)} F^{m-1}(y|x)[F(y|x)-\hat{F}_n(y|x)]dy\} \xrightarrow{a.s.} 0$. This gives $R_{m,n}(x) \xrightarrow{a.s.} 0$ since $\hat{F}_{X,n}(x)/F_X(x) \xrightarrow{a.s.} 1$. By applying again the classical LIL (see *e.g.* Serfling 1980, Theorem A, p.35), we obtain for either choice of sign

$$\limsup_{n\to\infty} \pm\frac{\sqrt{n}\Phi_{m,n}(x)}{(2\log\log n)^{1/2}} = \limsup_{n\to\infty} \pm\frac{(m/F_X(x))}{(2n\log\log n)^{1/2}} \sum_{i=1}^{n} \int_0^{\varphi(x)} F^{m-1}(y|x)\,[\mathbb{1}(X_i \le x)F(y|x) - \mathbb{1}(X_i \le x, Y_i \le y)]\,dy = \sigma(x,m)$$

with probability 1. Moreover $R_{m,n}(x)/(2\log\log n)^{1/2} \xrightarrow{a.s.} 0$ as $n \to \infty$. Thus, by combining these results, we get the desired LIL. $\square$

The following lemma will be needed to prove Theorem 4.2.

Lemma B.3. *Assume that the condition of Theorem 4.2 hold. For any $\alpha \in]\alpha_1, \alpha_2[$ and any $c \in \mathbb{R}$, let $\alpha_n = \alpha + c/\sqrt{n\hat{F}_{X,n}(x)} + o(1/\sqrt{n})$. Then*

$$\hat{q}_{\alpha_n,n}(x) \xrightarrow{a.s.} q_\alpha(x) \quad \textit{and} \quad \sqrt{n}(\hat{q}_{\alpha_n,n}(x) - \hat{q}_{\alpha,n}(x)) \xrightarrow{p} c/\sqrt{F_X(x)}f(q_\alpha(x)|x) \quad \textit{as} \quad n \to \infty.$$

Proof Following Serfling (1980, p.6), an equivalent condition for the convergence $Z_n \xrightarrow{a.s.} Z$ to hold is $\lim_{n\to\infty} \mathbb{P}(\sup_{m\ge n} |Z_m - Z| > \varepsilon) = 0$ for every $\varepsilon > 0$, where $Z_1, Z_2, \cdots$ and Z are random variables on $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\varepsilon > 0$. By the smoothness of $F(\cdot|x)$ at $q_\alpha(x)$ we have $F(q_\alpha(x) - \varepsilon|x) < \alpha < F(q_\alpha(x) + \varepsilon|x)$. Since $\alpha_n \xrightarrow{a.s.} \alpha$, we then have by applying the equivalent condition for the almost sure convergence

$$\mathbb{P}[\alpha_m < \alpha + \frac{F(q_\alpha(x)+\varepsilon|x) - \alpha}{2},\ \forall m \ge n] \to 1,$$
$$\mathbb{P}[\alpha - \frac{\alpha - F(q_\alpha(x)-\varepsilon|x)}{2} < \alpha_m,\ \forall m \ge n] \to 1 \quad \text{as} \quad n \to \infty.$$

On the other hand, since $\hat{F}_n(q_\alpha(x) \pm \varepsilon|x) \xrightarrow{a.s.} F(q_\alpha(x) \pm \varepsilon|x)$, we have

$$\mathbb{P}[\alpha + \frac{F(q_\alpha(x)+\varepsilon|x) - \alpha}{2} < \hat{F}_m(q_\alpha(x)+\varepsilon|x),\ \forall m \ge n] \to 1,$$
$$\mathbb{P}[\hat{F}_m(q_\alpha(x)-\varepsilon|x) < \alpha - \frac{\alpha - F(q_\alpha(x)-\varepsilon|x)}{2},\ \forall m \ge n] \to 1 \quad \text{as} \quad n \to \infty.$$

