T E C H N I C A L
R E P O R T

11039

On Projection-Type Estimators of Multivariate Isotonic Functions

DAOUIA, A. and B.U. PARK

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

On Projection-Type Estimators of Multivariate Isotonic Functions

Abdelaati Daouia* and Byeong U. Park†

* Institut de Statistique, Université catholique de Louvain, Belgium (daouia@cict.fr)
† Department of Statistics, Seoul National University, South Korea (bupark@stats.snu.ac.kr)

Abstract

Let M be an isotonic real-valued function on a compact subset of $\mathbb{R}^d$ and let $\hat{M}_n$ be an unconstrained estimator of M. A feasible monotonizing technique is to take the largest (smallest) monotone function that lies below (above) the estimator $\hat{M}_n$ or any convex combination of these two envelope estimators. When the process $r_n(\hat{M}_n - M)$ is asymptotically equicontinuous for some sequence $r_n > 0$, we show that these projected estimators are r_n-equivalent in probability to the original unrestricted estimator. Our first motivating application involves a monotone estimator of the conditional distribution function that has the distributional properties of the local linear regression estimator. Applications also include the estimation of econometric (probability-weighted moment, quantile-based) and biometric (mean remaining lifetime) functions.

AMS 2000 subject classification: 62G05; 60E20; 91B38

Key words : isotonic, asymptotic equicontinuity, limit process, conditional distribution function, local linear fitting, frontier modeling.

1 Introduction

Let M be a real-valued function defined on a domain $\mathbb{D} \subset \mathbb{R}^d$ to be estimated from a sample of size n. In many practical applications the function of interest M is believed to be isotonic nondecreasing with respect to the partial order in the sense that $x \leq x'$ componentwise implies $M(x) \leq M(x')$ (the nonincreasing case is similar). It is then natural to try to incorporate this prior information into an estimation procedure. Examples include analysis of monotone regression means (see, *e.g.*, Hall and Huang 2001; Mammen and Yu 2007), estimation of monotone conditional quantiles (see Mukerjee 1993), study of monotone failure rates (see, *e.g.*, Wang 1986), analysis of the mean residual life (see, *e.g.*, Kochar, Mukerjee and Samaniego 2000) and partial frontier modeling (see Daouia and Simar 2005).

Generally, monotonicity is not guaranteed when constructing estimators $\hat{M}_n$ with highly desirable asymptotic properties. In this paper, we consider isotonic target functions M and unconstrained estimators $\hat{M}_n$ for which the process $r_n(\hat{M}_n - M)$ converges weakly at a rate $r_n \to \infty$ as $n \to \infty$. We show that the monotonized versions of $\hat{M}_n$ that we describe below inherit the asymptotic distributional properties of $\hat{M}_n$. The basic idea of the monotonization here simply utilizes the fact that M is nondecreasing if and only if $M(x) = \sup_{x' \le x} M(x') = \inf_{x \le x'} M(x')$, where x' runs over $\mathbb{D}$. This consideration leads to the projected isotonic estimators

$$\hat{M}_n^u(x) = \sup_{x' \le x} \hat{M}_n(x'), \quad \hat{M}_n^\ell(x) = \inf_{x \le x'} \hat{M}_n(x'). \tag{1}$$

Their asymptotic properties may be driven from those of the original estimator $\hat{M}_n$. For example, if $\hat{M}_n$ is uniformly consistent on $\mathbb{D}$, then $\hat{M}_n^u$ and $\hat{M}_n^\ell$ are also uniformly consistent on $\mathbb{D}$. This follows since both $|\hat{M}_n^u(x) - M(x)|$ and $|\hat{M}_n^\ell(x) - M(x)|$ are less than or equal to $\sup_{x' \in \mathbb{D}} |\hat{M}_n(x') - M(x')|$.

Both the envelope estimators are monotone. Any convex combination of $\hat{M}_n^\ell(x)$ and $\hat{M}_n^u(x)$ also yields an isotonic estimator for M. Mukerjee and Stern (1994) favored the hybrid variant

$$\hat{M}_n^\star(x) = (\hat{M}_n^\ell(x) + \hat{M}_n^u(x))/2$$

to isotonize the Nadaraya-Watson kernel estimator of the regression function. They derived the strong uniform consistency of their isotonic estimator and demonstrated via Monte Carlo studies its inexpensiveness and superiority in terms of mean squared error. The same principle was employed by Daouia and Simar (2005) to construct monotonized nonparametric frontier estimators. The resulting frontier functions $\hat{M}_n^\star$ share the robustness and the complete uniform convergence properties of the original estimators $\hat{M}_n$. Hall and Müller (2003) considered the upper envelope $\hat{M}_n^u$ to monotonize an estimator $\hat{M}_n$ of the conditional distribution function obtained by local linear fitting. Likewise, Kochar et *al.* (2000) utilized the upper version $\hat{M}_n^u$ to isotonize the empirical estimator of a biometric function M.

In this paper we show that

$$\sup_{x \in \mathbb{D}} r_n |\hat{M}_n(x) - \hat{M}_n^\#(x)| \xrightarrow{p} 0 \tag{2}$$

as $n \to \infty$, for any convex combination $\hat{M}_n^\#$ of $\hat{M}_n^\ell$ and $\hat{M}_n^u$. Thus, we get monotonicity free of charge. In the particular case where M is the mean residual life function and $\hat{M}_n$ is Yang's (1978) estimator, Kochar *et al.* (2000) obtained a similar result for the process $n^{1/2}(\hat{M}_n - \hat{M}_n^u)$ indexed by $x \in [0, b]$, for any $b < T$ with T being the support endpoint of the life-length of the population. Our result (Theorem 1) provides a general framework that covers this special case as well as new applications of considerable statistical interest, with an extension to multivariate isotonic functions.

Our first motivating application involves a monotone estimator of the conditional distribution function that has the distributional properties of the local linear regression estimator (Theorem 2). The latter estimator is not order-preserving even in the limit, as Hall and

Müller (2003) stated: "*This failure is generally most serious at boundaries of the distribution of the explanatory variable, and ironically it is often in just those places that estimation is of great interest, because responses there imply constraints on the larger population*". Our improved estimator of the conditional distribution function, obtained by projected local linear fitting, is particularly advantageous if one desires to invert it to produce a more relevant estimator of the regression quantile function (Corollary 1).

Applications of our general results also include the estimation of two multi-argument econometric functions: a probability-weighted moment and a quantile-based frontier functions which are found to be useful descriptors of the optimal cost and production quantities (see, *e.g.*, the survey article by Simar and Wilson (2008) for a nice summary). Both frontier functions are believed to be isotonic nondecreasing. The great simplicity of their empirical estimators and their desirable robustness and asymptotic properties are quite appealing, but they do not automatically inherit the monotonicity property. This article contributes to the frontier modeling by ensuring the monotonicity 'free of charge' via the projection type technique (Theorems 3 and 4). It also revisits the monotonization of Yang's (1978) estimator for the mean remaining lifetime function (Theorem 5).

The next section provides our general result for (2) and discusses the ideas of the proof. Section 3 gives in details the applications. Section 4 demonstrates other statistical properties via Monte Carlo simulations. Section 5 returns to our motivating econometric application and explores isotonic estimation of the quantile-based frontiers through the Ecuadorian manufacturing sector. Section 6 concludes with some results and directions of future research.

2 General results

We denote the Euclidean norm in $\mathbb{R}^d$ by $\|\cdot\|_d$, and the sup-norm on the domain $\mathbb{D}$ by $\|\cdot\|_{\mathbb{D}}$. We assume that the compact $\mathbb{D}$ takes the form $\mathbb{D} = \prod_{j=1}^d [a_j, b_j]$, where $-\infty < a_j < b_j < \infty$.

Theorem 1. *Assume that*

(C1) the first partial derivatives $\{\partial_i M : i = 1, \ldots, d\}$ *of* M *exist and satisfy* $\partial_i M(x) \geq c_i$, $\forall x \in \mathbb{D}$, *for some constants* $c_1, \ldots, c_d > 0$;

(C2) the second partial derivatives $\{\partial^2_{ij} M : i, j = 1, \ldots, d\}$ *exist and are continuous on* $\mathbb{D}$;

(C3) for a sequence $r_n \uparrow \infty$, *the process* $\mathbb{Z}_n := r_n(\hat{M}_n - M)$ *satisfies: for every* $\varepsilon > 0$, *there exist* $\delta > 0$ *and* $n_0 \in \mathbb{N}$ *such that*

$$\mathbb{P}\left\{ \sup_{\substack{x,x' \in \mathbb{D} \\ \|x-x'\|_d < \delta}} |\mathbb{Z}_n(x) - \mathbb{Z}_n(x')| > \varepsilon \right\} < \varepsilon \quad \forall n \geq n_0;$$

(C4) $\|\mathbb{Z}_n\|_{\mathbb{D}} = O_{a.s.}(s_n)$ *as* $n \to \infty$ *for some sequence* $s_n \uparrow \infty$*, and there exists* $\delta_n \to 0$ *such that* $r_n^{1/2}\delta_n \to 0$ *and* $s_n^{-1} r_n \delta_n \to \infty$*.*

Then for any convex combination $\hat{M}_n^{\#}$ *of* $\hat{M}_n^{\ell}$ *and* $\hat{M}_n^{u}$*,*

$$r_n \, \|\hat{M}_n^{\#} - \hat{M}_n\|_{\mathbb{D}} \xrightarrow{p} 0 \quad as \quad n \to \infty.$$

Assumption (C1) requires the strict monotonicity of $M(x_1, \ldots, x_d)$ with respect to each component x_j. Assumption (C3) is the asymptotic equicontinuity of the process $\mathbb{Z}_n$. Typically the assumption holds if $\mathbb{Z}_n$ converges weakly as a process indexed by $x \in \mathbb{D}$. For instance, weak convergence in the space of continuous functions with the uniform topology implies (C3). Weak convergence in the space of càdlàg functions (Skorohod space) with the Skorohod J_1 topology also implies (C3) if the limit process has continous sample paths. In Section 3, we discuss some interesting examples where (C3) is satisfied so that our method can be applied. The assumption does not hold in general, however, for nonparametric estimation of density and regression functions, see the related discussions in Ruymgaart (1998) or in Nishiyama (2011), for example. In the latter cases, one can typically verify only a 'local' version of the asymptotic equicontinuity. For example, in the case of local polynomial boundary estimation with a bandwidth h that goes to zero as n tends to infinity, one may prove only

$$\mathbb{P}\left\{ \sup_{\substack{x,x' \in \mathbb{D} \\ \|x-x'\|_d < \delta h}} |\mathbb{Z}_n(x) - \mathbb{Z}_n(x')| > \varepsilon \right\} < \varepsilon \quad \forall n \geq n_0,$$

see Hall and Park (1998), for example. As for Assumption (C4), if r_n in (C3) is the usual scaling $\sqrt{n}$ for empirical processes and s_n in (C4) is $O\left(\sqrt{\log\log n}\right)$ obtained typically from a law of iterated logarithm, then taking $\delta_n \sim n^{-\rho}$ with $1/4 < \rho < 1/2$ satisfies the condition.

