

New designs to consistently estimate the isotonic regression

Ana Colubi
J. Santos Dominguez-Menchero
Gil Gonzalez-Rodriguez

Received: / Accepted:

Abstract The usual estimators of the regression under isotonicity assumptions are too sensitive at the tails. In order to avoid this problem, some new strategies for fixed designs are analyzed. The uniform consistency of certain estimators on a closed and bounded working interval are obtained. It is shown that the usual isotonic regression can be employed when the number of observations at the edges of the interval is suitably controlled. Moreover, two modifications are proposed which substantially improve the results. One modification is based on the reallocation of part of the edge observations, and the other one forces the isotonic regression to take values within some horizontal bands. The theoretical results are complemented with some examples and simulation studies that illustrate the performance of the proposed estimators in practice.

Keywords Isotonic regression · Estimation · Uniform consistency · Designs

1 Introduction

The problems of estimating and testing about the regression function when it is isotonic have been widely analyzed in the literature. Comprehensive books on these topics are [3] and [20]. The natural estimator in this context is the so-called isotonic regression, defined as the isotonic function that better fits the data with respect to the weighted empirical L_2 distance.

A. Colubi · J.S. Dominguez-Menchero · G. Gonzalez-Rodriguez
Dpto. de Estadística e I.O. y D.M.,
Universidad de Oviedo 33007 Oviedo, Spain
E-mail: colubi@uniovi.es

J.S. Dominguez-Menchero
E-mail: jsdm@uniovi.es

G. Gonzalez-Rodriguez
E-mail: gil@uniovi.es

In [4] it is proved that this estimator is uniformly consistent in a closed and compact interval strictly contained in the design interval under general conditions and fixed design. In [12] this result is improved by relaxing the moment assumptions in the preceding work, but restricting the situation handled in [4] to the particular case of uniform weights. However, it is well-known that the isotonic regression has a *spiking problem*, and it is usually too sensitive at the tails [18].

In order to alleviate the spiking problem, a more robust estimator could be considered, e.g., one based on the L_1 distance [24] or the M-estimators [1]. In [26] an estimator with good properties for testing isotonicity is considered. This estimator is based on a penalized loglikelihood function. The idea is to modify the response observations at the first and last design points, and then to obtain the corresponding isotonic regression. Some properties of consistency can be derived from [18]. Nevertheless, none of these estimators have been proved to be uniformly consistent under fully random sampling in general.

As an alternative, we explore ways to guarantee the consistency of the usual isotonic regression estimator by proposing some ways of designing the experiment. Specifically, we derive a generalization of the well-known consistently result in Brunk (1970) by allowing different weighting functions, and we show how to apply this result, in conjunction with an appropriate fixed design, to prove the uniform consistency avoiding, in this way, the spiking problem. Moreover, we will propose a slight modification of the original estimator (to restrict it to a band) that guarantees the uniform consistency with quite weak requirements on the number of observations on the tails. In details, we consider three approaches:

1. To suitable choose the number of observations at the first and last design point. In this case, we prove that the usual isotonic regression is uniformly consistent. In particular, we extend the results in [4] and [12] to the whole working interval and general weights;
2. To modify the first approach by reallocating part of the observations considered at the first and last design point at some artificial adjacent points;
3. To force the isotonic regression to take values within some horizontal band [8].

Guaranteeing the uniform consistency can be specially valuable in situations like those in [21], as the behaviour of the tails becomes under control. It is also important for deriving consistent estimators of the conditional variance function in heteroscedastic models. This could allow us to improve the different tests about isotonicity (see for instance, [5], [6], [26]).

It will be theoretically shown that the magnitude order of sample size required to obtain the consistency of the third estimator is lower than the necessary for the other two estimators. The way of distributing the total number of observations according to each design in practice is explained by considering some simulations. The empirical results indicate that the proposed methods outperform the classical approach and the estimators in [24], [1] and [26]. The second and the third estimators perform better in general.

The rest of the paper is organized as follows. In Section 2, notation concerning the isotonic regression is introduced. The basic conditions to obtain the previous consistency results are recalled, and the behaviour of the classical isotonic estimator at the tails is illustrated by means of some examples. The three new strategies to estimate the regression under isotonicity are presented in Section 3, and the conditions to guarantee the consistency are established. In Section 4, the experimental results are shown. In Section 5, extensions to the multidimensional case is discussed. Finally, the proofs are included in the Appendix.

2 Regression model and notation

We consider the regression model $Y(x) = m(x) + \varepsilon(x)$ where $x \in A \subseteq \mathbb{R}$ and the errors $\varepsilon(x)$ have 0 mean. The regression function m is assumed to be continuous and isotonic.

In order to estimate m , we consider a fixed design $\{x_{1,n}, \dots, x_{n,n}\} \subset A$ with $x_{i,n} < x_{j,n}$ ($1 \leq i < j \leq n$, $n \in \mathbb{N}$) and $r_n(i)$ independent observations $Y^1(x_{i,n}), \dots, Y^{r_n(i)}(x_{i,n})$ on each design point $x_{i,n}$ ($1 \leq i \leq n$, $n \in \mathbb{N}$). The observations $\{Y^j(x_{i,n})\}$ constitute a triangular array of row-wise independent random variables.

Consider a non-negative weighting function $w : A \rightarrow \mathbb{R}^+$ and let \mathcal{T}_n^I be the set of real isotonic functions defined on $\{x_{1,n}, \dots, x_{n,n}\}$. The argmin of

$$\sum_{i=1}^n \sum_{j=1}^{r_n(i)} w(x_{i,n}) (Y^j(x_{i,n}) - f(x_{i,n}))^2$$

on \mathcal{T}_n^I is the well-known isotonic regression estimator \widehat{m}_I , which can be computed by means of PAVA (see [2], and also [13] for an implementation in R).

In this ‘‘classical setting’’ the obtained estimator is just defined on the fixed design points. This situation has been widely analyzed in the literature both from the estimating as well as the testing view points (see, for instance, [4], [5], [6], [7], [9], [10], [17], [21], [25], [26]). As usual in this setting, this regression estimator can be defined on the whole set A by considering any isotonic extension. In this way, we will denote by \widehat{m}_I^* any isotonic extension of \widehat{m}_I on A verifying that $\widehat{m}_I^*(x) = \widehat{m}_I(x_{1,n})$ if $x \leq x_{1,n}$ and $\widehat{m}_I^*(x) = \widehat{m}_I(x_{n,n})$ if $x \geq x_{n,n}$.

As an alternative to the classical setting, the function to be minimized has been proposed to be modified by adding a penalty term in order to take into account the smoothness of the regression function (see, for instance, [15], [19], [23], [27]) or even different constraints are considered (see, for instance, [16], [22]). The introduction of a penalty term leads in many cases to a solution defined on the whole set A for which no isotonic extension between design points is needed. It is also possible to consider a continuous version of the isotonic regression problem (see, for instance, [11], [14]) by using an integrable weight function defined on the whole A . All these modifications rely on additional

assumptions on the smoothness of m . The main aim here is to analyze ways of designing the experiment making the classical isotonic regression estimator to be consistent (alternatives 1 and 2 mentioned in the introduction). Additionally, alternative 3 introduces a slight modification of the isotonic regression estimator (by simply adding a restriction to the original minimization problem in order to directly control the behaviour of the estimator at the tails). Nevertheless, no assumptions on the smoothness of m are imposed.

As in [12], we will consider $F : [0, \infty) \rightarrow [0, 1]$ to be the function given by

$$F(y) = \sup_{x \in A} P(|\varepsilon(x)| > y), \text{ for all } y \in [0, \infty).$$

In addition, $N_n(J)$ will stand for the number of observations at those design points contained in $J \subseteq A$ for each $n \in \mathbb{N}$. That is,

$$N_n(J) = \sum_{\{i: x_{i,n} \in J\}} r_n(i).$$

In this way, $N_n(A)$ will be the overall sample size.