It follows $\mathbb{P}[\alpha_m < \hat{F}_m(q_\alpha(x)+\varepsilon|x),\ \forall m \ge n] \to 1$ and $\mathbb{P}[\hat{F}_m(q_\alpha(x)-\varepsilon|x) < \alpha_m,\ \forall m \ge n] \to 1$ as $n \to \infty$. Whence $\mathbb{P}[\hat{F}_m(q_\alpha(x)-\varepsilon|x) < \alpha_m < \hat{F}_m(q_\alpha(x)+\varepsilon|x),\ \forall m \ge n] \to 1$ as $n \to \infty$. Thus, by applying the fundamental property that the event $\{\hat{F}_m(y|x) \ge \alpha_m\}$ is equivalent to $\{y \ge \hat{q}_{\alpha_m,m}(x)\}$, we get $\mathbb{P}[q_\alpha(x) - \varepsilon < \hat{q}_{\alpha_m,m}(x) \le q_\alpha(x) + \varepsilon,\ \forall m \ge n] \to 1$ as $n \to \infty$. Therefore $\mathbb{P}[|\hat{q}_{\alpha_m,m}(x) - q_\alpha(x)| \le \varepsilon,\ \forall m \ge n] \to 1$, which is equivalent to $\hat{q}_{\alpha_n,n}(x) \xrightarrow{a.s.} q_\alpha(x)$.

Let us now turn to the second result. Since $\hat{q}_{\alpha,n}(x)$ and $\hat{q}_{\alpha_n,n}(x) \xrightarrow{a.s.} q_\alpha(x)$ and $\alpha_n \xrightarrow{a.s.} \alpha$, the interval $[a,b]$ contains both $\hat{q}_{\alpha,n}(x)$ and $\hat{q}_{\alpha_n,n}(x)$ and the interval $[\alpha_1,\alpha_2]$ contains α_n, for n sufficiently large, with probability 1. Hence we have almost surely, for n large enough,

$$\sqrt{n}(\hat{q}_{\alpha_n,n}(x) - \hat{q}_{\alpha,n}(x)) = \sqrt{n}\{F(\hat{q}_{\alpha_n,n}(x)|x) - F(\hat{q}_{\alpha,n}(x)|x)\}/f(q_{\delta_n}(x)|x)$$

where $\min\{F(\hat{q}_{\alpha_n,n}(x)|x), F(\hat{q}_{\alpha,n}(x)|x)\} < \delta_n < \max\{F(\hat{q}_{\alpha_n,n}(x)|x), F(\hat{q}_{\alpha,n}(x)|x)\}$. Define the random function $g_n : L^\infty([\alpha_1,\alpha_2]) \mapsto \mathbb{R}$ by $g_n(z) = z(\alpha_n) - z(\alpha)$. Putting $z_n(\cdot) = \sqrt{n}\{F(\hat{q}_{\cdot,n}(x)|x) - F(q_\cdot(x)|x)\}$, we obtain with probability 1, for all n large enough,

$$\sqrt{n}(\hat{q}_{\alpha_n,n}(x) - \hat{q}_{\alpha,n}(x)) = [g_n(z_n) - \sqrt{n}(\alpha - \alpha_n)]/f(q_{\delta_n}(x)|x). \tag{B.5}$$

Let us show that z_n converges in distribution in $L^\infty([\alpha_1,\alpha_2])$ to a process z with continuous paths at α: let $\mathbb{D}_1$ be the set of all restrictions of distribution functions on $\mathbb{R}$ to $[a,b]$, and for any $G \in \mathbb{D}_1$, let $G^{-1} :]0,1[\longrightarrow \mathbb{R}$ denotes the generalized inverse map $\alpha \mapsto G^{-1}(\alpha) := \inf\{y|G(y) \geq \alpha\}$. Then by Lemma 3.3 in Daouia (2005), the inverse map $\phi_1 : G \mapsto G^{-1}$ as a map $\mathbb{D}_1 \subset D([a,b]) \longrightarrow L^\infty([\alpha_1,\alpha_2])$ is Hadamard differentiable at $F(\cdot|x)$ tangentially to $C([a,b])$ with derivative $\phi'_{1,F(\cdot|x)} : h \longmapsto -h(F^{-1}(\cdot|x))/f(F^{-1}(\cdot|x)|x)$. We also have

$$z_n = \sqrt{n}\{F(\phi_1(\hat{F}_n(\cdot|x))|x) - F(\phi_1(F(\cdot|x))|x)\} = \sqrt{n}\{\phi_2 \circ \phi_1(\hat{F}_n(\cdot|x)) - \phi_2 \circ \phi_1(F(\cdot|x))\}, \tag{B.6}$$