Let $k_n = [\delta_n^{-1}]$ be the integer part of δ_n^{-1}, where δ_n is as described in (C4), and define for $j = 1, \ldots, d$,

$$\Delta_j = (b_j - a_j)/k_n \quad \text{and} \quad a_{j,\ell} = a_j + \ell\Delta_j \quad \text{for} \quad \ell = 0, 1, \ldots, k_n.$$

The key element in the proof is introducing the linear interpolation L_n, defined for any function h on $\mathbb{D}$, by

$$\begin{aligned} L_n h(x) = h(a_{1,\ell_1}, \ldots, a_{d,\ell_d}) + \sum_{j=1}^{d} \frac{x_j - a_{j,\ell_j}}{\Delta_j} \\ \times \left\{ h(a_{1,\ell_1}, \ldots, a_{j-1,\ell_{j-1}}, (a_{j,\ell_j} + \Delta_j), a_{j+1,\ell_{j+1}}, \ldots, a_{d,\ell_d}) - h(a_{1,\ell_1}, \ldots, a_{d,\ell_d}) \right\}, \end{aligned}$$

for $x = (x_1, \ldots, x_d) \in \mathbb{D}_{\ell_1,\ldots,\ell_d}$, where

$$\mathbb{D}_{\ell_1,\ldots,\ell_d} = \prod_{j=1}^{d} [a_{j,\ell_j}, a_{j,(\ell_j+1)}]$$

is a partitioning set of the domain $\mathbb{D}$, with $0 \le \ell_1, \ldots, \ell_d \le k_n - 1$. Note that

$$L_n h(a_{1,\ell_1}, \ldots, a_{d,\ell_d}) = h(a_{1,\ell_1}, \ldots, a_{d,\ell_d}) \quad \text{for all} \quad 0 \le \ell_1, \ldots, \ell_d \le k_n.$$

The idea of the proof is then to show that the transformation $L_n \hat{M}_n$ of $\hat{M}_n$ is eventually isotonic *a.s.*

Lemma 1. *Given Assumptions (C1) and (C4), $L_n \hat{M}_n$ is nondecreasing on $\mathbb{D}$ a.s. for all sufficiently large n.*

Since the # operator is sup-norm contracting, Lemma 1 entails that for all large n

$$\|\hat{M}_n^{\#} - L_n \hat{M}_n\|_{\mathbb{D}} = \|\hat{M}_n^{\#} - (L_n \hat{M}_n)^{\#}\|_{\mathbb{D}} \le \|\hat{M}_n - L_n \hat{M}_n\|_{\mathbb{D}} \quad a.s.$$

This leads to $\|\hat{M}_n^{\#} - \hat{M}_n\|_{\mathbb{D}} \le 2\|\hat{M}_n - L_n \hat{M}_n\|_{\mathbb{D}} \quad a.s.$, for all n large enough. Hence to complete the proof of Theorem 1, it suffices to show

Lemma 2. *If Assumptions (C1)-(C4) hold, then $r_n \|\hat{M}_n - L_n \hat{M}_n\|_{\mathbb{D}} \xrightarrow{p} 0$ as $n \to \infty$.*

Note that Assumption (C2) guarantees the following.

Lemma 3. *Under (C2), $\|M - L_n M\|_{\mathbb{D}} = O(\delta_n^2)$ for the sequence δ_n described in (C4).*

Note also that in the one-dimensional case ($d = 1$), the condition (C2) can be replaced by the weaker assumption that the second derivative M'' exists and $\|M''\|_{\mathbb{D}} < \infty$.

3 Main applications

3.1 Estimation of conditional distribution functions

Estimation of the conditional distribution function $F(y|x) \equiv P(Y \le y | X = x)$ is a main task in many statistical problems. One important example is quantile regression, where one typically estimates the α-quantile function $q_\alpha(x) \equiv F^{-1}(\alpha|x)$ by inverting an estimator of $F(\cdot|x)$. This problem was tackled by, for example, Yu and Jones (1998), Hall, Wolff and Yao (1999), and Lee, Lee and Park (2006). A promising technique by Yu and Jones (1998) which is inspired by the approach of Fan, Yao and Tong (1996) is to employ the local linear approach to smoothed versions of the indicator responses $\mathbb{I}(Y_i \le y)$. This is based on the fact that $F(y|x) = E[\mathbb{1}(Y \le y)|X = x]$. The main difficulty with this method is that it produces a distribution function estimator that is not constrained to be monotone increasing as a function of y for each fixed x. This is illustrated in *e.g.* Hall and Müller (2003, Section 2).

Suppose we have a random sample $\{(X_i, Y_i) : 1 \le i \le n\}$ from (X, Y). Let K be a symmetric nonnegative function supported on a compact set, say $[-1, 1]$, and L be a distribution function of a symmetric density. For the bandwidths h and b, associated with the kernels K and L, the local linear approach of Yu and Jones (1998) for estimating $F(y|x)$ minimizes

$$\sum_{i=1}^{n} \left[L\left(\frac{y - Y_i}{b}\right) - \beta_0 - \beta_1(X_i - x) \right]^2 K\left(\frac{X_i - x}{h}\right) \tag{3}$$

to get $\tilde{F}(y|x) = \hat{\beta}_0$. If b tends to zero faster than h, then the first-order properties of $\tilde{F}(y|x)$ are the same as those of $\hat{F}(y|x) = \hat{\gamma}_0$, where $(\hat{\gamma}_0, \hat{\gamma}_1)$ minimizes

$$\sum_{i=1}^{n} \left[\mathbb{1}(Y_i \le y) - \beta_0 - \beta_1(X_i - x)\right]^2 K\left(\frac{X_i - x}{h}\right).$$

To simplify the discussion, we focus on the latter $\hat{F}(y|x)$. We can write

$$\hat{F}(y|x) = \sum_{i=1}^{n} w_i(x; X_1, \ldots, X_n)\, \mathbb{1}(Y_i \le y),$$

where $K_i = K((X_i - x)/h)$ and

$$w_i(x; X_1, \ldots, X_n) = \frac{\sum_{j=1}^{n}(X_j - x)^2 K_j - (X_i - x)\sum_{j=1}^{n}(X_j - x)K_j}{\left[\sum_{j=1}^{n} K_j\right]\left[\sum_{j=1}^{n}(X_j - x)^2 K_j\right] - \left[\sum_{j=1}^{n}(X_j - x)K_j\right]^2} \cdot K_i.$$

Note that $\sum_{i=1}^{n} w_i(x; X_1, \ldots, X_n) = 1$, but $w_i(x; X_1, \ldots, X_n)$ are not constrained to be nonnegative.

We let x be fixed in the interior of the support of X_i, and assume that $h \to 0$ and $nh \to \infty$ as n tends to infinity. Under some conditions on the conditional distribution function $F(\cdot|\cdot)$ and the density of X, denoted by f_X, one can prove that $\sqrt{nh}(\hat{F}(y|x) - F(y|x) - h^2 c(x, y))$ converges to a normal distribution for some function c, see Yu and Jones (1998) or Lee, Lee and Park (2006). Thus, $r_n = \sqrt{nh}$ is the right scaling and we consider the process $\mathbb{Z}_n$ defined by $\mathbb{Z}_n(y) = \sqrt{nh}(\hat{F}(y|x) - F(y|x))$. One may think that an application of existing weak convergence results for weighted empirical processes can verify the condition (C3). To do this, one may apply probability integral transformations to Y_i to get indicators for uniform random variables. This approach does not work in the present case since one should make the transformation conditionally on X_i for each i, i.e., take $F(\cdot|X_i)$ to get uniform random variables (conditionally). The latter gives $\mathbb{1}[F(Y_i|X_i) \le F(y|X_i)]$, so that one may not express $\mathbb{Z}_n$ as an empirical process of the conditionally uniform random variables $F(Y_i|X_i)$.

We define $\xi_i(y) = \mathbb{1}(Y_i \leq y) - F(y|X_i)$, and we write $w_i \equiv w_i(x; X_1, \ldots, X_n)$ and $\mathcal{X} \equiv (X_1, \ldots X_n)$ for simplicity. We decompose $\mathbb{Z}_n(y)$ into two terms

$$\mathbb{Z}_{1n}(y) = \sqrt{nh} \sum_{i=1}^{n} w_i \xi_i(y) \quad \text{and} \quad \mathbb{Z}_{2n}(y) = \sqrt{nh} \sum_{i=1}^{n} w_i \left[F(y|X_i) - F(y|x) \right],$$

so that $\mathbb{Z}_n = \mathbb{Z}_{1n} + \mathbb{Z}_{2n}$. The idea is then to show directly that both $\mathbb{Z}_{1n}$ and $\mathbb{Z}_{2n}$ satisfy (C3) by applying a chaining technique with a maximal inequality for partial sums, and the standard theory of kernel smoothing. Define $F^{ab}(y|u) = \partial^{a+b} F(y|u)/(\partial y^a \partial u^b)$. Let $\mathbb{D}$ be a compact set of the form $\mathbb{D} = [a, b]$, where $-\infty < a < b < \infty$.

Lemma 4. *Suppose that the bandwidth h is asymptotic to $n^{-\alpha}$ for some $1/5 \leq \alpha < 1/3$. Assume that $F^{10}(y|u)$ is bounded for y in $\mathbb{D}$ and for u in a neighborhood of x. Assume also that $F^{02}(y|u)$ is continuous at $u = x$ uniformly for $y \in \mathbb{D}$, that $F^{02}(\cdot|x)$ is continuous on $\mathbb{D}$, and that f_X is continuous at x with $f_X(x) > 0$. Then, for every $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\mathbb{P}\Bigg[\sup_{\substack{y, y' \in \mathbb{D} \\ |y'-y| \leq \delta}} \big| \mathbb{Z}_{1n}(y') - \mathbb{Z}_{1n}(y) \big| > \varepsilon \,\Big|\, \mathcal{X} \Bigg] \leq \varepsilon$$

with probability tending to one, and that

$$\mathbb{P}\Bigg[\sup_{\substack{y, y' \in \mathbb{D} \\ |y'-y| \leq \delta}} \big| \mathbb{Z}_{2n}(y') - \mathbb{Z}_{2n}(y) \big| > \varepsilon \Bigg] \leq \varepsilon$$

for sufficiently large n.

By using some exponential inequalities, we also prove that $\mathbb{Z}_n$ satisfies (C4).

Lemma 5. *Suppose that the bandwidth h is asymptotic to $n^{-\alpha}$ for some $1/5 \leq \alpha < 1$. Assume that $F^{02}(y|x)$ is bounded for $y \in \mathbb{D}$, and that f_X is continuous at x with $f_X(x) > 0$. Then,*

$$\|\mathbb{Z}_{1n}\|_{\mathbb{D}} = O_{a.s.}(\sqrt{\log n}) \quad \textit{and} \quad \|\mathbb{Z}_{2n}\|_{\mathbb{D}} = O_{a.s.}(1).$$

The above two lemmas give the asymptotic equivalence between $\hat{F}(\cdot|x)$ and the nondecreasing projected estimators $\hat{F}^{\#}(\cdot|x)$ as demonstrated below.

Theorem 2. *Assume the conditions of Lemma 4. Assume also that $\inf_{y \in \mathbb{D}} F^{10}(y|x) > 0$, and that $F^{20}(\cdot|x)$ is continuous on $\mathbb{D}$. Then, $\sqrt{nh}\, \|\hat{F}^{\#}(\cdot|x) - \hat{F}(\cdot|x)\|_{\mathbb{D}} \xrightarrow{p} 0$ as $n \to \infty$.*

For the double kernel estimator $\tilde{F}(y|x)$ which minimizes (3), it can be proved that $\sqrt{nh}(\tilde{F}(y|x) - F(y|x) - h^2 c_1(x, y) - b^2 c_2(x, y))$ converges to a normal distribution for some functions c_1 and c_2 under certain conditions. Thus, $r_n = \sqrt{nh}$ remains to be the proper scaling for $\tilde{F}(y|x)$. Along the lines of the proof for Theorem 2, one can prove that the theorem is also valid for $\tilde{F}(y|x)$.

A relevant application concerns the estimation of conditional quantiles $F^{-1}(\alpha|x) = \inf\{y : F(y|x) \geq \alpha\}$ for $\alpha \in (0, 1)$. Choosing

$$\hat{F}^{-1}(\alpha|x) = \inf\{y : \hat{F}(y|x) \geq \alpha\} \text{ and } \hat{F}^{\#-1}(\alpha|x) = \inf\{y : \hat{F}^{\#}(y|x) \geq \alpha\},$$

as estimates of $F^{-1}(\alpha|x)$, we get the following corollary.

Corollary 1. *Assume the conditions of Theorem 2. In addition, suppose that the inverse $F^{-1}(\alpha|x)$ of $F(\cdot|x)$ belongs to $\mathbb{D}$ for $\alpha \in [q, r]$ and $0 < q < r < 1$. Then*

$$\sqrt{nh} \sup_{q \leq \alpha \leq r} \left| \hat{F}^{-1}(\alpha|x) - \hat{F}^{\#-1}(\alpha|x) \right| \xrightarrow{p} 0 \quad as \quad n \to \infty.$$

3.2 Estimation of partial frontier functions

Partial frontier models find increasing usage in management, finance, economics, education, public policy, and other areas. When analyzing the productivity of firms, one may want to compare how the firms transform a set of inputs-usage $X \in \mathbb{R}_+^d$ (*e.g.* labor, energy, capital) into an output $Y \in \mathbb{R}_+$ (a quantity of produced goods or services). In this context, the joint support of (X, Y) is interpreted as the set of all possible firms and its upper boundary is viewed as the set of the most efficient ones. From an economic point of view, this optimal support boundary is supposed to be isotonic nondecresing (see, *e.g.*, Gijbels, Mammen, Park and Simar 1999).