We will consider the following conditions proposed in [12]:

- (C1) m is continuous and isotonic on A .
- (C2) $\lim_{y \rightarrow \infty} F(y) = 0$; $\int_0^\infty y |dF(y)| < \infty$.
- (C3) For all $x \in A$ so that $(x, \infty) \cap A \neq \emptyset$, $\min\{\{x_{1,n}, \dots, x_{n,n}\} \cap (x, \infty)\}$ tends to x as $n \rightarrow \infty$ and, analogously, for all $x \in A$ so that $(-\infty, x) \cap A \neq \emptyset$, $\max\{\{x_{1,n}, \dots, x_{n,n}\} \cap (-\infty, x)\}$ tends to x as $n \rightarrow \infty$.
- (C4) $\limsup_{n \rightarrow \infty} N_n(A)/N_n(J) < \infty$ for all $J \subseteq A$.

Conditions (C1) and (C2) are only connected with the regression model, while conditions (C3) and (C4) refer to the design. Specifically, (C3) indicates that the set of design points tends to be dense in A as n tends to infinity. Condition (C4) balances the ratio between the sample size per design point and the overall sample size (to avoid too much sample information concentrated in a subinterval of A). In particular, if $A = [0, 1]$ and considering the set of equally spaced design points $x_{i,n} = i/(n+1)$ for all $i \in \{1, \dots, n\}$ and $n \in \mathbb{N}$, then condition (C3) is fulfilled. Moreover, if $r_n(i) = 1$, then condition (C4) is also satisfied.

Under conditions (C1)-(C4), [12] proved the uniform consistency of \widehat{m}_I^* (with constant weighting function) on any closed and bounded subset $B \subsetneq A$. However, as it is well-known that the behaviour of the estimator at the tails can be poor. To illustrate this fact, in Figure 1 and 2 we have chosen $A = [-1, 1]$, $m(x) = 1$, $w(x) = 1$, $\varepsilon(x)$ distributed as a $\mathcal{N}(0, 1)$ for all $x \in A$ and equally spaced design points. We see that if $r_n(i) = 1$ for all $i \in \{1, \dots, n\}$, then the estimates are not suitable at the tails even if n is large (Situations a). However, with the same total sample sizes, the behaviour improves if more observations per design point at the edges are considered (Situations b). This fact is connected with the results in Section 3.

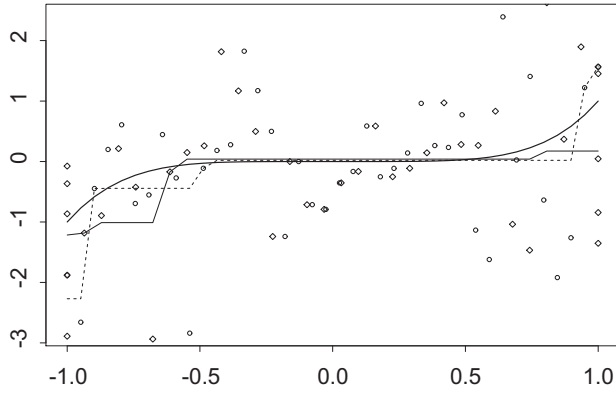


Fig. 1 Regression function (thick line). Situation a1): $n = 40$ and $r_{40}(i) = 1$ for all $i \in \{1, \dots, 40\}$. Simulated points (\circ), estimates (dash lines). Situation b1): $n = 32$, $r_{32}(1) = 5$, $r_{32}(32) = 5$ and $r_{32}(i) = 1$ for all $i \in \{2, \dots, 31\}$. Simulated points (\diamond), estimates (thin lines)

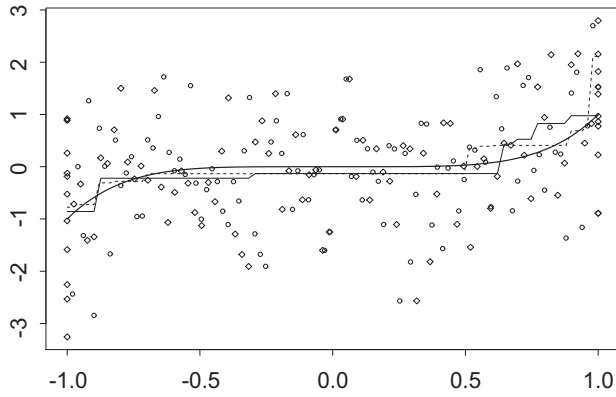


Fig. 2 Regression function (thick line). Situation a1): $n = 100$ and $r_{100}(i) = 1$ for all $i \in \{1, \dots, 100\}$. Simulated points (\circ), estimates (dash lines). Situation b1): $n = 80$, $r_{80}(1) = 11$, $r_{80}(80) = 11$ and $r_{80}(i) = 1$ for all $i \in \{2, \dots, 79\}$. Simulated points (\diamond), estimates (thin lines)

3 The results

As a starting point, in next theorem we generalize the consistency result in [12] by considering more general weighting functions. In order to achieve the result the following condition on the weighting function proposed by [4] will also be considered:

- **(C5)** w is bounded and bounded away from 0.

Obviously, (C5) is verified by the constant weighting function used in [12].

Theorem 1 *If conditions (C1)-(C5) are satisfied, then for each closed and bounded subinterval $B \subsetneq A$,*

$$\sup_{x \in B} |\widehat{m}_I^*(x) - m(x)| \xrightarrow{n \rightarrow \infty} 0 \quad a.s. - [P].$$

In this section we propose three different approaches to consistently estimate the regression function under isotonicity in a working interval $A = [a, b]$.

The first procedure, **P1**, is based on a design in which the number of observations at the ending design points is as follows:

$$- \text{(C6)} \quad \limsup_{n \rightarrow \infty} N_n(A) / \min(r_n(1), r_n(n)) < \infty.$$

The next theorem states the uniform consistency on the whole A of any isotonic extension \widehat{m}_1^* of $\widehat{m}_1 = \widehat{m}_I$ verifying that $\widehat{m}_1^*(x) = \widehat{m}_1(x_{1,n})$ if $x \leq x_{1,n}$ and $\widehat{m}_1^*(x) = \widehat{m}_1(x_{n,n})$ if $x \geq x_{n,n}$.

Theorem 2 *If conditions (C1)-(C5) and condition (C6) are satisfied, then*

$$\sup_{x \in A} |\widehat{m}_1^*(x) - m(x)| \xrightarrow{n \rightarrow \infty} 0 \quad a.s. - [P].$$

The second procedure, **P2**, is based on a simple reallocation of the observations at the ending points that allows us to make use of Theorem 1 in order to guarantee the consistency. For each $n \in \mathbb{N}$ let $k_n^a, k_n^b \in \mathbb{N}$ and consider k_n^a independent observations $Z_a^1, \dots, Z_a^{k_n^a}$ distributed as $Y(a)$ and k_n^b independent observations $Z_b^1, \dots, Z_b^{k_n^b}$ distributed as $Y(b)$. Let \widehat{m}_2 be the isotonic function computed from the design points $B_n = \cup_{j=1}^{k_n^a} \{a - j/k_n^a\} \cup \{x_{1,n}, \dots, x_{n,n}\} \cup \cup_{j=1}^{k_n^b} \{b + j/k_n^b\}$, associating the observation Z_a^j with the point $a - j/k_n^a$ for each $j = 1, \dots, k_n^a$, the observation Z_b^j with the point $b + j/k_n^b$ for each $j = 1, \dots, k_n^b$, and the rest of the observations with the original points. That is, \widehat{m}_2 is the argmin of

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^{r_n(i)} w(x_{i,n}) (Y^j(x_{i,n}) - f(x_{i,n}))^2 \\ & + \sum_{j=1}^{k_n^a} w(a) (Z_a^j - f(a - j/k_n^a))^2 + \sum_{j=1}^{k_n^b} w(b) (Z_b^j - f(b + j/k_n^b))^2 \end{aligned}$$

on the set of isotonic functions defined on B_n . The estimator \widehat{m}_2 is just an isotonic regression, then it can be computed by means of the PAVA algorithm in the usual way.