where $\phi_2 : G^{-1} \longmapsto F(\cdot|x) \circ G^{-1}$. Let us show that ϕ_2 as a map $\phi_1(\mathbb{D}_1) \subset L^\infty([\alpha_1,\alpha_2]) \longrightarrow L^\infty([\alpha_1,\alpha_2])$ is Hadamard differentiable at $\phi_1(F(\cdot|x)) = F^{-1}(\cdot|x) = q_\cdot(x)$ tangentially to $\phi'_{1,F(\cdot|x)}(C([a,b]))$. Let $H = \phi'_{1,F(\cdot|x)}(h)$ with $h \in C([a,b])$ and take an arbitrary converging path $H_t \to H$ in $L^\infty([\alpha_1,\alpha_2])$ such that $F^{-1}(\cdot|x) + tH_t \in \phi_1(\mathbb{D}_1)$ for all small $t > 0$. By the smoothness of $F(\cdot|x)$, it can be easily seen that

$$[F(F^{-1}(\beta|x) + tH_t(\beta)|x) - F(F^{-1}(\beta|x)|x)]/t \longrightarrow H(\beta)f(F^{-1}(\beta|x)|x) \quad \text{as} \quad t \to 0$$

uniformly in $\beta \in [\alpha_1,\alpha_2]$. Then ϕ_2 is Hadamard differentiable at $\phi_1(F(\cdot|x))$ with derivative $\phi'_{2,\phi_1(F(\cdot|x))} : H \longmapsto H \times f(q_\cdot(x)|x) = -h(q_\cdot(x))$. Hence by the chain rule (see van der Vaart 1998, Theorem 20.9, p.298), we have $\phi_2 \circ \phi_1 : \mathbb{D}_1 \longrightarrow L^\infty([\alpha_1,\alpha_2])$ is Hadamard differentiable at $F(\cdot|x)$ tangentially to $C([a,b])$ with derivative $(\phi_2 \circ \phi_1)'_{F(\cdot|x)} = \phi'_{2,\phi_1(F(\cdot|x))} \circ \phi'_{1,F(\cdot|x)}$. With this result and the representation (B.6) of z_n, we can apply immediately the functional delta method (van der Vaart 1998, Theorem 20.8, p.297) in conjunction with Theorem 3.1 in Daouia (2005) to obtain the convergence in distribution of z_n in $L^\infty([\alpha_1,\alpha_2])$ to

$$z = (\phi_2 \circ \phi_1)'_{F(\cdot|x)}\left(W \circ F(\cdot|x)/\sqrt{F_X(x)}\right) = -W/\sqrt{F_X(x)}$$

where $W(\cdot)$ denotes the standard Brownian bridge. Moreover the process z has continuous paths. Since $g_n(z_n) \xrightarrow{d} 0$ whenever $z_n \xrightarrow{d} z$ in $L^\infty([\alpha_1,\alpha_2])$ for a process z with continuous

paths at α (see van der Vaart 1998, Proof of Lemma 21.7, p.308), we conclude that $g_n(z_n)$ in (B.5) converges in distribution to 0. On the other hand, by the smoothness of $F(\cdot|x)$, we have $\delta_n \xrightarrow{a.s.} \alpha$ and $f(q_{\delta_n}(x)|x) \xrightarrow{a.s.} f(q_\alpha(x)|x)$. Finally, since $\sqrt{n}(\alpha - \alpha_n) \xrightarrow{a.s.} -c/\sqrt{F_X(x)}$, we get $\sqrt{n}(\hat{q}_{\alpha_n,n}(x) - \hat{q}_{\alpha,n}(x)) \xrightarrow{p} c/\sqrt{F_X(x)} f(q_\alpha(x)|x)$. $\square$

Proof of Theorem 4.2 Write $\sqrt{n}\{\hat{q}_{\alpha_{n1},n}(x) - (\hat{q}_{\alpha,n}(x) - z\sigma(\alpha,x)/\sqrt{n})\} = \sqrt{n}(\hat{q}_{\alpha_{n1},n}(x) - \hat{q}_{\alpha,n}(x)) + z\sigma(\alpha,x)$. It follows from Lemma B.3 that

$$\sqrt{n}\{\hat{q}_{\alpha_{n1},n}(x) - (\hat{q}_{\alpha,n}(x) - z\sigma(\alpha,x)/\sqrt{n})\} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty. \tag{B.7}$$