Let $F(\cdot, \cdot)$ and $F_X(\cdot)$, respectively, denote the joint and marginal distribution functions of (X, Y) and X. An important function in productivity analysis is given by $\varphi(x) = \sup\{y \in \mathbb{R}_+ : F(y|x) < 1\}$, where $F(y|x) = F(x, y)/F_X(x)$ assuming $F_X(x) > 0$. We note here that the definition of the 'conditional' distribution function $F(\cdot|\cdot)$ is different from the standard one in the previous section. Thus, generally speaking, $\varphi(x)$ is not the upper boundary of the support of (X, Y) at $X = x$, say $\phi(x)$, but equals $\sup_{x' \leq x} \phi(x')$. Thus, it is monotone nondecreasing and envelops the support boundary. In the case where the production frontier function ϕ is nondecreasing, φ coincides with ϕ. In the latter case, consideration of φ is advantageous since it affords estimation at a faster rate than ϕ. Because of the local nature of ϕ, one can use only the observations in a local strip around x to estimate it (see, *e.g.*, Gijbels and Peng (2000)), while it is not the case with estimation of φ.

Given a random sample $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$ of i.i.d. firms, and taking $\hat{F}_n(y|x) = \sum_{i=1}^n \mathbb{1}(X_i \leq x, Y_i \leq y) / \sum_{i=1}^n \mathbb{1}(X_i \leq x)$, a usual estimator of $\varphi(x)$ is the Free Disposal Hull

(FDH) estimator

$$\hat{\varphi}(x) = \sup\{y \geq 0 \mid \hat{F}_n(y|x) < 1\} = \max_{i:X_i \leq x} Y_i$$

which is the lowest step and monotone function that envelops all the data points (see, *e.g.*, Park *et al.* (2000)). When the joint support is assumed to be convex, one can use the conventional Data Envelopment Analysis (DEA) estimator defined as the smallest concave function covering the FDH estimator (see, *e.g.*, Gijbels *et al.* (1999)). Regrettably both the FDH and DEA estimators are, by construction, very non-robust.

Due to the predominance of outliers in production data, a robust approach is not to estimate the true frontier φ itself, but to estimate a *partial frontier* well inside the support of (X, Y) but lying close to the full frontier φ, as suggested by Cazals *et al.* (2002). Formally, they estimate the expected value of the maximum of m $(m = 1, 2, \ldots)$ independent random variables $Y_x^1, \cdots, Y_x^m$, drawn from the conditional distribution of Y given $X \leq x$, that is,

$$\varphi_m(x) := \mathbb{E}\left[\max(Y_x^1, \ldots, Y_x^m)\right] = \varphi(x) - \int_0^{\varphi(x)} F^m(y|x)\, dy.$$

While $\varphi(x)$ represents the maximum attainable output for a firm working at the level of inputs x, $\varphi_m(x)$ gives the expected maximum achievable production among a fixed number of m firms using less inputs than x. The limiting case where $m \to \infty$ is of particular interest: it achieves the monotone efficient frontier $\varphi(x)$.

In view of economic considerations, the chance of producing less than a value y decreases if a firm utilizes more inputs (*i.e.*, $F(y|x') \leq F(y|x)$ for all $x \leq x'$). Under this natural hypothesis, the frontier function $\varphi_m(x)$ is also nondecreasing in x. Its empirical estimator

$$\hat{\varphi}_{m,n}(x) = \hat{\varphi}(x) - \int_0^{\hat{\varphi}(x)} \hat{F}_n^m(y|x)\, dy$$

does not enjoy the desirable property of monotonicity, however. Cazals *et al.* (2002) showed the strong consistency of $\hat{\varphi}_{m,n}(x)$ and its functional convergence to a Gaussian process, provided that the joint support of (X, Y) is compact. Daouia and Gijbels (2011) have strengthened some of Cazals *et al.*'s results: assuming that the support of Y is bounded, they proved that for any $\mathbb{D} \subset \mathbb{R}_+^d$ such that $\inf_{x\in\mathbb{D}} F_X(x) > 0$, the process $\mathbb{Z}_n = \{\sqrt{n}(\hat{\varphi}_{m,n}(x) - \varphi_m(x)),\ x \in \mathbb{D}\}$ converges weakly in the space $L^\infty(\mathbb{D})$ of bounded functions on $\mathbb{D}$ to the centered Gaussian process $\mathbb{Z}$ defined by

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

with $\mathbb{F}(\cdot, \cdot)$ being a $(d+1)$-dimensional $F(\cdot, \cdot)$-Brownian bridge. By applying Theorem 1 in conjunction with this weak convergence to $M = \varphi_m$ and $\hat{M}_n = \hat{\varphi}_{m,n}$, we show here that any convex combination $\hat{\varphi}^{\#}_{m,n}$ of the envelope estimators $\hat{\varphi}^{u}_{m,n}$ and $\hat{\varphi}^{\ell}_{m,n}$ is asymptotically $\sqrt{n}$-equivalent in probability to the unrestricted $\hat{\varphi}_{m,n}$ estimator.

Theorem 3. *Suppose that the support of Y is bounded, and let $\mathbb{D}$ be any subset interior to the support of X of the form $\prod_{j=1}^{d}[a_j, b_j]$ such that $\inf_{x\in\mathbb{D}} F_X(x) > 0$. Also, assume that $F(\cdot,\cdot)$ and F_X are continuous on the supports. If φ_m satisfies the conditions (C1) and (C2) of Theorem 1, then $\sqrt{n}\,\|\hat{\varphi}^{\#}_{m,n} - \hat{\varphi}_{m,n}\|_{\mathbb{D}} \xrightarrow{p} 0$ as $n \to \infty$.*

Hendricks and Koenker (1992) stated, "*In the econometric literature on the estimation of production technologies, there has been considerable interest in estimating the so called frontier production models that correspond closely to models for extreme quantiles of a stochastic production surface*". The paper of Aragon *et al.* (2005) may be viewed as the first work to actually implement the idea of Hendricks and Koenker: they introduced an alternative partial frontier function defined by

$$\psi_\alpha(x) := \inf\{y \in \mathbb{R}_+ \mid F(y|x) \geq \alpha\}$$

for $\alpha \in (0, 1)$. This conditional quantile function converges to the monotone efficient frontier $\varphi(x) \equiv \psi_1(x)$ as $\alpha \to 1$. It is also isotonic nondecreasing in x under the economic hypothesis that $F(y|x)$ is monotone nonincreasing in x. Its empirical estimator

$$\hat{\psi}_{\alpha,n}(x) = \inf\{y \in \mathbb{R}_+ \mid \hat{F}_n(y|x) \geq \alpha\}$$

satisfies very similar statistical properties to those of the sample m-trimmed frontier $\hat{\varphi}_{m,n}(x)$. In particular, for any subset $\mathbb{D} \subset \mathbb{R}^d_+$ interior to the support of X such that

(Q1) $\inf_{x\in\mathbb{D}} F_X(x) > 0$, and $f(\cdot|x) = F'(\cdot|x)$ exists and is continuous for all $x \in \mathbb{D}$,

(Q2) $\inf_{x\in\mathbb{D}} \inf_{\varepsilon\leq\alpha\leq 1-\varepsilon} f(\psi_\alpha(x)|x) > 0$ for all $0 < \varepsilon < 1/2$,

Horváth, Horváth and Zhou (2008) showed that, for all $\varepsilon \in (0, 1/2)$,

$$\sup_{\varepsilon\leq\alpha\leq 1-\varepsilon} \|\hat{\psi}_{\alpha,n} - \psi_\alpha\|_{\mathbb{D}} = O_{a.s.}\left((\log\log n/n)^{1/2}\right).$$

They also proved that the process $\sqrt{n}(\hat{\psi}_{\alpha,n} - \psi_\alpha)$ converges weakly in the space $L^\infty(\mathbb{D})$ to the centered Gaussian process

$$\left\{\frac{\alpha\mathbb{F}(x,\infty) - \mathbb{F}(x,\psi_\alpha(x))}{f(\psi_\alpha(x)|x)F_X(x)},\, x \in \mathbb{D}\right\}.$$

Regrettably, the empirical step function $\hat{\psi}_{\alpha,n}$ has no guarantee of being monotone even if ψ_α is so. The isotonic estimator $\hat{\psi}^{\#}_{\alpha,n}$ is shown here to be asymptotically $\sqrt{n}$-equivalent in probability to the unconstrained $\hat{\psi}_{\alpha,n}$ estimator.

Theorem 4. *Suppose that the support of Y is bounded, and let $\mathbb{D}$ be any subset interior to the support of X of the form $\prod_{j=1}^{d}[a_j, b_j]$ fulfilling the conditions (Q1)-(Q2). Also, assume that $F(\cdot,\cdot)$, F_X and $f(\cdot|\cdot)$ are continuous on their supports. If ψ_α satisfies the conditions (C1) and (C2) of Theorem 1, then $\sqrt{n}\,\|\hat{\psi}^{\#}_{\alpha,n} - \hat{\psi}_{\alpha,n}\|_{\mathbb{D}} \xrightarrow{p} 0$ as $n \to \infty$.*

Note that, although the classes $\{\varphi_m(\cdot),\, m \geq 1\}$ and $\{\psi_\alpha(\cdot),\, \alpha \in (0,1]\}$ have emerged in the econometric literature as two different appealing concepts of partial production functions, recently Daouia and Gijbels (2011) established that they are closely linked in the sense that, for each $m \geq 1$, there exists a well-specified order $\alpha = (1/2)^{1/m}$ such that the pointwise values $\varphi_m(x)$ and $\psi_\alpha(x)$ are respectively the theoretical mean and median of the same distribution, namely $F^m(\cdot|x)$. This confirms the well-known advantage of the quantile type α-frontiers over the probability weighted moment m-frontiers in terms of finite sample breakdown point and gross-error sensitivity, but such a robust proposal may sacrifice efficiency.

3.3 Estimation of biometric functions

A function of prime importance in many statistical studies involving survival analysis, biometric mortality data, failure data and actuarial data is the mean residual lifetime (MRL) function. It is defined by

$$M(x) = \mathbb{E}[X - x | X > x] = 1\!\!I(S_X(x) > 0) \int_x^\infty S_X(y)dy/S_X(x) \quad \text{for} \quad x \in [0, \infty),$$

where X is a continuous non-negative random variable with finite mean $M(0)$, representing the life length of the population, and S_X is its survival function with support $[0, T]$ for a possibly infinite endpoint T. For life tables, $M(x)$ is called the life expectancy at age x. It describes the average remaining life among those population members who have survived until time x. In many cases there are reasons to believe that M is nondecreasing or nonincreasing in x due to monotonic improvement or deterioration of the system life with age, see, *e.g.*, Guess and Proschan (1988) and the references therein.

Let $X_1, \ldots, X_n$ be independent copies of the random variable X. Denoting by $\hat{S}_{X,n}$ the corresponding empirical survival function and by $X_{(n)}$ the sample maximum, Yang (1978) introduced the empirical estimator

$$\hat{M}_n(x) = 1\!\!I(x < X_{(n)}) \int_x^\infty \hat{S}_{X,n}(y)dy/\hat{S}_{X,n}(x),$$

and established its uniform strong consistency on $[0, b]$, for any fixed $b < T$, and that $\mathbb{Z}_n = \sqrt{n}(\hat{M}_n - M)$ converges weakly to a Gaussian process. Assuming that $\mathbb{E}(X^p) < \infty$ for some $p > 2$, Hall and Wellner (1979) strengthened Yang's result showing that $\mathbb{Z}_n$ converges weakly in the space $D([0,T))$ in the Skorohod topology J_1 to the centered Gaussian process $\mathbb{Z}$ defined by

$$\mathbb{Z}(x) = (\sigma(0)/\sigma(x))\,\mathbb{B}(U(x)), \tag{4}$$

where $\sigma^2(x) = Var[X - x|X > x]$, $U(x) = S_X(x)\sigma^2(x)/\sigma^2(0)$, and $\mathbb{B}$ is a standard Brownian motion. Under the same condition, Kochar *et al.* (2000) showed that

$$\sup_{x\in[0,b]} |\mathbb{Z}_n(x)|/(\log\log n)^{1/2} = O_{a.s.}(1). \tag{5}$$

We obtain the following theorem for the asymptotic equivalence of the non-monotone piecewise linear estimator $\hat{M}_n$ and its projected estimators $\hat{M}_n^{\#}$ in the nondecreasing case.

Theorem 5. *Assume that F_X is continuous on $[0,T]$. If M satisfies the conditions (C1)-(C2) of Theorem 1 with $\mathbb{D} = [0,b]$ and $\mathbb{E}(X^p) < \infty$ for some $p > 2$, then $\sqrt{n}\,\|\hat{M}_n^{\#} - \hat{M}_n\|_{\mathbb{D}} \xrightarrow{p} 0$ as $n \to \infty$.*

This result is not new. It can be found in Kochar *et al.* (2000) under the decreasing constraint, where the weaker assumption that the second derivative M'' exists and $\|M''\|_{\mathbb{D}} < \infty$ is used instead of Condition (C2). The simulations in Kochar *et al.* (2000) seem to indicate that the restricted $\hat{M}_n^{\star}$ version is uniformly superior to the initial MRL estimator $\hat{M}_n$ in terms of mean squared error, although it has uniformly a higher negative bias.