Remark 1 We have kept the notation for the number of observations at the original design points for the sake of simplicity. However, it should be underlined that we propose to **reallocate** the observations obtained at the ending points, not to obtain extra observations at the new (artificially considered) design points. This fact will be illustrated in Section 4. \square

Under the next condition about the number of reallocated observations, it is possible to prove the uniform consistency on the whole A of any isotonic extension \widehat{m}_2^* of \widehat{m}_2 .

$$- \text{(C7)} \quad \limsup_{n \rightarrow \infty} \frac{\max(k_n^a, k_n^b)}{N_n(A)} < \infty; \quad \limsup_{n \rightarrow \infty} \frac{N_n(A)}{\min(k_n^a, k_n^b)} < \infty.$$

The first part of condition (C7) along with condition (C4) implies that the latter condition is also satisfied for the overall sample size $N_n(A) + k_n^a + k_n^b$ (and not only for $N_n(A)$).

Theorem 3 *If conditions (C1)-(C5) and condition (C7) are satisfied, then*

$$\sup_{x \in A} |\widehat{m}_2^*(x) - m(x)| \xrightarrow{n \rightarrow \infty} 0 \quad \text{a.s.} - [P].$$

Conditions (C6) and (C7) state almost the same in connection with the observations considered for the ending design points. Actually, the number of observations required at the ending design points are of the same order. The main difference is as follows. In the first case, all the information at the ending points is used together in order to obtain a global consistent estimator of the regression function. In contrast, in the second one, different estimators can be used at the boundaries. In general, the reallocated observations considered at the tails are used to control the behaviour of the estimator there, by constraining in some sense the possible values.

Lastly, we consider an alternative design, **P3**, that allows us to reduce the number of observations at the ending design points with respect to the previous designs. Based on this design we propose a new estimator of the isotonic regression by combining in some sense the main ideas of the previous situations.

On the one hand, the estimator will use all the information together for obtaining a global estimator (as in the first situation). On the other hand, the idea of constraining the possible values of the estimator is handled by directly restricting the isotonic regression to take values in an appropriate region as follows:

For any $i \in \{1, \dots, n\}$, let

$$\overline{Y}(x_{i,n}) = \frac{1}{r_n(i)} \sum_{j=1}^{r_n(i)} Y^j(x_{i,n}),$$

$$\overline{Y} = \frac{1}{\sum_{i=1}^n r_n(i)} \sum_{i=1}^n \sum_{j=1}^{r_n(i)} Y^j(x_{i,n}),$$

and let \widehat{m}_3 be the isotonic regression estimator restricted to take values in

$$[\min(\overline{Y}(x_{1,n}), \overline{Y}), \max(\overline{Y}, \overline{Y}(x_{n,n}))].$$

That is, \widehat{m}_3 is the argmin of

$$\sum_{i=1}^n \sum_{j=1}^{r_n(i)} w(x_{i,n}) (Y^j(x_{i,n}) - f(x_{i,n}))^2$$

on the set of isotonic functions defined on \mathcal{T}_n^I and taking values restricted to $[\min(\overline{Y}(x_{1,n}), \overline{Y}), \max(\overline{Y}, \overline{Y}(x_{n,n}))]$ (see [8]). Thus, the uniform consistency on the whole A of any isotonic extension \widehat{m}_3^* of \widehat{m}_3 can be proved by considering the following condition concerning the ending design points:

- (C8) $\lim_{n \rightarrow \infty} \min(r_n(1), r_n(n)) = \infty$.

Theorem 4 *If conditions (C1)-(C5) and condition (C8) are satisfied, then*

$$\sup_{x \in A} |\widehat{m}_3^*(x) - m(x)| \xrightarrow{n \rightarrow \infty} 0 \quad a.s. - [P].$$

Condition (C7) obviously implies condition (C8). Nevertheless, there are many possible ways of choosing the number of observations at the ending design points verifying conditions (C4) and (C8) for which condition (C7) fails. In this way, if $A = [0, 1]$, and the situation with equally spaced design points introduced in Section 2 is considered, we have that

- If $r_n(1) = r_n(n) = n$, then conditions (C4) and (C7) are satisfied, and consequently condition (C8) is satisfied.
- If $r_n(1) = r_n(n) = \lceil \log(n) \rceil$, then conditions (C4) and (C8) are satisfied but (C7) is not satisfied.

In order to clarify the differences between the approaches, we show a simulated situation in Example 1 in which $P1$, $P2$ and $P3$ are applied.

Example 1 We consider the heteroscedastic regression model represented in Figure 3(a), where $A = [1, 6]$ and $m(x)$ is the cubic-shaped regression function represented by a solid black line. The errors are Gaussian with increasing variance in A . The ‘intensity’ of the conditional density function at each point is represented by means of a color scale.

In order to estimate the regression function, $N = 20$ observations are taken at random from this model according to the different designs proposed in Section 3. Concretely, we have chosen $n = 16$ equally spaced design points in A (with $x_1 = 1$ and $x_n = 6$) in all cases. For $P1$ and $P3$, $r_n(i) = 1$ for all $i \in \{2, \dots, n-1\}$ and $r_n(1) = r_n(n) = 3$ (see Figure 3 (c)). Regarding $P2$, we have selected $r_n(i) = 1$ for all $i \in \{1, \dots, n\}$, and we have fixed $k_n^a = k_n^b = 2$. Then, the total number of observations for the conditional distribution at the point $x_1 = 1$ (respectively $x_n = 6$) is 3 in all procedures, although for $P2$ one of the observations is reallocated at 6.5, and other one at 7 (respectively 0.5 and 0), as shown in Figure 3(d). This is equivalent to extending the regression model to the interval $[0, 7]$ in a constant way (Figure 3(b)), and taking one observation of the conditional distribution at 6.5 and 7 (respectively 0.5 and 0).

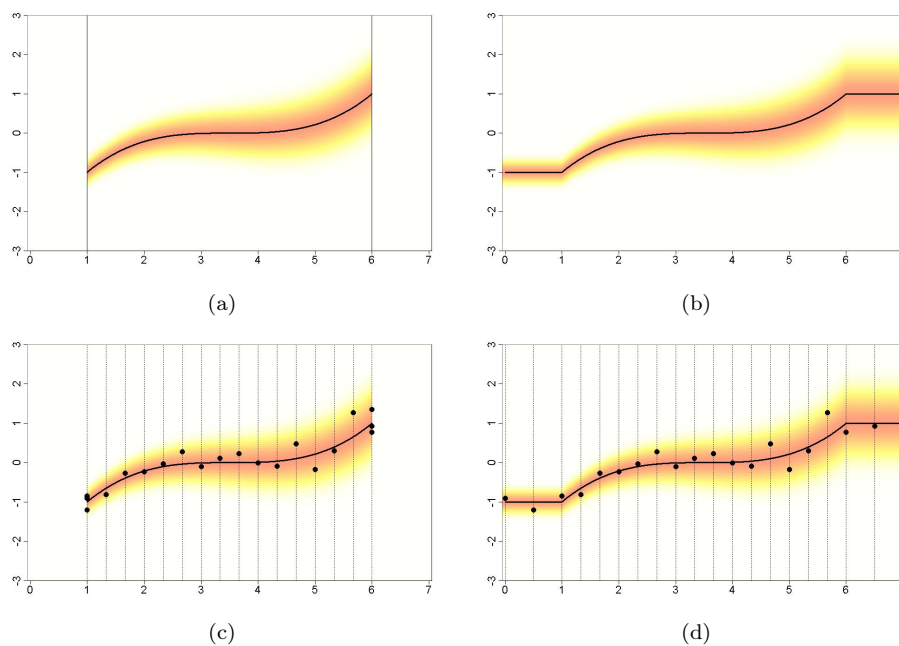


Fig. 3 A simulated example. (a) Regression model. (b) Extended regression model. (c) Fixed design for P_1 , P_2 and P_3 , and observations for P_1 and P_3 . (d) Fixed design for the extended regression model and reallocated observations for P_2 .