Likewise $\sqrt{n}\{\hat{q}_{\alpha_{n2},n}(x) - (\hat{q}_{\alpha,n}(x) + z\sigma(\alpha,x)/\sqrt{n})\} \xrightarrow{p} 0$ as $n \to \infty$. Hence $\sqrt{n}\{(\hat{q}_{\alpha_{n2},n}(x) - \hat{q}_{\alpha_{n1},n}(x)) - 2z\sigma(\alpha,x)/\sqrt{n}\} \xrightarrow{p} 0$ as $n \to \infty$. On the other hand, we have

$$\mathbb{P}[q_\alpha(x) \in C_n] = 1 - \{\mathbb{P}(q_\alpha(x) \le \hat{q}_{\alpha_{n1},n}(x)) + \mathbb{P}(q_\alpha(x) \ge \hat{q}_{\alpha_{n2},n}(x))\}.$$

By using (B.7), we obtain $\mathbb{P}(q_\alpha(x) \le \hat{q}_{\alpha_{n1},n}(x)) = \mathbb{P}\{\sqrt{n}(\hat{q}_{\alpha,n}(x) - q_\alpha(x)) + o_p(1) \ge z\sigma(\alpha,x)\}$. By the asymptotic normality, we have $\lim_{n\to\infty} \mathbb{P}(q_\alpha(x) \le \hat{q}_{\alpha_{n1},n}(x)) = 1 - \Phi(z)$. Likewise $\lim_{n\to\infty} \mathbb{P}(q_\alpha(x) \ge \hat{q}_{\alpha_{n2},n}(x)) = 1 - \Phi(z)$. Therefore $\lim_{n\to\infty} \mathbb{P}[q_\alpha(x) \in C_n] = 2\Phi(z) - 1$. $\square$

Proof of Proposition 4.2 (i) Let $\sigma_n = \sigma(\alpha_n, x)/\sqrt{n} = \sqrt{\alpha_n(1-\alpha_n)}/f(q_{\alpha_n}(x)|x)\sqrt{nF_X(x)}$. We shall prove for any real $y \in \mathbb{R}$ that $\mathbb{P}[\sigma_n^{-1}(\hat{q}_{\alpha_n,n}(x) - q_{\alpha_n}(x)) \le y] \to \Phi(y)$ as $n \to \infty$. Let n be large enough so that $\hat{q}_{\alpha_n,n}(x)$ belongs to the left neighborhood of $\varphi(x)$ on which $F(\cdot|x)$ is differentiable with a strictly positive derivative $f(\cdot|x)$. We have $\mathbb{P}[\sigma_n^{-1}(\hat{q}_{\alpha_n,n}(x) - q_{\alpha_n}(x)) \le y] = \mathbb{P}[\hat{q}_{\alpha_n,n}(x) \le q_{\alpha_n}(x) + \sigma_n y] = \mathbb{P}[\hat{F}(q_{\alpha_n}(x) + \sigma_n y|x) \ge \alpha] = \mathbb{P}[A_n \ge a_n]$, where

$$a_n = \frac{\sqrt{nF_X(x)}}{\sqrt{\alpha_n(1-\alpha_n)}}\{\alpha_n - F(q_{\alpha_n}(x) + \sigma_n y|x)\},$$

$$\begin{aligned} A_n &= \frac{\sqrt{nF_X(x)}}{\sqrt{\alpha_n(1-\alpha_n)}}\{\hat{F}(q_{\alpha_n}(x) + \sigma_n y|x) - F(q_{\alpha_n}(x) + \sigma_n y|x)\} \\ &= \frac{F_X(x)}{\hat{F}_X(x)}\left(\frac{F(q_{\alpha_n}(x) + \sigma_n y|x)[1 - F(q_{\alpha_n}(x) + \sigma_n y|x)]}{\alpha_n(1-\alpha_n)}\right)^{1/2} \sum_{i=1}^n \frac{W_{n,i}}{\sqrt{n}\sigma(W_{n,i})} \end{aligned}$$