4 Some simulation evidence

This section provides Monte Carlo evidence that the monotonized hybrid estimators $\hat{M}_n^{\star}$ of the conditional distribution and econometric functions are efficient relative to the unconstrained estimators $\hat{M}_n$. To compute the constrained estimators, we took a discrete grid $\mathbb{D}_n$ instead of the whole domain $\mathbb{D}$ in the equation (1). For our Monte Carlo exercises, we used $2n$ grid points evenly distributed across the entire sample space of $\{Y_i\}$ for the conditional distribution function, and of $\{X_i\}$ for the econometric functions.

4.1 Estimation of conditional distribution functions

We consider the following model:

$$Y_i = g(X_i) + \varepsilon_i, \quad X_i \sim U(0,1), \quad \varepsilon_i \sim N(0,\sigma^2), \quad i = 1,\ldots,n,$$

where $\{X_i\}$ and $\{\varepsilon_i\}$ are two independent sequences of independent random variables. We first chose $g(x) = \exp(x)$, $\sigma = 0.5$ and the kernel function K to be the triweigth kernel $K(t) = \frac{35}{32}(1-t^2)^3 I(-1 \le t \le 1)$. The local linear estimator $\hat{F}(y|x)$ of the conditional distribution function $F(y|x)$ would be ideally monotone increasing in y, but this is unlikely the case especially at the endpoints $x \in \{0,1\}$ and near them (see Hall and Müller (2003)). This vexing defect can be circumvented by using, for instance, the projected hybrid variant $\hat{F}^{\star}(\cdot|x)$ which keeps the same convergence properties as the unconstrained estimator $\hat{F}(y|x)$. Monte Carlo experiments were performed over 5000 simulations in order to compare $\hat{F}(\cdot|x)$ and $\hat{F}^{\star}(\cdot|x)$ for different quantiles x in $[0,1]$. Sample sizes of 100 and 500 were used. We chose $h = n^{-\alpha}$ for $\alpha = \frac{1}{5}$ and $\alpha = \frac{2}{5}$. The measures of efficiency for each simulation were the mean squared error (MSE) and the mean absolute deviation error (MADE)

$$\text{MSE}\ [\pi(\cdot|x)] = \frac{1}{n}\sum_{i=1}^{n}\{\pi(Y_i|x) - F(Y_i|x)\}^2, \quad \text{MADE}\ [\pi(\cdot|x)] = \frac{1}{n}\sum_{i=1}^{n}|\pi(Y_i|x) - F(Y_i|x)|,$$

where $\pi(\cdot|x)$ is either $\hat{F}(\cdot|x)$ or $\hat{F}^{\star}(\cdot|x)$. The Monte Carlo averages of MSE$[\pi(\cdot|x)]$ and of MADE $[\pi(\cdot|x)]$ over the 5000 replications are reported in Table 1 for the sample sizes $n = 100$ and $n = 500$. It is seen that $\hat{F}^{\star}(\cdot|x)$ performs at least as well as the empirical $\hat{F}(\cdot|x)$ in terms of MSE in all cases and in terms of MADE in almost all cases (the gray cells indicate where $\hat{F}^{\star}(\cdot|x)$ is slightly inferior). The constrained estimator $\hat{F}^{\star}(\cdot|x)$ gets better with larger margin as x approaches to the endpoints.

Table 1: *Results for* 5000 *Monte-Carlo simulations with* $n = 100$, 500.

	$n=100$ & $\alpha=1/5$				$n=100$ & $\alpha=2/5$			
	MSE		MADE		MSE		MADE	
x	$\hat{F}(\cdot\|x)$	$\hat{F}^{\star}(\cdot\|x)$	$\hat{F}(\cdot\|x)$	$\hat{F}^{\star}(\cdot\|x)$	$\hat{F}(\cdot\|x)$	$\hat{F}^{\star}(\cdot\|x)$	$\hat{F}(\cdot\|x)$	$\hat{F}^{\star}(\cdot\|x)$
0	.0149	.0136	.0779	.0740	.0365	.0324	.1183	.1117
.01	.0128	.0118	.0730	.0698	.0284	.0258	.1051	.1005
.05	.0079	.0076	.0584	.0569	.0107	.0104	.0670	.0660
.1	.0049	.0047	.0467	.0460	.0063	.0062	.0534	.0528
.25	.0028	.0028	.0373	.0372	.0067	.0065	.0571	.0564
.5	.0030	.0030	.0408	.0408	.0073	.0071	.0619	.0612
.75	.0029	.0029	.0388	.0388	.0065	.0064	.0565	.0558
.9	.0041	.0039	.0406	.0399	.0051	.0050	.0457	.0450
.95	.0062	.0058	.0475	.0459	.0079	.0076	.0528	.0518
.99	.0093	.0083	.0552	.0516	.0197	.0176	.0762	.0722
1	.0104	.0092	.0573	.0533	.0258	.0224	.0852	.0793

	$n=500$ & $\alpha=1/5$				$n=500$ & $\alpha=2/5$			
	MSE		MADE		MSE		MADE	
x	$\hat{F}(\cdot\|x)$	$\hat{F}^{\star}(\cdot\|x)$	$\hat{F}(\cdot\|x)$	$\hat{F}^{\star}(\cdot\|x)$	$\hat{F}(\cdot\|x)$	$\hat{F}^{\star}(\cdot\|x)$	$\hat{F}(\cdot\|x)$	$\hat{F}^{\star}(\cdot\|x)$
0	.0040	.0039	.0413	.0403	.0137	.0128	.0747	.0721
.01	.0033	.0032	.0376	.0369	.0076	.0073	.0561	.0553
.05	.0017	.0017	.0276	.0274	.0022	.0022	.0313	.0312
.1	.0010	.0009	.0208	.0208	.0021	.0021	.0312	.0310
.25	.0007	.0007	.0190	.0190	.0024	.0024	.0346	.0345
.5	.0008	.0008	.0211	.0212	.0027	.0027	.0376	.0374
.75	.0008	.0008	.0203	.0204	.0024	.0024	.0344	.0343
.9	.0008	.0008	.0184	.0184	.0018	.0018	.0274	.0273
.95	.0013	.0013	.0224	.0221	.0017	.0017	.0253	.0252
.99	.0024	.0023	.0291	.0281	.0054	.0052	.0425	.0418
1	.0029	.0027	.0314	.0301	.0094	.0087	.0541	.0518

We also compared the projected local linear estimator $\hat{F}^{\star}(\cdot|x)$ with various regression estimators of $F(\cdot|x)$ in Table 2. Those were the Nadaraya-Watson estimator, the least squares (LS) polynomial estimator, the LS spline estimator and the smoothing spline estimator. The standard Nadaraya-Watson estimator is defined by

$$\hat{F}_{NW}(y|x) = \sum_{i=1}^{n} K((x - X_i)/h) Z_i^y / \sum_{i=1}^{n} K((x - X_i)/h),$$

with $Z_i^y = I(Y_i \leq y)$. For both the Nadaraya-Watson and local linear estimators, we used bandwidths h ranging over a refined grid of 200 points regularly distributed between $h_{min} = \max_{1 \leq i < n}\{X_{(i+1)} - X_{(i)}\}$ and $h_{max} = \{X_{(n)} - X_{(1)}\}$, where $X_{(1)} \leq \cdots \leq X_{(n)}$ are the ordered observations. The LS polynomial estimator of $F(y|\cdot) \equiv \mathbb{E}(Z^y|X = \cdot)$ is given by

$$\hat{F}_{LP}(y|\cdot) = \underset{P \in \mathcal{P}_N}{\operatorname{argmin}} \sum_{i=1}^{n} (Z_i^y - P(X_i))^2,$$

where $\mathcal{P}_N$ stands for the set of polynomials of degree less than or equal to N, a smoothing parameter ranging over $\{0, \ldots, n-1\}$. If $\{P_0, \ldots, P_N\}$ is a basis of the space $\mathcal{P}_N$, then $\hat{F}_{LP}(y|x) = \sum_{j=0}^{N} \hat{\theta}_j(y) P_j(x)$, with $\{\hat{\theta}_j(y)\}$ being the solution of

$$\min_{(\theta_0, \ldots, \theta_N) \in \mathbb{R}^{N+1}} \sum_{i=1}^{n} \left(Z_i^y - \sum_{j=0}^{N} \theta_j P_j(X_i) \right)^2.$$

The LS cubic spline estimator of the regression function $F(y|\cdot)$ is defined by

$$\hat{F}_{RS}(y|\cdot) = \underset{S \in \mathcal{S}_4(x^1, \ldots, x^k)}{\operatorname{argmin}} \sum_{i=1}^{n} (Z_i^y - S(X_i))^2,$$

where $\mathcal{S}_4(x^1, \ldots, x^k)$ is the set of cubic splines with given knots $x^1, \ldots, x^k$ in $[0, 1]$. The smoothing parameter k ranges over $\{0, \ldots, n-4\}$ and the knots are distributed regularly or at the empirical quantiles of X. If $\{S_1, \ldots, S_{k+4}\}$ is a basis of the vector space $\mathcal{S}_4(x^1, \ldots, x^k)$, then $\hat{F}_{RS}(y|x) = \sum_{j=1}^{k+4} \hat{\theta}_j(y) S_j(x)$, where the $\{\hat{\theta}_j(y)\}$ is the solution of

$$\min_{(\theta_1, \ldots, \theta_{k+4}) \in \mathbb{R}^{k+4}} \sum_{i=1}^{n} \left(Z_i^y - \sum_{j=1}^{k+4} \theta_j S_j(X_i) \right)^2.$$

The smoothing cubic spline estimator of $F(y|\cdot)$ is given by

$$\hat{F}_{SS}(y|\cdot) = \underset{W \in \mathcal{W}^{2,2}(0,1)}{\operatorname{argmin}} \sum_{i=1}^{n} (Z_i^y - W(X_i))^2 + \lambda \int_0^1 (W''(t))^2 dt,$$

where $\mathcal{W}^{2,2}(0,1)$ is the Sobolev space of order 2 on $(0,1)$ and $\lambda > 0$ is the smoothing parameter. If $\{W_1, \ldots, W_n\}$ is a basis of the natural cubic splines of $\mathcal{S}_4(X_1, \ldots, X_n)$, then $\hat{F}_{SS}(y|x) = \sum_{j=1}^{n} \hat{\theta}_j(y) W_j(x)$, where $\{\hat{\theta}_j(y)\}$ is the solution of

$$\min_{(\theta_1,\ldots,\theta_n)\in\mathbb{R}^n} \sum_{i=1}^{n} \left(Z_i^y - \sum_{j=1}^{n} \theta_j W_j(X_i) \right)^2 + \int_0^1 \left(\sum_{j=1}^{n} \theta_j W_j''(t) \right)^2 dt.$$

To guarantee a fair comparison among the different methods, we used for each estimator $\pi \in \{\hat{F}^\star, \hat{F}, \hat{F}_{NW}, \hat{F}_{LP}, \hat{F}_{RS}, \hat{F}_{SS}\}$ the smoothing parameter which minimizes the respective mean squared error MSE $[\pi(\cdot|x)]$. The Monte Carlo averages of MSE$[\pi(\cdot|x)]$ and of MADE $[\pi(\cdot|x)]$, computed over 600 samples of size $n = 100$, confirm the superiority of the monotonized local linear estimator over the unconstrained version, especially near and at the endpoints x. Compared with the Nadaraya-Watson estimator, we see that both the initial and monotonized local linear estimators are clearly the winners in the right tail values of x. However, $\hat{F}_{NW}(\cdot|x)$ seems to outperform in this particular example the unrestricted estimator in the left tails (this is probably due to the disadvantage of producing distribution function estimates $\hat{F}(y|x)$ that are not constrained to be positive), while the projected version $\hat{F}^\star(\cdot|x)$ reduces considerably both MSE and MADE. Compared to the least squares spline estimator, $\hat{F}^\star(\cdot|x)$ keeps the big margins that $\hat{F}(\cdot|x)$ takes over $\hat{F}_{RS}(\cdot|x)$. Finally, $\hat{F}^\star(\cdot|x)$ performs in almost all cases at least as well as $\hat{F}_{SS}(\cdot|x)$ and $\hat{F}_{LP}(\cdot|x)$, especially in the tail values of x.