The estimate in P_1 is just the isotonic regression of the corresponding data (Figure 4). For P_2 , firstly the isotonic regression for the reallocated data is computed (Figure 5(a)); secondly, the estimate is obtained by constraining the overall estimate to the working interval A (Figure 5(b)).

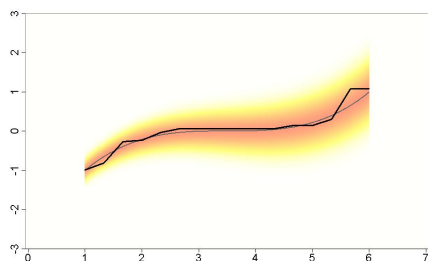


Fig. 4 A simulated example. Estimate for P_1 .

Finally, for P_3 , firstly a band to constrain the values of the estimator (by basically averaging the observations at the ending design points) is considered (Figure 6(a)); secondly the isotonic regression restricted to this band is computed as estimate of the regression (Figure 6(b)).

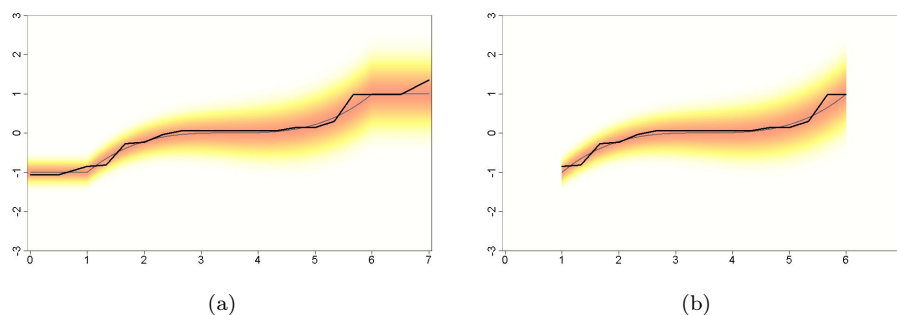


Fig. 5 A simulated example. (a) Estimate in the extended interval for $P2$. (b) Final estimate for $P2$.

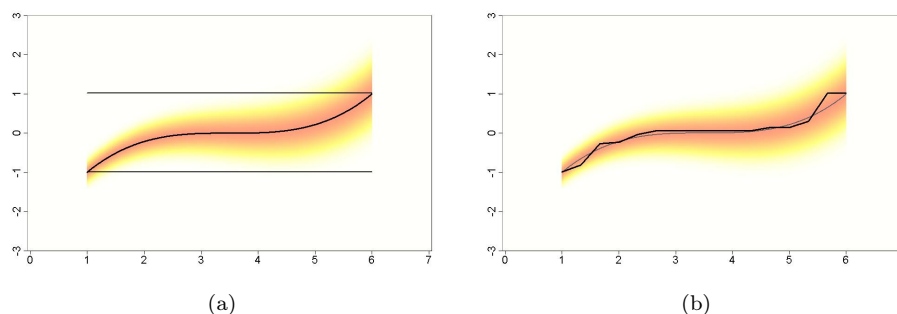


Fig. 6 A simulated example. (a) Band of restricted regression values for $P3$. (b) Estimate for $P3$.

A comparison of the three obtained estimates is shown in Figure 7. As expected, the differences between the estimates mainly concern the tails. In this particular example, the same result is given for $P1$ and $P3$ in the left hand tail, although far from the one obtained with $P2$. Regarding the right hand tail, the results of the three methods are different. \square

4 Simulations

Simulations are carried out in order to compare the 3 proposed procedures ($P1, P2, P3$) for uniformly estimating an isotonic regression function. At the same time, they are compared to the classical isotonic regression estimator with one observation per design point ($P0$), the median isotonic regression ($L1$) of [24], the M-estimator (M) of [1], and the penalized estimator (W) of [26]. As mentioned in the introduction, $L1$ and M are robust estimators, whereas W attempts to alleviate the spiking effect by modifying the observations at the boundaries of A .

For all the simulations we have fixed $A = [-1, 1]$. We have considered $m(x) = x^p$ for values of p in $\{0, 1, 3, 5\}$ in order to cover different situations regarding the increasing rate of m , as well as its behaviour on the tails. We have chosen identically distributed errors with different distributions, namely,

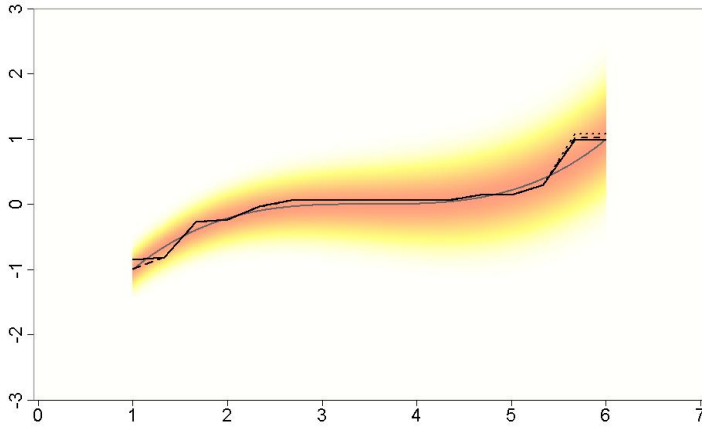


Fig. 7 A simulated example. Comparison of the different procedures ($P1$ in dotted line, $P2$ in solid line and $P3$ in dashed line).

centred $\mathcal{N}(0, 1)$, $U(-1, 1)$, $Exp(1)$, $Beta(0.3, 0.2)$ and Student's t with 3 and 2 degrees of freedom, respectively denoted by D_1, \dots, D_6 . Obviously conditions (C1) and (C2) holds. Note that D_1, D_2, D_5 and D_6 are symmetric. D_2 is bounded and its density is equally distributed over its support. D_4 is also bounded, but asymmetric and its density is concentrated near the boundaries of its support. D_5 does not have finite variance, and it is used in [1] to illustrate the robustness of the M-estimator. D_6 , however, has finite variance and, thus, lighter tails than D_5 . Finally D_3 is asymmetric with higher density near 0.

We have considered the same overall sample size N for all the procedures and equally spaced design points, particularly $x_{i,n} = 2(i-1)/(n-1) - 1$ for any $i \in \{1, \dots, n\}$ and any $n \in \mathbb{N}$. For $P0, L1$ and W , $n = N$ and $r_n(i) = 1$ are taken for all $i \in \{1, \dots, N\}$ (Situation introduced in Section 2). For $P1$ and $P3$, $r_n(1) = r_n(n) = \lceil N/10 \rceil + 1$ with $n = N - 2\lceil N/10 \rceil$ and $r_n(i) = 1$ for all $i \in \{2, \dots, n-1\}$, and for $P2$, $k_n^a = k_n^b = \lceil N/10 \rceil$ with $n = N - 2(\lceil N/10 \rceil)$ and $r_n(i) = 1$ for all $i \in \{1, \dots, n\}$. In this way, the number of observations considered at the boundary of A is the same for all the procedures. Conditions (C3) and (C4) hold, as well as the corresponding conditions (C6), (C7) and (C8) in each case. Finally, we select $w(x) = 1$ for all $x \in A$, so condition (C5) is also fulfilled.