with $W_{n,i} = \mathbb{1}(X_i \le x, Y_i \le q_{\alpha_n}(x) + \sigma_n y) - F(q_{\alpha_n}(x) + \sigma_n y|x)\mathbb{1}(X_i \le x)$ and $\sigma^2(W_{n,i}) = F_X(x)F(q_{\alpha_n}(x) + \sigma_n y|x)[1 - F(q_{\alpha_n}(x) + \sigma_n y|x)]$. We first need to prove that $A_n \xrightarrow{d} \mathcal{N}(0,1)$ and second we shall show that $a_n \to -y$ as $n \to \infty$. It is easy to see from (4.4) that $q_{\alpha_n}(x) = \varphi(x) - \left(\frac{1-\alpha_n}{\ell(x)}\right)^{1/\rho_x}$ for n large enough. Likewise since $f(y|x) = \rho_x \ell(x)\{\varphi(x) - y\}^{\rho_x - 1}$ as $y \uparrow \varphi(x)$, we get $f(q_{\alpha_n}(x)|x) = \rho_x \ell(x)\{\varphi(x) - q_{\alpha_n}(x)\}^{\rho_x - 1} = \rho_x \ell(x)^{1/\rho_x}(1-\alpha_n)^{(\rho_x - 1)/\rho_x}$ for n large enough. Then $\sigma_n/(\varphi(x) - q_{\alpha_n}(x)) = \sqrt{\alpha_n}/\rho_x\sqrt{n(1-\alpha_n)F_X(x)} \to 0$ since $n(1-\alpha_n) \to$

∞. It follows that $[1-F(q_{\alpha_n}(x)+\sigma_n y|x)]/[1-F(q_{\alpha_n}(x)|x)] = 1-\sigma_n y/(\varphi(x)-q_{\alpha_n}(x)) \to 1$. Therefore $F(q_{\alpha_n}(x)+\sigma_n y|x)[1-F(q_{\alpha_n}(x)+\sigma_n y|x)] \sim \alpha_n(1-\alpha_n)$ as $n\to\infty$. We also have $F_X(x)/\hat{F}_X(x) \stackrel{a.s.}{\to} 1$. Hence to check that $A_n \stackrel{d}{\to} \mathcal{N}(0,1)$, it is enough to show according Loève's criterion (1963, p.295) that $\lim_{n\to\infty} n\int_{|z|\geq\varepsilon} z^2 dF_{n,1}(z) = 0$ for all $\varepsilon>0$, where $F_{n,1}$ is the common distribution function of the random variables $W_{n,i}/\sqrt{n}\sigma(W_{n,1})$. We have

$$\varepsilon\int_{|z|\geq\varepsilon} z^2 dF_{n,1}(z) \leq \int_{\mathbb{R}} |z|^3 \mathbb{1}(|z|\geq\varepsilon) dF_{n,1}(z) = \mathbb{E}\left[\left|\frac{W_{n,i}}{\sqrt{n}\sigma(W_{n,1})}\right|^3 \mathbb{1}\left(\left|\frac{W_{n,i}}{\sqrt{n}\sigma(W_{n,1})}\right|\geq\varepsilon\right)\right]$$
$$\leq \frac{\mathbb{P}[|W_{n,1}|\geq\varepsilon\sqrt{n}\sigma(W_{n,1})]}{\{\sqrt{n}\sigma(W_{n,1})\}^3} \leq 1/n\varepsilon^2\{\sqrt{n}\sigma(W_{n,1})\}^3$$

by Chebyshev's inequality. Since $\sigma^2(W_{n,1}) \sim \alpha_n(1-\alpha_n)F_X(x)$ and $n(1-\alpha_n)\to\infty$, we get $n\int_{|z|\geq\varepsilon} z^2 dF_{n,1}(z) \to 0$ and so $\sum_{i=1}^n \frac{W_{n,i}}{\sqrt{n}\sigma(W_{n,i})} \stackrel{d}{\to} \mathcal{N}(0,1)$. Whence $A_n \stackrel{d}{\to} \mathcal{N}(0,1)$. Therefore the monotone function $S_n(\cdot) = \mathbb{P}[A_n \geq \cdot]$ converges pointwise to $1-\Phi(\cdot)$ which is continuous. By Dini's Theorem, S_n also converges uniformly to $1-\Phi$. Finally it suffices to show that $a_n \to -y$ to conclude that $\mathbb{P}[A_n \geq a_n] \to \Phi(y)$. First we have $a_n = -y\sigma_n\sqrt{nF_X(x)}f(\delta_n|x)/\sqrt{\alpha_n(1-\alpha_n)} = -yf(\delta_n|x)/f(q_{\alpha_n}(x)|x)$ for a real δ_n lying between $q_{\alpha_n}(x)$ and $q_{\alpha_n}(x)+\sigma_n y$. Second, since $\frac{f(\delta_n|x)}{f(q_{\alpha_n}(x)|x)} = \left\{1+\frac{q_{\alpha_n}(x)-\delta_n}{\varphi(x)-q_{\alpha_n}(x)}\right\}^{\rho_x-1}$ for all n large enough, and $\left|\frac{q_{\alpha_n}(x)-\delta_n}{\varphi(x)-q_{\alpha_n}(x)}\right| \leq \frac{|y|\sigma_n}{\varphi(x)-q_{\alpha_n}(x)} \to 0$, we get $f(\delta_n|x)/f(q_{\alpha_n}(x)|x)\to 1$ and $a_n\to -y$.