Table 2: *Results for* 600 *Monte-Carlo simulations with* $n = 100$.

	MSE$[\pi(\cdot\|x)]$						MADE$[\pi(\cdot\|x)]$					
x	$\hat{F}^\star$	$\hat{F}$	$\hat{F}_{NW}$	$\hat{F}_{RS}$	$\hat{F}_{SS}$	$\hat{F}_{LP}$	$\hat{F}^\star$	$\hat{F}$	$\hat{F}_{NW}$	$\hat{F}_{RS}$	$\hat{F}_{SS}$	$\hat{F}_{LP}$
0	.0046	.0051	.0045	.6992	.0067	.0087	.0444	.0472	.0425	.7955	.0617	.0722
.01	.0045	.0049	.0044	.2693	.0062	.0078	.0440	.0464	.0422	.3435	.0592	.0676
.05	.0039	.0041	.0037	.0097	.0047	.0049	.0409	.0426	.0389	.0568	.0505	.0511
.1	.0030	.0031	.0029	.0037	.0036	.0034	.0374	.0384	.0358	.0423	.0441	.0421
.25	.0021	.0021	.0020	.0030	.0019	.0018	.0319	.0321	.0315	.0402	.0319	.0318
.5	.0022	.0022	.0023	.0026	.0024	.0024	.0350	.0351	.0355	.0384	.0361	.0373
.75	.0023	.0023	.0024	.0030	.0022	.0021	.0346	.0346	.0348	.0404	.0346	.0342
.9	.0026	.0028	.0035	.0031	.0033	.0029	.0320	.0334	.0367	.0379	.0387	.0379
.95	.0031	.0036	.0046	.0053	.0045	.0042	.0332	.0357	.0397	.0456	.0452	.0449
.99	.0041	.0049	.0071	.0275	.0068	.0077	.0355	.0394	.0471	.0856	.0569	.0633
1	.0038	.0047	.0072	.0607	.0079	.0097	.0340	.0381	.0468	.1269	.0613	.0732

We also performed a Monte Carlo comparison with a linear function $g(x) = x$. The results were similar to the first example, so they are not reported here. The benefits of our restricted estimator $\hat{F}^\star(\cdot|x)$ are acheived especially near and at the support endpoints 'x'.

4.2 Estimation of partial frontier functions

Here, we first considered the Cobb-Douglas model:

$$Y_i = X_i^{1/2} \exp(-U_i), \quad i = 1, \ldots, n,$$

where X_i is uniformly distributed on $(0,1)$ and U_i, independent of X_i, has an exponential distribution with mean 1/3. In this case, we have $\psi_\alpha(x) = x^{1/2}\left\{\cos\left(\frac{\arccos(1-2\alpha)+4\pi}{3}\right)+\frac{1}{2}\right\}$ and $\varphi_m(x) = x^{1/2}\{1-\Phi_m\}$, where $\Phi_m = \sum_{j=0}^{m}\binom{m}{j}3^j(-2)^{m-j}/(3m-j+1)$. Both the partial frontiers are increasing and log-linear in $x \in (0,1]$. They coincide if and only if $\alpha = \alpha_m = \frac{1}{2}(1-\cos[3\arccos(\frac{1}{2}-\Phi_m)-4\pi])$. For example, $\alpha_{10} = 0.9242$ and $\alpha_{20} = 0.9612$. In this case, the frontier function $\varphi_m \equiv \psi_{\alpha_m}$ can be estimated by both $\hat{\varphi}_{m,n}$ and $\hat{\psi}_{\alpha_m,n}$ as well as their projected monotone versions $\hat{\varphi}^\star_{m,n}$ and $\hat{\psi}^\star_{\alpha_m,n}$.

We simulated 5000 samples according to the Cobb-Douglas scenario to analyze the bias and the MSE of the four estimators $\hat{\varphi}_{m,n}(x)$, $\hat{\psi}_{\alpha_m,n}(x)$, $\hat{\varphi}^\star_{m,n}(x)$, $\hat{\psi}^\star_{\alpha_m,n}(x)$, for x ranging over $\{0.1, 0.2, \ldots, 0.9\}$ and for $m \in \{10, 20\}$. The sample sizes were 100 and 500. The results are reported in Table 3 for $n = 100$. Those for $n = 500$ were qualitatively similar to the case where $n = 100$, so they are omitted.

We can see that $\hat{\varphi}^\star_{m,n}(x)$ and $\hat{\psi}^\star_{\alpha_m,n}(x)$ have uniformly smaller MSEs than $\hat{\varphi}_{m,n}(x)$ and $\hat{\psi}_{\alpha_m,n}(x)$, respectively, but not by much as is expected from their asymptotic $\sqrt{n}$-equivalence in probability. In terms of bias, the monotonized estimators also perform better than the original ones, except for certain values of m and x where the squared bias is negligible compared to the value of MSE. On the other hand, the difference in their performance between the empirical order-m and order-α_m frontier functions as estimators of the same monotone function $\varphi_m = \psi_{\alpha_m}$ remains for their constrained variants.

Table 4 contains the results when the 5000 Monte Carlo samples were contaminated by an outlier at each point x. Clearly, the monotone versions $\hat{\varphi}^\star_{m,n}(x)$ and $\hat{\psi}^\star_{\alpha_m,n}(x)$ were superior to the initial estimators $\hat{\varphi}_{m,n}(x)$ and $\hat{\psi}_{\alpha_m,n}(x)$, respectively, in terms of both bias and MSE. This robustness property of the hybrid projected estimators to outliers is highly desirable.

We also performed a Monte Carlo comparison with another scenario: (X,Y) was generated from a uniform distribution on the triangle $\{(x,y)|0 \le y \le x \le 1\}$. Here, both $\psi_\alpha(x) = x(1-\sqrt{1-\alpha})$ and $\varphi_m(x) = x(1-A_m)$ are linear, where $A_m = \sum_{j=0}^{m}\binom{m}{j}2^j(-1)^{m-j}/(2m-j+1)$. They coincide when $\alpha = \alpha_m = 1-A_m^2$. For example, $\alpha_{10} = 0.9270$ and $\alpha_{20} = 0.9622$. The lessons were similar to those from the first scenario, hence the results are not reported here. The estimators $\hat{\varphi}^\star_{m,n}(x)$ and $\hat{\psi}^\star_{\alpha_m,n}(x)$ were more efficient and more resistant to anomalous data than the unrestricted $\hat{\varphi}_{m,n}(x)$ and $\hat{\psi}_{\alpha_m,n}(x)$ estimators.

Table 3: *Cobb-Douglas model,* $5,000$ *Monte-Carlo samples of size* $n = 100$.

$m = 10$

	MSE$\times 10^3$		Bias$\times 10^3$	
x	$\hat{\varphi}_{m,n}(x)$	$\hat{\varphi}^{\star}_{m,n}(x)$	$\hat{\varphi}_{m,n}(x)$	$\hat{\varphi}^{\star}_{m,n}(x)$
0.1	1.0295	1.0063	-14.8786	-14.8738
0.2	0.7675	0.7477	-10.5927	-10.4787
0.3	0.7017	0.6848	-8.4244	-8.3872
0.4	0.6657	0.6540	-7.0854	-7.0422
0.5	0.6411	0.6324	-6.4184	-6.4092
0.6	0.6017	0.5942	-5.6570	-5.6435
0.7	0.6144	0.6074	-4.9594	-4.9267
0.8	0.6155	0.6070	-5.1687	-5.1372
0.9	0.6108	0.6062	-4.6875	-4.6491

$\alpha_m = \alpha_{10} = 0.9242$

MSE$\times 10^3$		Bias$\times 10^3$	
$\hat{\psi}_{\alpha_m,n}(x)$	$\hat{\psi}^{\star}_{\alpha_m,n}(x)$	$\hat{\psi}_{\alpha_m,n}(x)$	$\hat{\psi}^{\star}_{\alpha_m,n}(x)$
1.0206	0.9643	-5.3451	-5.7296
1.0150	0.9356	-6.1262	-5.7438
0.9890	0.9307	-5.3119	-5.0679
0.9901	0.9399	-4.9526	-4.8225
1.0025	0.9677	-4.1223	-4.0929
1.0129	0.9781	-3.9310	-3.8007
1.0136	0.9836	-3.8630	-3.7386
0.9890	0.9612	-3.0135	-2.8831
0.9874	0.9559	-0.7336	-1.2649

$m = 20$

x	$\hat{\varphi}_{m,n}(x)$	$\hat{\varphi}^{\star}_{m,n}(x)$	$\hat{\varphi}_{m,n}(x)$	$\hat{\varphi}^{\star}_{m,n}(x)$
0.1	1.4078	1.3679	-22.1218	-22.0722
0.2	0.9515	0.9323	-14.5829	-14.5409
0.3	0.7839	0.7685	-11.2985	-11.3003
0.4	0.7198	0.7070	-9.3657	-9.3515
0.5	0.6642	0.6537	-8.1365	-8.1391
0.6	0.6449	0.6362	-7.6981	-7.6645
0.7	0.6069	0.5977	-7.0837	-7.0818
0.8	0.6039	0.5955	-6.3146	-6.2928
0.9	0.6312	0.6235	-6.4389	-6.4009

$\alpha_m = \alpha_{20} = 0.9612$

$\hat{\psi}_{\alpha_m,n}(x)$	$\hat{\psi}^{\star}_{\alpha_m,n}(x)$	$\hat{\psi}_{\alpha_m,n}(x)$	$\hat{\psi}^{\star}_{\alpha_m,n}(x)$
1.2505	1.2324	-17.3504	-17.2979
0.9295	0.8806	-3.0042	-3.6409
1.0688	0.9425	-10.5546	-9.6092
0.9501	0.9339	-4.9056	-5.0865
0.9998	0.9181	-4.0747	-4.1922
1.0434	0.9992	-7.6179	-6.9052
0.9360	0.9048	-1.7786	-2.2854
1.0173	0.9419	-6.8133	-6.0797
0.9547	0.9477	-4.9118	-4.8289

Table 4: *Cobb-Douglas model with 9 outliers,* $5,000$ *Monte-Carlo samples of size* $n = 109$.

$m = 10$

	MSE$\times 10^3$		Bias$\times 10^3$	
x	$\hat{\varphi}_{m,n}(x)$	$\hat{\varphi}^{\star}_{m,n}(x)$	$\hat{\varphi}_{m,n}(x)$	$\hat{\varphi}^{\star}_{m,n}(x)$
0.1	21.5602	3.5950	144.2780	57.1822
0.2	18.1104	8.1085	133.0708	88.5048
0.3	15.2115	8.4576	122.2741	90.7982
0.4	14.6204	8.8624	119.9818	93.0775
0.5	15.4200	9.7648	123.3490	97.8631
0.6	16.6763	11.2345	128.3916	105.1204
0.7	18.6371	13.0264	135.8233	113.3201
0.8	19.7960	14.7578	140.0798	120.7571
0.9	20.4651	15.8865	142.4835	125.3767

$\alpha_m = \alpha_{10} = 0.9242$

MSE$\times 10^3$		Bias$\times 10^3$	
$\hat{\psi}_{\alpha_m,n}(x)$	$\hat{\psi}^{\star}_{\alpha_m,n}(x)$	$\hat{\psi}_{\alpha_m,n}(x)$	$\hat{\psi}^{\star}_{\alpha_m,n}(x)$
44.9776	1.5383	191.2005	26.5049
15.3029	9.8166	117.5463	92.0706
5.6520	2.6643	63.2375	49.3343
5.8782	3.3670	67.1900	49.7857
4.3340	3.3848	62.5612	54.7157
4.8955	3.5222	63.7662	53.2989
4.5543	3.5433	63.5704	55.6157
5.5598	4.2718	69.2568	59.8991
5.1996	4.2869	68.3840	61.7524

$m = 20$

x	$\hat{\varphi}_{m,n}(x)$	$\hat{\varphi}^{\star}_{m,n}(x)$	$\hat{\varphi}_{m,n}(x)$	$\hat{\varphi}^{\star}_{m,n}(x)$
0.1	34.8457	5.5081	185.8157	72.4978
0.2	32.6953	14.4042	180.0621	119.4244
0.3	28.2507	16.0439	167.5092	126.0675
0.4	27.6809	16.5960	165.9073	128.3727
0.5	30.1919	18.7323	173.3699	136.4944
0.6	33.1486	22.0431	181.7376	148.1432
0.7	37.9023	26.1291	194.4146	161.3750
0.8	40.0504	29.8317	199.9093	172.4920
0.9	41.0195	31.9189	202.3611	178.4723

$\alpha_m = \alpha_{20} = 0.9612$

$\hat{\psi}_{\alpha_m,n}(x)$	$\hat{\psi}^{\star}_{\alpha_m,n}(x)$	$\hat{\psi}_{\alpha_m,n}(x)$	$\hat{\psi}^{\star}_{\alpha_m,n}(x)$
48.9564	9.4020	221.2610	94.0031
54.6848	13.0256	226.0735	112.2760
30.1573	23.9229	172.2698	151.4351
35.1036	13.2740	185.6419	111.9611
24.3216	15.4538	150.1006	123.7398
27.8948	14.3823	166.7902	115.4965
32.4609	12.8179	171.6178	113.0296
29.6673	22.2731	172.1110	146.7524
43.9957	15.4813	208.7249	124.3710

5 Data example

In this section we discuss a real data example which involves isotonic econometric functions. The dataset consisted of $n = 406$ firms in the Petroleum, Chemical and Plastics industries in Ecuador in 2002. For each firm one observed the capital K in thousands of USD, the average number of employees L and the value-added real output Y in thousands of USD. Despite its popularity, the FDH estimator $\hat{\varphi}(\cdot)$ of the production frontier has the obvious drawback of being too sensitive to isolated extremes. In this particular example, the FDH surface was determined by a small fraction (12.56%) of extreme firms, and some of these FDH firms were atypical and/or outliers. For example, the FDH frontier corresponding to 99 observations (24.38% of the data) was constant having the value 273, 000 and was determined by only one extremal firm. Clearly, such a suspicious firm influenced dramatically the FDH frontier.