For the penalized isotonic regression, W , firstly the isotonic regression is used in order to compute $\hat{\sigma}_n$, the standard deviation of the residuals $Y^1(x_{i,N}) - \hat{m}(x_{i,N})$, $i \in \{1, \dots, N\}$. It should be noted that by the independence assumption, only $\gamma(0)$ must be estimated in Equation (11) of [26]. Once $\hat{\sigma}_n$ is computed, the observations $Y^1(x_{1,N})$ and $Y^1(x_{N,N})$ are replaced by $Y^1(x_{1,N}) + 0.15 \hat{\sigma}_n \sqrt{N}$ and $Y^1(x_{N,N}) - 0.15 \hat{\sigma}_n \sqrt{N}$ respectively. The estimator W is just the isotonic regression of the modified data set.

For each procedure, the estimation of the regression function has been extended to the whole interval A by linear interpolation between design points (being constant in the tails). As the interest relies on the uniform consistency of the estimators and its behaviour on the tails, the supremum norm error (SNE) is considered. Note that this value is usually attained at the tails due to the spiking problem. The SNE has been approximated on the basis of a grid of size $G = 1000$, $z_i = (i - 1)/(G - 1)$ for $i = 1, \dots, G$. Specifically, if m^* stands for any of the obtained isotonic estimators,

$$SNE(m^*) = \sup_{i=1, \dots, G} |m^*(z_i) - m(z_i)|.$$

Its expected value has been estimated by Monte Carlo method on the basis of 10000 simulations for all procedures except for M for which only 200 simulations were obtained. In this respect it should be noted that the dependence of the M estimator on $max - min$ formulae involves a high computational effort in order to compute the M estimator. In the following tables a point-wise estimator of the expected value of SNE in the different scenarios is collected. In addition, a percentile bootstrap confidence interval with 1000 iterations was also computed and used to rank the estimators by determining those with smaller expected value of SNE in each case (highlighted in blue).

$N = 20$						
	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$
M	0.888	0.538	0.751	0.401	1.344	1.773
$P0$	0.894	0.519	0.842	0.357	1.363	2.016
$L1$	0.925	0.563	0.872	0.411	1.302	1.760
W	0.602	0.335	0.584	0.226	0.974	1.525
$P1$	0.557	0.321	0.536	0.221	0.882	1.322
$P2$	0.468	0.274	0.452	0.192	0.730	1.056
$P3$	0.456	0.265	0.439	0.182	0.701	1.017
$N = 100$						
	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$
M	0.926	0.505	0.897	0.400	1.134	1.773
$P0$	0.889	0.518	0.861	0.354	1.389	2.055
$L1$	0.922	0.564	0.899	0.416	1.324	1.742
W	0.367	0.202	0.400	0.139	0.663	1.142
$P1$	0.309	0.180	0.307	0.125	0.506	0.838
$P2$	0.234	0.136	0.233	0.094	0.381	0.627
$P3$	0.230	0.133	0.230	0.093	0.372	0.594
$N = 500$						
	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$
$P0$	0.893	0.518	0.855	0.357	1.387	2.066
$L1$	0.926	0.565	0.897	0.420	1.329	1.755
W	0.191	0.109	0.208	0.075	0.365	0.729
$P1$	0.151	0.088	0.150	0.061	0.253	0.439
$P2$	0.109	0.063	0.108	0.043	0.183	0.320
$P3$	0.108	0.063	0.107	0.043	0.180	0.304

Table 1 Expected value of SNE. Case $p = 0$.

$N = 20$						
	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$
M	1.088	0.683	0.969	0.476	1.432	1.822
$P0$	1.040	0.648	1.002	0.458	1.561	2.313
$L1$	1.074	0.700	1.005	0.490	1.477	1.975
W	0.876	0.568	0.879	0.427	1.309	1.959
$P1$	0.829	0.561	0.817	0.423	1.192	1.724
$P2$	0.799	0.555	0.785	0.422	1.102	1.519
$P3$	0.827	0.562	0.805	0.423	1.126	1.480
$N = 100$						
	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$
M	0.997	0.586	0.907	0.437	1.319	1.923
$P0$	0.946	0.574	0.909	0.402	1.493	2.164
$L1$	0.982	0.631	0.957	0.463	1.409	1.818
W	0.567	0.375	0.599	0.294	0.936	1.435
$P1$	0.531	0.375	0.531	0.300	0.765	1.112
$P2$	0.515	0.371	0.514	0.298	0.713	0.996
$P3$	0.523	0.371	0.523	0.297	0.744	0.993
$N = 500$						
	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$
$P0$	0.917	0.539	0.859	0.373	1.425	2.192
$L1$	0.952	0.588	0.917	0.445	1.351	1.842
W	0.340	0.235	0.358	0.186	0.550	1.039
$P1$	0.331	0.239	0.331	0.192	0.462	0.701
$P2$	0.327	0.238	0.326	0.192	0.448	0.644
$P3$	0.326	0.237	0.326	0.191	0.452	0.660

Table 2 Expected value of SNE. Case $p = 1$.

According to Tables 1, 2 and 3, the estimators $P2$ and $P3$ provide, in general, the best results in all the scenarios. Also, the estimator $P1$ is very competitive, being one of the best ones in some cases. It is shown that for the three proposed estimators, the expected SNE reduces as the sample size increases, as it is expected according to Theorems 2, 3 and 4.

The robust estimators $L1$ or M outperform the isotonic regression when the distribution of the errors has heavy tails, nevertheless, they are not useful to face the spiking problem in general. On the other hand, the estimator W exhibits a good behaviour in all scenarios, being the estimated expected SNE decreasing as the sample size increases. It is particularly good when the distribution of the errors has bounded support ($D2$ and $D4$), and concretely, in this situation when $p = 1$ it slightly outperforms the proposed procedures in some cases. This situation changes when the regression is constant ($p = 0$) or when it has a high increasing rate ($p = 5$). This behaviour can be related with the selection of the penalizing constant $c = 0.15$, that could be somehow tuned depending on the conditional distributions, and the increasing rate of the regression at the tails. In this respect it has to be mentioned that in the proposed procedures it was fixed a 10% of sampling effort in each terminal point that can also be tuned to improve the results.

Overall, according to the obtained results, it is not advisable to use the robust estimators $L1$ or M in order to face the spiking problem. The estimator

$N = 20$						
	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$
M	1.079	0.687	0.961	0.484	1.429	1.745
$P0$	1.077	0.684	1.023	0.475	1.592	2.192
$L1$	1.098	0.711	1.018	0.488	1.518	1.902
W	0.925	0.636	0.921	0.492	1.344	1.839
$P1$	0.841	0.579	0.821	0.429	1.211	1.657
$P2$	0.830	0.598	0.806	0.442	1.124	1.489
$P3$	0.831	0.575	0.799	0.428	1.128	1.430
$N = 100$						
	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$
M	1.026	0.632	0.906	0.454	1.398	1.695
$P0$	0.993	0.610	0.958	0.431	1.506	2.212
$L1$	1.023	0.661	0.987	0.472	1.418	1.876
W	0.662	0.467	0.696	0.379	0.971	1.480
$P1$	0.600	0.442	0.593	0.357	0.819	1.153
$P2$	0.592	0.443	0.585	0.359	0.775	1.029
$P3$	0.580	0.433	0.574	0.353	0.776	1.022
$N = 500$						
	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$
$P0$	0.941	0.556	0.888	0.392	1.436	2.126
$L1$	0.977	0.610	0.938	0.455	1.362	1.801
W	0.432	0.316	0.448	0.258	0.640	1.013
$P1$	0.410	0.304	0.407	0.249	0.538	0.752
$P2$	0.405	0.303	0.404	0.248	0.529	0.703
$P3$	0.401	0.301	0.399	0.247	0.519	0.698

Table 3 Expected value of SNE. Case $p = 5$.