(ii) We know by the proof of Theorem 4.1 (see Equation (B.3)) that

$$\sqrt{n}(\hat{\xi}_{m,n}(x)-\xi_m(x)) \stackrel{a.s.}{=} (F_X(x)/\hat{F}_{X,n}(x))\sqrt{n}\Phi_{m,n}(x)$$
$$-\sqrt{n}(m/2)(m-1)\int_0^{\varphi(x)}[\eta_{x,n}(y)]^{m-2}[\hat{F}_n(y|x)-F(y|x)]^2 dy$$

and that $\sup_y|\hat{F}_n(y|x)-F(y|x)| \stackrel{a.s.}{=} O\left((\log\log n/n)^{1/2}\right)$ in view of (B.4). For $y\in]0,\varphi(x)[$ we have $0<\eta_{x,n}(y)<1$ and $[\eta_{x,n}(y)]^{m(n)-2} \stackrel{a.s.}{\to} 0$ when $n\to\infty$, so using the dominated convergence theorem we get $\int_0^{\varphi(x)}[\eta_{x,n}(y)]^{m-2}dy \stackrel{a.s.}{\to} 0$. Since $\sqrt{n}m(m-1)/\sigma(x,m) = O(n/\log\log n)$, we obtain

$$\frac{\sqrt{n}}{\sigma(x,m)}(m/2)(m-1)\int_0^{\varphi(x)}[\eta_{x,n}(y)]^{m-2}[\hat{F}_n(y|x)-F(y|x)]^2 dy$$
$$\stackrel{a.s.}{\leq} \frac{\sqrt{n}}{\sigma(x,m)}m(m-1)O(\log\log n/n)\int_0^{\varphi(x)}[\eta_{x,n}(y)]^{m-2}dy \stackrel{a.s.}{\longrightarrow} 0.$$

On the other hand,

$$\frac{\sqrt{n}\Phi_{m,n}(x)}{\sigma(x,m)} = \sum_{i=1}^n \frac{Z_{n,i}}{\sqrt{n}\sigma(Z_{n,i})}$$

where $Z_{n,i} = (m/F_X(x))\mathbb{1}(X_i\leq x)\int_0^{\varphi(x)} F^{m-1}(y|x)[F(y|x)-\mathbb{1}(Y_i\leq y)]dy$ and its variance $\sigma^2(Z_{n,i}) = \sigma^2(x,m)$. We have $n\mathbb{E}[|Z_{n,1}|^3]/\{n\sigma^2(Z_{n,1})\}^{3/2} \leq m\varphi(x)/F_X(x)\sqrt{n}\sigma(Z_{n,1}) \to$

0 since $m/\sqrt{n}\sigma(x,m) \to 0$. Hence Lyapounov's Theorem gives $\sqrt{n}\sigma^{-1}(x,m)\Phi_{m,n}(x) \xrightarrow{d} \mathcal{N}(0,1)$. Therefore $\sqrt{n}\sigma^{-1}(x,m)(\hat{\xi}_{m,n}(x) - \xi_m(x)) \xrightarrow{d} \mathcal{N}(0,1)$. $\square$

Acknowledgements. The authors thank the Editor, Associate Editor and reviewers for their valuable comments which led to a considerable improvement of the manuscript. This research was supported by the Research Fund KULeuven (GOA/07/04-project) and by the IAP research network P6/03, Federal Science Policy, Belgium. Support from the French "Agence Nationale pour la Recherche" (under grant ANR-08-BLAN-0106-01/EPI project) is also acknowledged.

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., W.W. Cooper and E. Rhodes (1981), Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through, *Management Science*, 27, 668-697.