In order to capture the shape of the sample boundary in a more robust way, the use of the partial order-m and order-α frontier approaches may be favored in this case. The practical question is then how to choose m and α in such a way that $\hat{\varphi}_{m,n}$ and $\hat{\psi}_{\alpha,n}$ provide reasonable estimates of large partial production functions φ_m and ψ_α lying close to φ. This could be achieved empirically by looking to Figure 1 which indicates how the percentage of the points lying outside the curves of $\hat{\varphi}_{m,n}$ and $\hat{\psi}_{\alpha(m),n}$ decreases with the order m, where $\alpha(m) = (1/2)^{1/m}$ and $m \geq 1$. The idea is to choose values of m for which the partial order-m and order-$\alpha(m)$ frontier estimators are sensitive to the magnitude of valuable extreme firms and are simultaneously resistant to the influence of the outlying observations. The evolution of the percentage should show an "L" structure whatever the studied dataset is. Looking to Figure 1, the percentage falls rapidly until the value, say, $m = 70$ and then it slows down. So, we took for our illustrations $m = 70$ to get a partial boundary lying close to the unreliable FDH frontier but well inside the sample. For the sake of conciseness, we restricted our analysis to the quantile-type frontier functions.

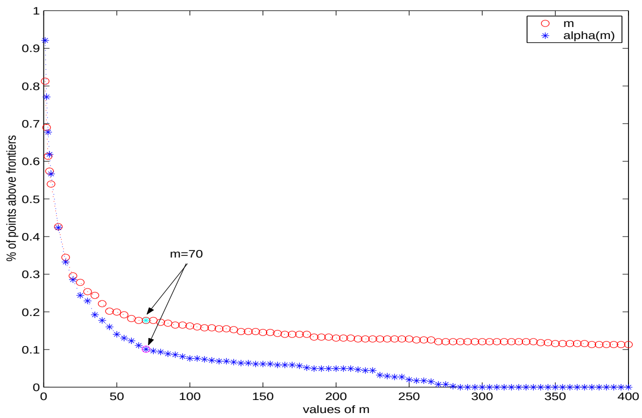

Figure 1: *Evolution of the percentage of the observations lying outside the partial order-m and order-$\alpha(m)$ frontiers as m varies (Ecuadorian manufacturing sector).*

Figure 2 (top panel) shows the resulting values $\hat{\psi}_{\alpha(m),n}(x_i)$ for 60 randomly chosen grid points among the 99 observed inputs-usage $x_i = (K_i, L_i)$ having the same FDH value $\hat{\varphi}(x_i) = 273,000$. As expected, there are many violations of monotonicity. Figure 2 (bottom panel) displays the values of $\hat{\psi}^{\star}_{\alpha(m),n}(x_i)$ for the same 60 points, showing that the partial order-$\alpha(m)$ production function is now isotonic nondecreasing. This is good news to the practitioners, whose concern, which potential outliers to eliminate before estimating the econometric function, might thus become less urgent.

For the computation of the isotonic multi-argument function $\hat{\psi}^{\star}_{\alpha(m),n}(x)$, we considered the minimal rectangular set with edges parallel to the coordinate axes covering all the observations X_i, and then chose a discrete grid $\mathbb{D}_n$ in this rectangular set. The discrete set $\mathbb{D}_n$ consisted of the observation points $\{X_i\}$ and the minimal and maximal (with respect to the partial order induced by "$\leq$") points of the minimal envelopment rectangular set.

6 Concluding Remarks

This paper contributes to the literature on isotonic estimation of a multivariate monotone function $M(\cdot)$ defined on a compact subset $\mathbb{D}$ of $\mathbb{R}^d$. We discussed an easy isotonization technique. The method yields a monotonized estimator $\hat{M}_n^{\#}$ which keeps the consistency property of the unconstrained version $\hat{M}_n$ and outperforms it in the sense that $\sup_{x\in\mathbb{D}} |\hat{M}_n^{\#}(x) - M(x)| \leq \sup_{x\in\mathbb{D}} |\hat{M}_n(x) - M(x)|$. Under the assumption that the process $r_n(\hat{M}_n - M)$ is asymptotically equicontinuous for some sequence $r_n > 0$, we show that the projected estimator $\hat{M}_n^{\#}$ is asymptotically r_n-equivalent in probability to the non-monotone estimator $\hat{M}_n$. Thus, the first-order properties of the latter are valid for the former.

A first motivating application involves a monotone estimator of the conditional distribution function, obtained by projected local linear fitting. Our simulations indicate that the hybrid estimator $\hat{F}^{\star}(\cdot|x)$ of the conditional distribution function corrects for monotonicity and improves the initial local linear estimator without losing its superior bias properties, especially near and at the support endpoints of the explanatory variable. We found that $\hat{F}^{\star}(\cdot|x)$ decreased both the MSE and the MADE. These attractive properties are particularly advantageous if one desires to invert the improved local linear estimator $\hat{F}^{\star}(\cdot|x)$ to produce a more relevant estimator of the conditional quantile function.

This article also contributes to frontier modeling by ensuring the monotonicity 'free of charge' of the empirical order-m probability weighted moment (Cazals *et al.* (2002)) and order-α quantile type (Aragon *et al.* (2005)) frontier functions. In absence of outliers, our experiments suggest that the constrained hybrid frontiers are uniformly the winners in terms of MSE and provide competitive performance even in terms of bias. In presence of outliers, it does appear that the monotonized estimators perform appreciably better than the empirical ones in terms of both bias and MSE in all cases. One way to extend our results may be to consider the regularized case by looking into the 'trimming' orders m and α as appropriate

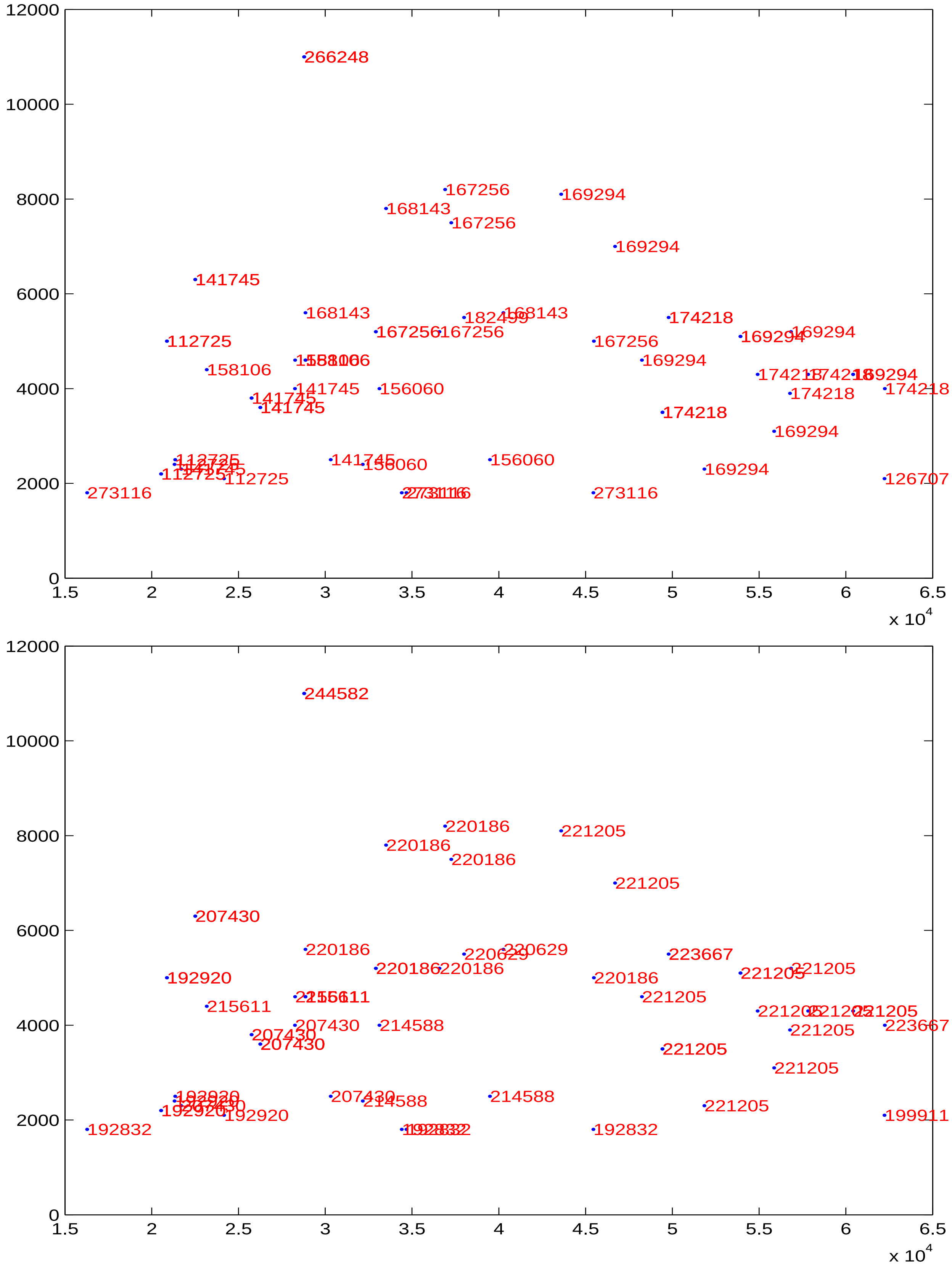

Figure 2: *Unconstrained estimates* $\hat{\psi}_{\alpha(m),n}$ *(top) and isotonic estimates* $\hat{\psi}^{\star}_{\alpha(m),n}$ *(bottom).*

functions of the sample size n. The intriguing question of whether the weak convergence of the regularized processes also holds is a topic of interest for future research.

Another interesting application, especially in the field of reliability and survival analysis, is the monotonization of Yang's (1978) estimator for the mean remaining life function. Kochar *et al.* (2000) were the first to implement the idea that a projection type estimator is 'free of charge' in the decreasing case, which is of genuine interest. They raised the question of possible extensions of the asymptotic $\sqrt{n}$-equivalence in probability from a fixed compact subinterval $\mathbb{D} = [0, b]$ to the whole positive half-line $[0, \infty)$ by making use of weighted empiricals (Hall and Wellner (1979)). Unfortunately, posing the question in this way involves some mathematical difficulties that we have not yet succeeded in overcoming.

It should be also pointed out that we restrict ourselves to rectangular regions $\mathbb{D}$ of $\mathbb{R}^d$. We do not discuss the extensions of our theorems to more general compact sets (*e.g.*, closed convex sets), but they are of interest.

Appendix: proofs.