W could be interesting to be used, particularly when it is not possible to design the experiment in the directions proposed for procedures $P1$, $P2$ or $P3$. Finally, the three proposed estimators offer a solution to handle the spiking problem and have a good behaviour in all the analysed scenarios, being, in general, $P2$ and $P3$ the best ones.

To finalise the analysis, a detailed comparison of the proposed estimators $P1$, $P2$ and $P3$ has been carried out on the basis of the performed simulations. In Figures 8, 9 and 10 we have summarized the distributions of SNE by means of box-plots.

When the regression is constant, the empirical distributions of the SNE associated with both $P2$ and $P3$ are usually strictly dominated by the one associated with $P1$, which implies that $P1$ is worse. Regarding $P2$ and $P3$, the higher quantiles (approximately those in $[0.85, 1]$) of the empirical distribution of SNE for $P2$ are smaller than the corresponding ones of $P3$. The difference is larger when the distribution of the errors has non-bounded support. On the contrary, the median of the empirical distribution for $P3$ is almost always smaller than the median for $P2$.

On the other hand, when the regression function is non-constant, the situation is rather different. The empirical distribution of SNE for $P1$ again dominates almost always the corresponding distribution for $P3$, and also the one associated with $P2$ when the error distribution has non-bounded support,

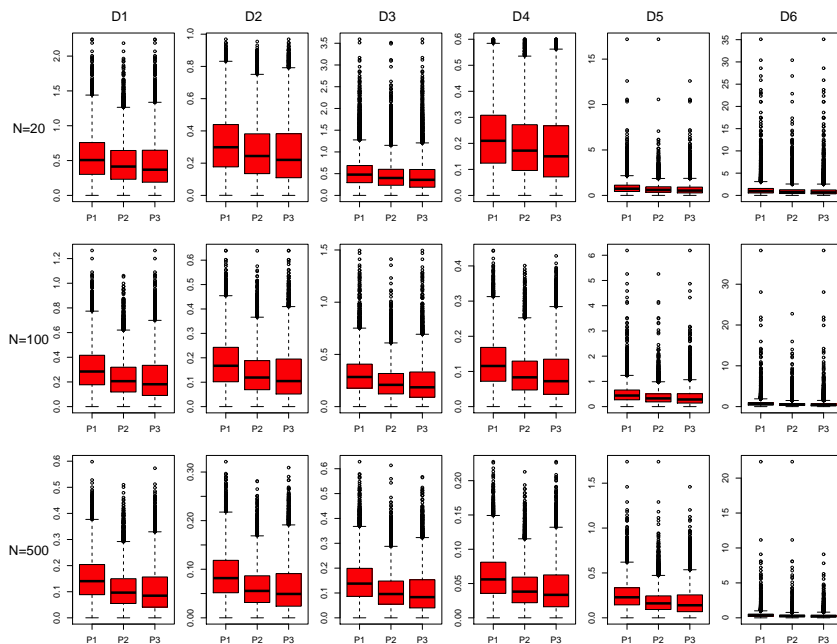


Fig. 8 Distributions of SNE for constant regression ($p = 0$)

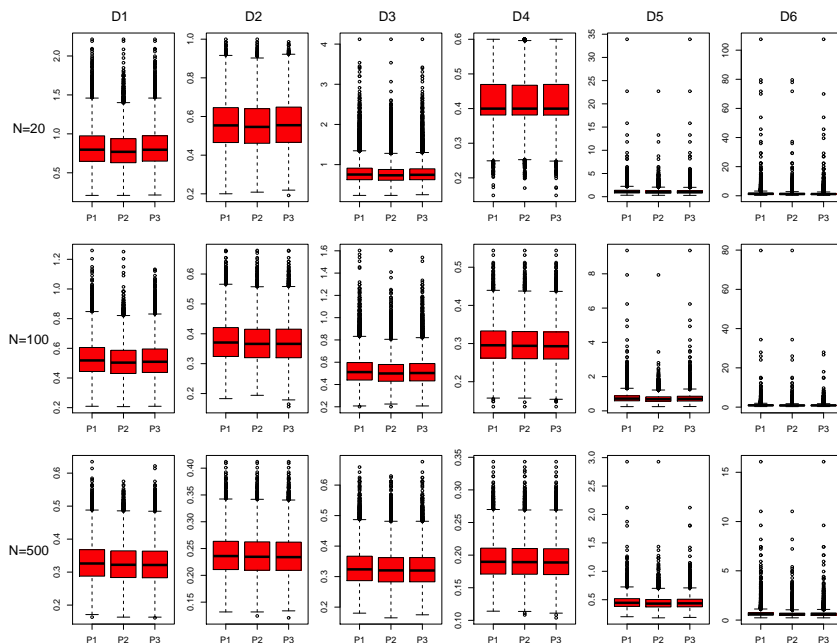


Fig. 9 Distributions of SNE for linear regression ($p = 1$)

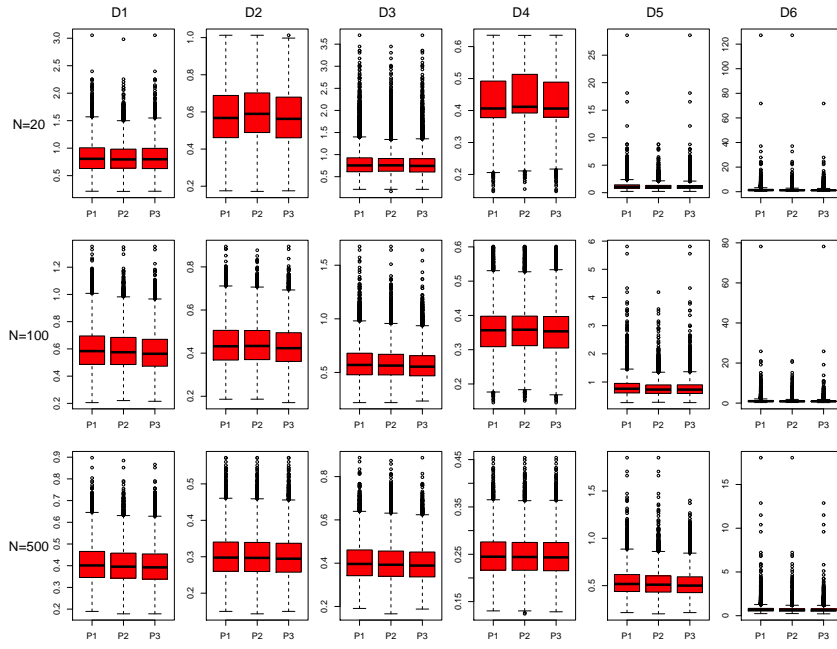


Fig. 10 Distributions of SNE for non-constant regression ($p = 5$)

although the difference is not so important. In general, for bounded errors cases, the empirical distributions of SNE for $P2$ and $P3$ are almost identical in the linear case, whereas for the case of marked slopes ($p = 5$) the quantiles corresponding to $P3$ are slightly smaller than the ones associated with $P2$. For the unbounded case, if $N = 20$ or $N = 100$ we can see a slight improvement of $P2$ with respect to $P3$, although this difference is balanced (and reversed in some cases) when $N = 500$.