[4] Daouia, A. (2005), Asymptotic Representation Theory for Nonstandard Conditional Quantiles, *Journal of Nonparametric Statistics*, 17(2), 253–268.

[5] Daouia, A., J-P. Florens and L. Simar (2008), Functional Convergence of Quantile-type Frontiers with Application to Parametric Approximations, *Journal of Statistical Planning and Inference*, 138, 708–725.

[6] Daouia, A., J-P. Florens and L. Simar (2010), Frontier Estimation and Extreme Value Theory, *Bernoulli*, 16(4), 1039–1063.

[7] Daouia, A. and A. Ruiz-Gazen (2006), Robust Nonparametric Frontier Estimators: Influence Function and Qualitative Robustness, *Statistica Sinica*, 16 (4), 1233–1253.

[8] Daouia, A. and L. Simar (2007), Nonparametric efficiency analysis: A multivariate conditional quantile approach, *Journal of Econometrics*, 140, 375–400.

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

[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] Donoho, D. L. and Huber, P. J. (1983). The notion of breakdown point, in A Festschrift for Erich L. Lehmann (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 157–184, Wadsworth, Belmont, CA.

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

[13] Florens, J.P. and L. Simar, (2005), Parametric Approximations of Nonparametric Frontier, *Journal of Econometrics*, 124 (1), 91–116.

[14] Gijbels, I. and Mammen, E. and Park, B. U. and Simar, L. (1999). On estimation of monotone and concave frontier functions. *J. Amer. Statistical Assoc.*, 94(445), 220–228.

[15] Hampel, F. R. (1968). Contributions to the Theory of Robust Estimation. PhD thesis, University of California, Berkeley.

[16] Loève, M. (1963). *Probability Theory*, Third Edition. Princeton: Van Nostrand.

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

[18] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons Inc.

[19] Simar, L. (2003), Detecting Outliers in Frontiers Models: a Simple Approach, *Journal of Productivity Analysis*, 20, 391–424.

[20] 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, H. Fried, C.A.K Lovell and S. Schmidt, editors, Oxford University Press.

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

[22] van der Vaart, A. W. and J.A. Wellner, (1996) *Weak Convergence and Empirical Processes. With Applications to Statistics.* Springer Series in Statistics, Springer-Verlag, New York.

[23] Wilson, P.W. (1993), Detecting outliers in deterministic nonparametric frontier models with multiple outputs, *Journal of Business and Economic Statistics*, 11, 319–323.

[24] Wilson, P.W. (1995), Detecting influential observations in data envelopment analysis, *Journal of Productivity Analysis*, 6, 27–45.

Table 1: *The values* $n \times RB(\tilde{\xi}_{m,n}(x))$ *with* $n = 105$.

x	$m=1$	$m=5$	$m=10$	$m=15$	$m=20$	$m=25$
0.1	5	2	1	1	1	1
0.3	15	4	2	2	1	1
0.5	26	7	4	3	2	2
0.7	37	10	5	4	3	2
0.9	49	13	7	5	4	3

Table 2: 1000 *Monte-Carlo simulations,* $n = 1000$ *(l-h.s) with* 5 *outliers added (r-h.s).*

$n = 1000$

	MSE		Bias	
x	$\hat{\xi}_{10,n}(x)$	$\tilde{\xi}_{10,n}(x)$	$\hat{\xi}_{10,n}(x)$	$\tilde{\xi}_{10,n}(x)$
0.15	0.0001	0.0001	-0.0011	-0.0009
0.35	0.0001	0.0001	-0.0010	-0.0007
0.55	0.0001	0.0001	-0.0002	0
0.75	0.0001	0.0001	-0.0009	-0.0007
0.95	0.0001	0.0001	-0.0008	-0.0007
x	$\hat{\xi}_{15,n}(x)$	$\tilde{\xi}_{15,n}(x)$	$\hat{\xi}_{15,n}(x)$	$\tilde{\xi}_{15,n}(x)$
0.15	0.0001	0.0001	-0.0012	-0.0008
0.35	0.0001	0.0001	-0.0009	-0.0008
0.55	0.0001	0.0001	-0.0011	-0.0009
0.75	0.0001	0.0001	-0.0006	-0.0004
0.95	0	0.0001	-0.0004	0.0002
x	$\hat{\xi}_{20,n}(x)$	$\tilde{\xi}_{20,n}(x)$	$\hat{\xi}_{20,n}(x)$	$\tilde{\xi}_{20,n}(x)$
0.15	0.0001	0.0001	-0.0011	-0.0008
0.35	0.0001	0.0001	-0.0011	-0.0008
0.55	0.0001	0.0001	-0.0004	-0.0001
0.75	0	0.0001	-0.0007	-0.0006
0.95	0.0001	0.0001	-0.0007	-0.0006