To simplify the notation throughout this section, we write $A_{\ell,0}$ and $A_{\ell,j}$, $j = 1, \ldots, d$, for the vectors

$$A_{\ell,0} = (a_{1,\ell_1}, \ldots, a_{d,\ell_d}) \quad \text{with} \quad 0 \leq \ell_1, \ldots, \ell_d \leq k_n,$$

$$A_{\ell,j} = (a_{1,\ell_1}, \ldots, a_{j-1,\ell_{j-1}}, (a_{j,\ell_j} + \Delta_j), a_{j+1,\ell_{j+1}}, \ldots, a_{d,\ell_d}) \quad \text{with} \quad 0 \leq \ell_1, \ldots, \ell_d \leq k_n - 1.$$

Note that the definition of $L_n h$ reads then as

$$L_n h(x) = h(A_{\ell,0}) + \left(\frac{h(A_{\ell,1}) - h(A_{\ell,0})}{\Delta_1}, \ldots, \frac{h(A_{\ell,d}) - h(A_{\ell,0})}{\Delta_d} \right) (x - A_{\ell,0})$$

for $x \in \mathbb{D}_{\ell_1,\ldots,\ell_d}$ and $0 \leq \ell_1, \ldots, \ell_d \leq k_n - 1$.

Proof of Lemma 1. The monotonicity being defined with respect to the partial order induced on $\mathbb{R}^d$, it suffices to show that the piecewise linear transformation $L_n \hat{M}_n$ is nondecreasing *a.s.*, for all large n, on each partitioning set $\mathbb{D}_{\ell_1,\ldots,\ell_d}$, $0 \leq \ell_1, \ldots, \ell_d \leq k_n - 1$. Clearly this holds if for each $j = 1, \ldots, d$,

$$\hat{M}_n(A_{\ell,j}) - \hat{M}_n(A_{\ell,0}) > 0 \quad a.s. \quad \text{for all large}\, n.$$

By Condition (C4) we have $|\hat{M}_n(A_{\ell,0}) - M(A_{\ell,0})| = O_{a.s.}(s_n/r_n)$ uniformly for $0 \leq \ell_1, \ldots, \ell_d \leq k_n$, and by Condition (C1) we have for all $\ell_1, \ldots, \ell_d \leq k_n - 1$ and $j = 1, \ldots, d$,

$$M(A_{\ell,0}) - M(A_{\ell,j}) = -\Delta_j \, \partial_j M(a_{1,\ell_1}, \ldots, a_{j-1,\ell_{j-1}}, (a_{j,\ell_j} + \xi_j), a_{j+1,\ell_{j+1}}, \ldots, a_{d,\ell_d}) \leq -c_j \Delta_j$$

for some $\xi_j \in (0, \Delta_j)$. Then

$$\hat{M}_n(A_{\ell,0}) - \hat{M}_n(A_{\ell,j}) = \left[\hat{M}_n(A_{\ell,0}) - M(A_{\ell,0})\right] + [M(A_{\ell,0}) - M(A_{\ell,j})] + \left[M(A_{\ell,j}) - \hat{M}_n(A_{\ell,j})\right] \leq -c_j\Delta_j + O_{a.s.}(s_n/r_n), \quad n \to \infty.$$

It follows from $s_n/r_n = o(\Delta_j)$ that

$$\hat{M}_n(A_{\ell,0}) - \hat{M}_n(A_{\ell,j}) \leq \Delta_j \left(-c_j + o_{a.s.}(1)\right) < -c_j\Delta_j/2 \quad a.s.,$$

for all large n. This ends the proof. $\square$

Proof of Lemma 2. Let $\varepsilon > 0$ be arbitrary and choose δ to satisfy (C3). Denote by Δ the vector $(\Delta_1, \ldots, \Delta_d)^\top$. Because $k_n \uparrow \infty$, there exists $n_1 \in \mathbb{N}$ such that $\|\Delta\|_d < \delta$ for all $n \geq n_1$. Then, for all $n \geq \max(n_0, n_1)$, both terms

$$\begin{aligned} U_n &:= \max_{0 \leq \ell_1, \ldots, \ell_d \leq k_n - 1} \sum_{j=1}^{d} |\mathbb{Z}_n(A_{\ell,j}) - \mathbb{Z}_n(A_{\ell,0})| \\ W_n &:= \max_{0 \leq \ell_1, \ldots, \ell_d \leq k_n - 1} \sup_{x \in \mathbb{D}_{\ell_1, \ldots, \ell_d}} |\mathbb{Z}_n(x) - \mathbb{Z}_n(A_{\ell,0})| \end{aligned}$$

are smaller than or equal to

$$d \times \sup_{\substack{x, x' \in \mathbb{D} \\ \|x - x'\|_d \leq \|\Delta\|_d}} |\mathbb{Z}_n(x) - \mathbb{Z}_n(x')| \leq d \times \sup_{\substack{x, x' \in \mathbb{D} \\ \|x - x'\|_d < \delta}} |\mathbb{Z}_n(x) - \mathbb{Z}_n(x')|.$$

Consequently, we obtain in view of (C3)

$$U_n \xrightarrow{p} 0 \quad \text{and} \quad W_n \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty. \tag{A.1}$$

Now, define a piecewise shifted variation $V_n M$ of M on $\mathbb{D}$ by

$$V_n M(x) := M(x) + \left[\hat{M}_n(A_{\ell,0}) - M(A_{\ell,0})\right]$$

for $x \in \mathbb{D}_{\ell_1, \ldots, \ell_d}$ with $x \notin \{A_{\ell,1}, \cdots, A_{\ell,d}, (A_{\ell,0} + \Delta)\}$, with $0 \leq \ell_1, \ldots, \ell_d \leq k_n - 1$, and by

$$V_n M(x) := \hat{M}_n(x) \quad \text{for} \quad x \in \{A_{\ell,1}, \cdots, A_{\ell,d}, (A_{\ell,0} + \Delta)\}.$$

Note that $V_n M(A_{\ell,0}) = \hat{M}_n(A_{\ell,0})$ for $0 \leq \ell_1, \ldots, \ell_d \leq k_n$. Note also that

$$r_n \left[\hat{M}_n(x) - V_n M(x)\right] = r_n \left[\hat{M}_n(x) - M(x)\right] - r_n \left[\hat{M}_n(A_{\ell,0}) - M(A_{\ell,0})\right] = \mathbb{Z}_n(x) - \mathbb{Z}_n(A_{\ell,0})$$

for $x \in \mathbb{D}_{\ell_1, \ldots, \ell_d}$ with $x \notin \{A_{\ell,1}, \cdots, A_{\ell,d}, (A_{\ell,0} + \Delta)\}$. Thus

$$r_n \|\hat{M}_n - V_n M\|_{\mathbb{D}} \leq W_n. \tag{A.2}$$

On the other hand, by the definitions of L_n and V_n, we have $L_n\hat{M}_n(x) = L_nV_nM(x)$ for all $x \in \mathbb{D}$, so that

$$\hat{M}_n(x) - L_n\hat{M}_n(x) = \left[\hat{M}_n(x) - V_nM(x)\right] + \left[V_nM(x) - L_nV_nM(x)\right]. \tag{A.3}$$

Since for $x \in \mathbb{D}_{\ell_1,\ldots,\ell_d}$ with $x \notin \{A_{\ell,1}, \cdots, A_{\ell,d}, (A_{\ell,0} + \Delta)\}$ and $0 \le \ell_1, \ldots, \ell_d \le k_n - 1$,

$$\begin{aligned}&[V_nM(x) - L_nV_nM(x)] - [M(x) - L_nM(x)]\\ &\quad= \sum_{j=1}^{d} \frac{x_j - a_{j,\ell_j}}{\Delta_j} \left\{\left[M(A_{\ell,j}) - \hat{M}_n(A_{\ell,j})\right] - \left[M(A_{\ell,0}) - \hat{M}_n(A_{\ell,0})\right]\right\}\end{aligned}$$

we get $\|V_nM - L_nV_nM\|_{\mathbb{D}} \le \|M - L_nM\|_{\mathbb{D}} + U_n/r_n$. It follows from (A.2) and (A.3) that

$$r_n\|\hat{M}_n - L_n\hat{M}_n\|_{\mathbb{D}} \le W_n + r_n\|M - L_nM\|_{\mathbb{D}} + U_n$$

which converges in probability to zero by (A.1) and Lemma 3. $\square$

Proof of Lemma 3. For any n and $\ell_1, \ldots, \ell_d \le k_n - 1$, we have by Taylor expansion that, uniformly for $x \in \mathbb{D}_{\ell_1,\ldots,\ell_d}$,

$$M(x) = M(A_{\ell,0}) + \left[\partial_j\, M(A_{\ell,0})\right]^\top (x - A_{\ell,0}) + O\left(\delta_n^2\right).$$

Thus, $M(x) - L_nM(x) = \sum_{j=1}^{d}(x_j - a_{j,\ell_j})\, T_j + O\left(k_n^{-2}\right)$, where

$$T_j = \partial_j\, M(A_{\ell,0}) - \frac{M(A_{\ell,j}) - M(A_{\ell,0})}{\Delta_j}.$$

A Taylor expansion of $M(A_{\ell,j})$ with $\|\partial_{jj}^2\, M\|_{\mathbb{D}} < \infty$ for all $j = 1, \ldots, d$ gives $T_j = O(\Delta_j)$, which leads to $M(x) - L_nM(x) = O\left(\delta_n^2\right)$ uniformly for $x \in \mathbb{D}_{\ell_1,\ldots,\ell_d}$ and $\ell_1, \ldots, \ell_d \le k_n - 1$. Therefore

$$\|M - L_nM\|_{\mathbb{D}} = \max_{0 \le \ell_1,\ldots,\ell_d \le k_n - 1} \|M - L_nM\|_{\mathbb{D}_{\ell_1,\ldots,\ell_d}} = O\left(\delta_n^2\right).$$

$\square$

Proof of Lemma 4. Note first that $nh\sum_{i=1}^{n} w_i^2$ and $n^3h^3\sum_{i=1}^{n} w_i^4$ are bounded with probability tending to one. Since $F^{10}(y|u)$ is bounded for $y \in \mathbb{D}$ and for u in a neighborhood of x, there exists a constant $0 < c_1 < \infty$ such that

$$E\left[(\xi_i(y') - \xi_i(y))^2 \middle| \mathcal{X}\right] \le c_1[|y' - y| + (y' - y)^2].$$

Since $[1\!\!1(Y_i \le y) - 1\!\!1(Y_i \le y')]^4 = [1\!\!1(Y_i \le y) - 1\!\!1(Y_i \le y')]^2$, there also exists a constant $0 < c_2 < \infty$ such that

$$E\left[(\xi_i(y') - \xi_i(y))^4 \middle| \mathcal{X}\right] \le c_2[|y' - y| + (y' - y)^4].$$

Thus, there exist constants C_i, $i = 1, 2$, such that

$$
\begin{aligned}
&E\left[|\mathbb{Z}_{1n}(y') - \mathbb{Z}_{1n}(y)|^4 \middle| \mathcal{X}\right] \\
&\quad \le C_1[(y'-y)^2 + (y-y')^4]\left(nh\sum_{i=1}^{n} w_i^2\right)^2 + C_2[|y'-y| + (y-y')^4]n^2h^2\sum_{i=1}^{n} w_i^4.
\end{aligned}
$$

This implies that there exists a constant $0 < C < \infty$ such that, for all $y, y' \in \mathbb{D}$ with $n^{-1}h^{-1} \le |y - y'| \le 1$, the following inequalities hold with probability tending to one:

$$
E\left[|\mathbb{Z}_{1n}(y') - \mathbb{Z}_{1n}(y)|^4 \middle| \mathcal{X}\right] \le C(y'-y)^2. \tag{A.4}
$$