To summarize, in general, it seems that $P2$ is slightly better than $P3$ (and both outperform $P1$). The difference between both procedures is clearer in the constant regression case. On the other hand, if the regression is non constant, $P2$ seems to be preferable to $P3$ only when few observations are available and the distribution of the errors has non-bounded support, while in those cases with marked slopes at the tails, $P3$ outperforms $P2$. This means that, in general, using the isotonic regression to stabilize the behaviour of the estimator at the tails is comparable to using the overall mean of the data at the tails. The main difference relies on the fact that theoretically $P3$ allows for fewer observations at the boundaries of A than $P2$, although in this analysis, for comparative purposes, exactly the same number of observations in a and b was considered.

5 Possible Extensions

Some ideas regarding the generalization of the obtained results to higher dimensions are outlined. Concretely we focus on the 2-dimensional Euclidean space case considered in [12], where some rates on the consistency of the isotonic regression in a closed and compact “interval” strictly contained in the design “interval” are obtained under certain conditions. These conditions are quite more complex than those presented for the 1-dimensional case, so we refer to the reader to Section 4 of [12].

As we have concluded from the simulation study, $P2$ and $P3$ are in general comparable and better than $P1$, so we will focus our comments on $P2$ and $P3$. For the sake of simplicity, we will consider $A = [0, 1] \times [0, 1]$ and a uniform weighting function as in [12].

Concerning $P2$, the situation can be handled by applying the ideas in Section 3 together with the trick used in the proof of Theorem 3 in the following way. First of all, the working rectangle A is slightly expanded (how much is irrelevant as in the 1-dimensional case), say $\tilde{A} = [-1, 2] \times [-1, 2]$, and consider the function $\Pi : \tilde{A} \rightarrow A$ that associates to each $(\tilde{x}_1, \tilde{x}_2) \in \tilde{A}$ the closest point in A . In this way, the original regression model $Y(x_1, x_2) = m(x_1, x_2) + \varepsilon(x_1, x_2)$ can be extended to \tilde{A} as follows:

$$\tilde{Y}(\Pi(\tilde{x}_1, \tilde{x}_2)) = m(\Pi(\tilde{x}_1, \tilde{x}_2)) + \varepsilon(\Pi(\tilde{x}_1, \tilde{x}_2)).$$

Consider a set of design points in \tilde{A} so that the corresponding conditions in Section 4 of [12] are fulfilled (these conditions regard the distribution of the errors in the model, as well as the set of design points chosen). Then, Theorems 5 and 6 in [12], providing the corresponding rates of consistency for the isotonic regression, can be applied. It should be noted that expanding the working rectangle A only affects the design points in the following way: given any design point $(\tilde{x}_1, \tilde{x}_2)$ not belonging to A , it is associated with the point $\Pi(\tilde{x}_1, \tilde{x}_2)$ in the frontier of A , so according to the mentioned conditions, this leads to an “elevated” concentration of design points at the frontier of A , with a particular “extra” concentration at the corners of A (the same happened in the 1-dimensional case presented in Section 3). In particular, the number of design points in any frontier segment J of A excluding the corners (like for instance $J = [a, b] \times \{0\}$ with $0 < a < b < 1$) verifies that $\liminf_{n \rightarrow \infty} N_n(J)/N_n(A) > 0$ and, in addition, every corner C of A (as for instance $C = \{0\} \times \{0\}$) verifies that $\liminf_{n \rightarrow \infty} N_n(C)/N_n(A) > 0$.

The key idea in procedure $P3$ is firstly to consider an appropriate design allowing the derivation of a uniformly consistent estimator of the regression function at the frontier of A . And, secondly, to use the estimate as a restriction in the isotonic regression problem (so that the values of the isotonic regression at the “tails” are under control). One of the fundamental points in this strategy is that it is not necessary to allocate too many sample information in the frontier of A , as condition (C8) is just a guarantee for the convergence of the “frontier estimator”. Consequently, we can proceed as follows: Fix the amount

of sample information to be used for the “frontier estimator“ (any function of the sample size growing to infinity as the sample size increases is enough, as for instance $p \times N_n(A)$ with $0 < p < 1$ or even lower rates as an amount proportional to $\log N_n(A)$), the remaining points (it is recommendable to keep an amount proportional to the overall sample size $N_n(A)$) are allocated in the interior of A so that the corresponding conditions in Section 4 of [12] are fulfilled for any rectangle in the interior of A .

Now we face the main problem: to define a feasible restriction set through a consistent estimator of m in the frontier, by allocating the design points in the frontier of A in an appropriate way. First of all, consider the mapping $g : [0, 1] \rightarrow \{0\} \times [0, 1] \cup [0, 1] \times \{0\}$ defined for each $x \in [0, 1]$ as

$$g(x) = (0, 2x)I_{[0,1/2]}(x) + (2x - 1, 1)I_{(1/2,1]}.$$

Obviously, for the “transformed” regression model $\tilde{Y}(x) = Y(g(x))$, the regression function $\tilde{m}(x) = m(g(x))$ (with $x \in [0, 1]$) is a monotone non-decreasing function and, consequently, we can consider any of the design strategies proposed in Section 3 in order to estimate uniformly \tilde{m} by means of \hat{m}_1 . For the other part of the frontier of A ($[0, 1] \times \{0\} \cup \{0\} \times [0, 1]$), we proceed in a similar way, allocating the rest of sample points and obtaining the estimator \hat{m}_2 . It should be noted that the corners $(0, 0)$ and $(1, 1)$ are common to both parts, so the sample points allocated there (if any) are also shared. Now, we define the estimator of m in the frontier of A as follows:

$$\hat{g}(x_1, x_2) = \begin{cases} \min\{\hat{m}_1(1/2 + x_1/2), \hat{m}_2(x_1/2)\} & \text{if } 0 < x_1 < 1, x_2 = 0, \\ \min\{\hat{m}_1(x_2/2), \hat{m}_2(1/2 + x_2/2)\} & \text{if } 0 < x_2 < 1, x_1 = 0, \\ \hat{m}_1(1/2 + x_1/2) & \text{if } 0 < x_1 < 1, x_2 = 1, \\ \hat{m}_2(1/2 + x_2/2) & \text{if } 0 < x_2 < 1, x_1 = 1, \\ \min\{\hat{m}_1(0), \hat{m}_2(0)\} & \text{if } x_1 = x_2 = 0, \\ \min\{\hat{m}_1(1), \hat{m}_2(1)\} & \text{if } x_1 = x_2 = 1. \end{cases}$$

The uniform consistency of \hat{m}_1 and \hat{m}_2 ensures the uniform consistency in the frontier of A of \hat{g} . In addition, by construction, it is easy to check that \hat{g} is a monotone and non-decreasing function. Finally, we define the estimator of m on A as the corresponding isotonic regression restricted by \hat{g} , that is \hat{m} is the argmin of

$$\sum_{i=1}^n \sum_{j=1}^{r_n(i)} w(x_{i,n}) (Y^j(x_{i,n}) - f(x_{i,n}))^2$$

on the set of isotonic functions (in the 2-dimensional Euclidean Space) defined on the set of design points $\{x_{n,1}, \dots, x_{n,n}\}$ and so that

$$\max\{\hat{g}(x_{i,n}^1, 0), \hat{g}(0, x_{i,n}^2)\} \leq f(x_{i,n}^1, x_{i,n}^2) \leq \min\{\hat{g}(x_{i,n}^1, 1), \hat{g}(1, x_{i,n}^2)\}.$$

This solution can be found by considering the extension of the isotonic regression problem analyzed in [8].

Acknowledgements

The research in this paper has been partially supported by MTM2013-44212-P, GRUPIN14-005 and the COST Action IC1408.

6 Appendix. Proofs

The following lemma is a slightly generalization of Lemma 3 in [12] that takes into account weights. We will focus on the part of this lemma that will be used for the different proofs.