$n = 1005$

MSE		Bias	
$\hat{\xi}_{10,n}(x)$	$\tilde{\xi}_{10,n}(x)$	$\hat{\xi}_{10,n}(x)$	$\tilde{\xi}_{10,n}(x)$
0.0002	0.0001	0.0105	0.0023
0.0002	0.0001	0.0103	0.0045
0.0001	0.0001	0.0066	0.0040
0.0001	0.0001	0.0058	0.0021
0.0001	0.0001	0.0022	0.0017
$\hat{\xi}_{15,n}(x)$	$\tilde{\xi}_{15,n}(x)$	$\hat{\xi}_{15,n}(x)$	$\tilde{\xi}_{15,n}(x)$
0.0003	0.0001	0.0147	0.0032
0.0002	0.0001	0.0128	0.0037
0.0001	0.0001	0.0087	0.0045
0.0001	0.0001	0.0076	0.0029
0.0001	0.0001	0.0028	0.0022
$\hat{\xi}_{20,n}(x)$	$\tilde{\xi}_{20,n}(x)$	$\hat{\xi}_{20,n}(x)$	$\tilde{\xi}_{20,n}(x)$
0.0004	0.0001	0.0186	0.0030
0.0003	0.0001	0.0159	0.0047
0.0001	0.0001	0.0099	0.0036
0.0001	0.0001	0.0085	0.0025
0.0001	0.0001	0.0033	0.0031

Table 3: *Average Lengths (avl) and Coverages (cov) of the* 95% *confidence intervals* C_n *and* Q_n*, sample sizes* $n = 100$ *and* $n = 1000$.

$n = 100$ $m = 5$ and $\alpha = .8557$

x	cov_{Q_n}	cov_{C_n}	avl_{Q_n}	avl_{C_n}
0.4	0.9120	0.9540	0.0903	0.1347
0.5	0.9310	0.9510	0.0903	0.1302
0.6	0.9360	0.9500	0.0907	0.1291
0.7	0.9440	0.9560	0.0910	0.1299
0.8	0.9450	0.9470	0.0913	0.1283
0.9	0.9400	0.9540	0.0914	0.1273
1	0.9380	0.9540	0.0913	0.1295

$n = 100$ $m = 10$ and $\alpha = .9242$

x	cov_{Q_n}	cov_{C_n}	avl_{Q_n}	avl_{C_n}
0.65	0.9200	0.9640	0.0889	0.1392
0.70	0.9080	0.9460	0.0883	0.1324
0.75	0.9310	0.9510	0.0888	0.1274
0.80	0.9140	0.9550	0.0878	0.1257
0.85	0.9110	0.9510	0.0881	0.1300
0.90	0.9110	0.9500	0.0886	0.1298
0.95	0.9250	0.9490	0.0891	0.1239

$n = 1000$ $m = 5$ and $\alpha = .8557$

x	cov_{Q_n}	cov_{C_n}	avl_{Q_n}	avl_{C_n}
0.4	0.9540	0.9560	0.0293	0.0401
0.5	0.9470	0.9360	0.0293	0.0397
0.6	0.9570	0.9540	0.0293	0.0402
0.7	0.9490	0.9530	0.0293	0.0400
0.8	0.9450	0.9470	0.0294	0.0400
0.9	0.9360	0.9510	0.0294	0.0401
1	0.9420	0.9640	0.0293	0.0403

$n = 1000$ $m = 10$ and $\alpha = .9242$

x	cov_{Q_n}	cov_{C_n}	avl_{Q_n}	avl_{C_n}
0.65	0.9560	0.9480	0.0290	0.0394
0.70	0.9280	0.9410	0.0291	0.0393
0.75	0.9350	0.9390	0.0291	0.0395
0.80	0.9320	0.9500	0.0290	0.0390
0.85	0.9380	0.9420	0.0290	0.0393
0.90	0.9280	0.9470	0.0290	0.0392
0.95	0.9480	0.9470	0.0290	0.0393