Choose a sequence $p \equiv p_n$ such that $p \ge n^{-1}h^{-1}$ and $p\sqrt{n/h} \to 0$ as n tends to infinity. This is possible since we assume h is asymptotic to $n^{-\alpha}$ for some $1/5 \le \alpha < 1/3$. Applying Theorem 12.2 of Billingsley (1968) to the partial sums $\sum_{l=1}^{i}[\mathbb{Z}_{1n}(y + lp) - \mathbb{Z}_{1n}(y + (l-1)p)]$ with (A.4), we obtain that there exists a constant C_3 such that, for any integer $m \ge 1$ possibly depending on n,

$$
P\left[\max_{0 \le i \le m} \left|\mathbb{Z}_{1n}(y + ip) - \mathbb{Z}_{1n}(y)\right| > \frac{\varepsilon}{12} \middle| \mathcal{X}\right] \le \frac{C_3}{\varepsilon^4}(mp)^2 \tag{A.5}
$$

with probability tending to one. To analyze the difference between $\mathbb{Z}(y')$ and $\mathbb{Z}(y+ip)$ when $y' \in [y + ip, y + (i+1)p]$ for some $0 \le i \le (m-1)$, define $\mathbb{Z}_{1n}^{+}(y) = \sqrt{nh}\sum_{w_i \ge 0} w_i \xi_i(y)$ and $\mathbb{Z}_{1n}^{-}(y) = \sqrt{nh}\sum_{w_i < 0} w_i \xi_i(y)$. Then, the maximal inequality (A.5) also holds for the processes $\mathbb{Z}_{1n}^{+}$ and $\mathbb{Z}_{1n}^{-}$. Furthermore, for any $y'' \in \mathbb{D}$ and $y' \in [y'', y'' + p]$, it follows that

$$
\begin{aligned}
&|\mathbb{Z}_{1n}(y') - \mathbb{Z}_{1n}(y'')| \\
&\quad \le |\mathbb{Z}_{1n}^{+}(y''+p) - \mathbb{Z}_{1n}^{+}(y'')| + |\mathbb{Z}_{1n}^{-}(y''+p) - \mathbb{Z}_{1n}^{-}(y'')| + C_4\, p\sqrt{nh}\sum_{i=1}^{n}|w_i|
\end{aligned} \tag{A.6}
$$

for some constant C_4. Since w_i are bounded by $C_5 n^{-1}h^{-1}$ for some constant C_5, the last term on the right hand side of (A.6) is less than $\varepsilon/12$ for sufficiently large n. Thus, the inequality (A.6) and the maximal inequality (A.5) together with its versions for $\mathbb{Z}_{1n}^{+}$ and $\mathbb{Z}_{1n}^{-}$ give

$$
P\left[\sup_{y \le y' \le y+mp} \left|\mathbb{Z}_{1n}(y') - \mathbb{Z}_{1n}(y)\right| > \frac{\varepsilon}{3} \middle| \mathcal{X}\right] \le \frac{3\,C_3}{\varepsilon^4}(mp)^2 \tag{A.7}
$$

with probability tending to one. Choose $\delta \le \varepsilon^5/(3C_3)$ and $m = \delta/p$. Then, the right hand side of the inequality (A.7) can be replaced by $\varepsilon\delta$. From this, one can prove

$$
P\left[\sup_{\substack{y,y' \in \mathbb{D} \\ |y'-y| \le \delta}} \left|\mathbb{Z}_{1n}(y') - \mathbb{Z}_{1n}(y)\right| > \varepsilon \middle| \mathcal{X}\right] \le \varepsilon
$$

with probability tending to one, where we assume $\mathbb{D} = [0,1]$ without loss of generality. This concludes the proof of the first part of the lemma.

For $\mathbb{Z}_{2n}$, note that $\mathbb{Z}_{2n}(y) = \sqrt{nh}\sum_{i=1}^n w_i\left[F(y|X_i) - F(y|x) - (X_i - x)F^{01}(y|x)\right]$ since $\sum_{i=1}^n w_i(X_i - x) = 0$. By a Taylor expansion, it follows that

$$\begin{aligned}\sup_{y\in\mathbb{D}}&\left|\mathbb{Z}_{2n}(y) - \sqrt{nh}\sum_{i=1}^n w_i(X_i - x)^2 F^{02}(y|x)/2\right|\\ &\leq \sqrt{nh^5}\sum_{i=1}^n |w_i| \sup_{|x'-x|\leq h}\sup_{y\in\mathbb{D}}\left|F^{02}(y|x') - F^{02}(y|x)\right|/2. \qquad \text{(A.8)}\end{aligned}$$

Since nh^5 is bounded, $\sum_{i=1}^n |w_i| = O_p(1)$, and $\sup_{y\in\mathbb{D}}|F^{02}(y|x') - F^{02}(y|x)|$ converges to zero as x' approaches to x, we have

$$\mathbb{Z}_{2n}(y) = \sqrt{nh}\sum_{i=1}^n w_i(X_i - x)^2 F^{02}(y|x)/2 + r_n(y),$$

where r_n satisfies $P\left(\sup_{y\in\mathbb{D}}|r_n(y)| > \varepsilon\right) \to 0$ for any $\varepsilon > 0$. The second part of the lemma then follows from the continuity of $F^{02}(\cdot|x)$ on $\mathbb{D}$. $\square$

Proof of Lemma 5. From (A.8) and the fact $\sum_{i=1}^n |w_i| = O_{a.s.}(1)$ resulted from an application of the strong law of large number, it follows that $\sup_{y\in\mathbb{D}}|\mathbb{Z}_{2n}(y)| = O_{a.s.}(1)$. To prove the first part of the lemma, let $\Delta_n = nh\sum_{i=1}^n w_i^2$. It follows that there exists a constant $C_6 > 0$ such that for all $a \geq 1$

$$P\left[\sup_{y\in\mathbb{D}}|\mathbb{Z}_{1n}(y)| \geq a\Delta_n^{1/2}\,\middle|\, \mathcal{X}\right] \leq C_6\, a\sqrt{\frac{n}{h}}\Delta_n^{-1/2}\exp\left(-2a^2\right). \qquad \text{(A.9)}$$

The exponential inequality (A.9) may be obtained by applying the lemma in Singh (1975) conditionally on $\mathcal{X}$. Also, one can obtain the following exponential inequality for Δ_n: there exist constants $0 < \tau_1 < \tau_2 < \infty$ and $0 < C_7, C_8 < \infty$ such that

$$P\left(\Delta_n \notin [\tau_1, \tau_2]\right) \leq C_7\exp(-C_8 nh). \qquad \text{(A.10)}$$

The two inequalities (A.9) and (A.10) imply

$$P\left[\sup_{y\in\mathbb{D}}|\mathbb{Z}_{1n}(y)| \geq a\Delta_n^{1/2}\right] \leq C_6\, a\sqrt{\frac{n}{h}}\tau_1^{-1/2}\exp\left(-2a^2\right) + C_7\exp(-C_8 nh). \qquad \text{(A.11)}$$

Choosing $a = a_n = \sqrt{\log n}$ gives $P\left[\sup_{y\in\mathbb{D}}|\mathbb{Z}_{1n}(y)| \geq \sqrt{\log n}\Delta_n^{1/2}\text{ i.o.}\right] = 0$. This together with the fact that $\Delta_n = O_{a.s.}(1)$ concludes $\sup_{y\in\mathbb{D}}|\mathbb{Z}_{1n}(y)| = O_{a.s.}(\sqrt{\log n})$. $\square$

Proof of Theorem 2. The conditions (C1) and (C2) of Theorem 1 follow from the assumption $\inf_{y\in\mathbb{D}} F^{10}(y|x) > 0$, and the one that $F^{20}(\cdot|x)$ is continuous on $\mathbb{D}$, respectively. The conditions (C3) and (C4) follow from Lemmas 4 and 5, respectively. $\square$

Proof of Theorem 3. It is enough to check the conditions (C3)-(C4) of Theorem 1. Condition (C3) follows from the weak convergence of the process $\mathbb{Z}_n$ in the space of bounded functions on $\mathbb{D}$ to the Gaussian process $\mathbb{Z}$. Condition (C4) is also satisfied with $s_n = (\log\log n)^{1/2}$. Indeed, if ν is the upper bound of the support of Y, we know from Daouia and Simar (2005, Equation (8)) that $\mathbb{P}[\hat{\varphi}(x) \le \varphi(x) \le \nu,$ for all $x \in \mathbb{D}] = 1$. Thus, we have

$$\begin{aligned}\hat{\varphi}_{m,n}(x) - \varphi_m(x) &= \int_0^\nu \Big([F(y|x)]^m - [\hat{F}_n(y|x)]^m\Big)\,dy \\ &= \int_0^\nu \Big(F(y|x) - \hat{F}_n(y|x)\Big)\sum_{j=0}^{m-1}[F(y|x)]^{m-1-j}[\hat{F}_n(y|x)]^j\,dy\end{aligned}$$

for all $x \in \mathbb{D}$, almost surely. This gives $\|\hat{\varphi}_{m,n} - \varphi_m\|_{\mathbb{D}} \le m\,\nu \sup_{(x,y)}|\hat{F}_n(y|x) - F(y|x)|$ almost surely. By the uniform law of the iterated logarithm for the empirical processes $\sqrt{n}(\hat{F}_{X,n} - F_X)$ and $\sqrt{n}(\hat{F} - F)$, we have $\sup_{(x,y)}|\hat{F}_n(y|x) - F(y|x)| = O\left((\log\log n/n)^{1/2}\right)$ almost surely, which completes the proof. $\square$

Proof of Theorem 4. The desired result follows immediately by applying Theorem 1 in conjunction with the results of Horváth, Horváth and Zhou (2008). $\square$

Proof of Theorem 5. Under the assumption that $\mathbb{E}(X^p) < \infty$ for some $p > 2$, both conditions (C3) and (C4) of Theorem 1 are fulfilled in view of (4) and (5). $\square$

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] Billingsley, P. (1968), *Convergence of Probability Measures*, John Wiley & Sons, New-York.

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

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

[5] Daouia, A. and I. Gijbels (2011), Robustness and inference in nonparametric partial frontier modeling, *Journal of Econometrics*, 161, 147–165.

[6] Daouia, A., and L. Simar (2005), Robust nonparametric estimators of monotone boundaries, *Journal of Multivariate Analysis*, 96, 311–331.

[7] Fan, J., Q. Yao and H. Tong (1996), Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems, *Biometrika*, 83, 189–206.

[8] Gijbels, I., E. Mammen, B.U. Park and L. Simar (1999), On estimation of monotone and concave frontier functions, *Journal of American Statistical Association*, 94, 220–228.

[9] Gijbels, I. and L. Peng (2000), Estimation of a Support Curve via Order Statistics, *Extremes*, 3(3), 251–277.

[10] Guess, F., and F. Proschan (1988), Mean residual life : theory and applications, In *Handbook of Statistics*, P.R. Krishnaiah and C.R. Rao, editors, North-Holland, New York, 7, 215–224.

[11] Hall, P., and L.S. Huang (2001), Nonparametric kernel regression subject to monotonicity constraints, *Annals of Statistics*, 29, 624–647.

[12] Hall, P., and H.G. Müller (2003), Order-preserving nonparametric regression with applications to conditional distribution and quantile function estimation, *Journal of American Statistical Association*, 98, 598–608.

[13] Hall, P. and B.U. Park (1998), On polynomial estimators of frontiers and boundaries, *Journal of Multivariate Analysis*, 66, 71-98.

[14] Hall, P., and J.A. Wellner (1979), Estimation of mean residual life. Unpublished manuscript.

[15] Hall, P., R.C.L. Wolff and Q. Yao (1999), Methods for estimating a conditional distribution function, *Journal of American Statistical Association*, 94, 154–163.

[16] Hendricks, W., and R. Koenker (1992), Hierarchical spline models for conditional quantiles and the demand for electricity, *Journal of American Statistical Association*, 87, 58–68.

[17] Horváth, L., Z. Horváth and W. Zhou (2008), Asymptotic properties of nonparametric frontier estimators, *Econometric Theory*, 24, 1607–1627.

[18] Kochar, S.C, H. Mukerjee and F.J. Samaniego (2000), Estimation of a monotone mean residual life, *Annals of Statistics*, 28, 905–921.

[19] Lee, Y.K., E.R. Lee and B.U. Park (2006), Conditional quantile estimation by local logistic regression, *Journal of Nonparametric Statistics*, 18, 357–373.

[20] Mammen, E. and K. Yu (2007), Additive isotone regression, *IMS Lecture Notes-Monograph Series*, 55, 179–195.

[21] Mukerjee, H. (1993), An improved monotone conditional quantile estimator, *Annals of Statistics*, 21, 924–942.

[22] Mukerjee, H. and S. Stern (1994), Feasible nonparametric estimation of multiargument monotone functions, *Journal of American Statistical Association*, 89, 77–80.

[23] Nishiyama, Y. (2011), Impossibility of weak convergence of kernel density estimators to a non-degenerate law in $L_2(\mathbb{R}^d)$, *Journal of Nonparametric Statistics*, 23, 129-135.

[24] Park, B.U., L. Simar and Ch. Weiner (2000), The FDH estimator for productivity efficiency scores: Asymptotic properties, *Econometric Theory*, 16, 855-877.

[25] Ruymgaart, F.H. (1998), A note on weak convergence of density estimators in Hilbert spaces, *Statistics*, 30, 331-343.

[26] 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, Oxford University Press.

[27] Singh, R.S. (1975), On the Glivenko-Cantelli theorem for weighted empiricals based on independent random variables, *Annals of Probability*, 3, 371–374.

[28] Yang, G.L. (1978), Estimation of a biometric function, *Annals of Statistics*, 6, 112–116.

[29] Yu, K. and M.C. Jones (1998), Local linear quantile regression, *Journal of American Statistical Association*, 93, 228–237.