Lemma 1 *Let F be a real-valued function on $[0, \infty)$ satisfying condition (C2) and let $\{Z_{i,n}\}_{i=1}^n$ ($n \in \mathbb{N}$) be a triangular array of row-wise independent random variables so that $E(Z_{i,n}) = 0$ and $P(|Z_{i,n}| \geq z) \leq F(z)$ for all $i = 1, \dots, n$, $n \in \mathbb{N}$ and $z \geq 0$. Let $\{w_{i,n}\}_{i=1}^n$ ($n \in \mathbb{N}$) be a family of positive and bounded real numbers, then*

$$\max_{1 \leq k \leq n} \left| \sum_{i=1}^k Z_{i,n} w_{i,n} \right| / n \xrightarrow{n \rightarrow \infty} 0 \quad a.s. - [P].$$

Proof: Fix $0 < M < \infty$ so that $w_{i,n} \leq M$ for all $i = 1, \dots, n$ and $n \in \mathbb{N}$. Obviously,

$$P(|Z_{i,n} w_{i,n}| \geq z) \leq F_M(z) = F(z/M)$$

for all $z \geq 0$. In addition, it is easy to check that both $\lim_{z \rightarrow \infty} F_M(z) = 0$ and $\int_0^\infty z |dF_M(z)| < \infty$. Consequently, we can apply Lemma 3 in [12] to the triangular array of row-wise independent random variables $\{Z_{i,n} w_{i,n}\}_{i=1}^n$ ($n \in \mathbb{N}$) to prove the result.

Proof Theorem 1: As in [12], the result can be proven by combining the reasoning of Theorem 4.1 of [4] with Lemma 1.

Proof Theorem 2: As Conditions (C1)-(C5) are satisfied, Theorem 1 guarantees the pointwise $a.s. - [P]$ convergence of \hat{m}_I^* in (a, b) . Thus, in order to prove the result, it is enough to check the pointwise $a.s. - [P]$ convergence in a and b .

First of all, we focus on the pointwise $a.s. - [P]$ convergence in a . In this sense, taking into account the well-known max-min formula and the isotonicity of m , it is easy to check that

$$|\hat{m}_I^*(a) - m(a)| \leq |m(x_{1,n}) - m(a)| + \max_{k=1, \dots, n} \left| \frac{\sum_{i=1}^k \sum_{j=1}^{r_n(i)} w(x_{i,n}) \varepsilon^j(x_{i,n})}{\sum_{i=1}^k w(x_{i,n}) r_n(i)} \right|.$$

The continuity of m ensures that the first term in the preceding sum converges to 0 as $n \rightarrow \infty$. On the other hand, the second term is bounded by

$$\left(\frac{1}{\inf_{x \in A} w(x)} \right) \left(\frac{N_n(A)}{r_n(1)} \right) \max_{k=1, \dots, n} \left| \frac{\sum_{i=1}^k \sum_{j=1}^{r_n(i)} w(x_{i,n}) \varepsilon^j(x_{i,n})}{N_n(A)} \right|$$

which converges *a.s.* $- [P]$ to 0 as $n \rightarrow \infty$ as a consequence of Lemma 1 and Conditions (C5) and (C6). Consequently, the pointwise *a.s.* $- [P]$ convergence in a follows. The pointwise *a.s.* $- [P]$ convergence in b can be deduced in the same way.

Proof Theorem 3: By considering $A = [a - 1, b - 1]$, $m(x) = m(a)$, $w(x) = w(a)$ and $\varepsilon(x)$ distributed as $\varepsilon(a)$ for all $x \in [a - 1, a)$ (analogous at the point b) and the set of design points B_n , we have that the conditions (C1)-(C5) are satisfied. Thus, the result follows directly from Theorem 1.

In order to prove the next two theorems, it should be noted that the increasing of \hat{m}_I^* together with Conditions (C1) and (C3) guarantee that it is enough to prove the pointwise *a.s.* $- [P]$ convergence in $[a, b]$.

Proof Theorem 4: Let $I_n = \min(\overline{Y(x_{1,n})}, \overline{Y})$ and $S_n = \max(\overline{Y}, \overline{Y(x_{n,n})})$. Based on [8] it can be deduced that

$$\hat{m}_3^*(x) = \max\{\min\{\hat{m}_I^*(x), S_n\}, I_n\}$$

for all $x \in [a, b]$ for certain isotonic extension \hat{m}_I^* of the non-restricted isotonic regression \hat{m}_I . In addition, $\hat{m}_3^*(a) = I_n$ and $\hat{m}_3^*(b) = S_n$ which, taking into account Condition (C8) and the SLLN, converges *a.s.* $- [P]$ to $m(a)$ and $m(b)$, respectively. Finally, note that as Conditions (C1)-(C5) are satisfied, Theorem 1 guarantees the pointwise *a.s.* $- [P]$ convergence of \hat{m}_I^* to m in (a, b) , and consequently the pointwise *a.s.* $- [P]$ convergence of \hat{m}_3^* to m in $[a, b]$ is obtained, which finishes the proof.

References

1. Alvarez, E., Yohai, V.: M-estimators for isotonic regression. *Journal of Statistical Planning and Inference* **142**, 2351–2368 (2012)
2. Ayer, M., Brunk, H.D., Ewing, W.T., Reid, W.T., Silverman, E.: An empirical distribution function for sampling with incomplete information. *The Annals of Mathematical Statistics* **26**, 641–647 (1955)
3. Barlow, R.E., Bartholomew, D., Bremner, J.M., Brunk, H.D.: *Statistical Inference Under Order Restrictions: The Theory and Applications of Isotonic Regression*. Wiley, New York (1972)
4. Brunk, H.D.: *Nonparametric Techniques in Statistical Inference*, chap. Estimation of isotonic regression, pp. 177–195. Cambridge University Press, London (1970)
5. Colubi, A., Domínguez-Menchero, J.S., González-Rodríguez, G.: Testing constancy for isotonic regressions. *Scandinavian Journal of Statistics* **33**, 463–475 (2006)
6. Colubi, A., Domínguez-Menchero, J.S., González-Rodríguez, G.: A test for constancy of isotonic regressions using the L_2 -norm. *Statistica Sinica* **17**, 713–724 (2007)
7. Cuesta-Albertos, J.A., Domínguez-Menchero, J.S., Matrán, C.: Consistency of L_p -best monotone approximations. *Journal of Statistical Planning and Inference* **47**, 295–318 (1995)

8. Domínguez-Menchero, J.S., González-Rodríguez, G.: Analyzing an extension of the isotonic regression problem. *Metrika* **66**, 19–30 (2007)
9. Domínguez-Menchero, J.S., González-Rodríguez, G., López-Palomo, M.J.: An L_2 point of view in testing monotone regression. *Nonparametric Statistics* **17**, 135–153 (2005)
10. Domínguez-Menchero, J.S., López-Palomo, M.J.: On the estimation of monotone uniform approximations. *Statistics and Probability Letters* **35**, 355–362 (1995)
11. Groeneboom, P., Jongbloed, G.: Generalized continuous isotonic regression. *Statistics and Probability Letters* **80**, 248–253 (2010)
12. Hanson, D.L., Pledger, G., Wright, F.T.: On consistency in monotonic regression. *Annals of Statistics* **1**, 401–421 (1973)
13. de Leeuw, J., Hornik, K., Mair, P.: Isotone optimization in R: Pool-adjacent-violators algorithm (PAVA) and active set methods. *Journal of Statistical Software* **32**, 1–24 (2009)
14. Mammen, E.: Estimating a smooth monotone regression function. *Annals of Statistics* **19**, 724–740 (1991)
15. Mammen, E., Thomas-Agnan, C.: Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics* **26**, 239–252 (1999)
16. Meyer, M., Woodroffe, M.: On the degrees of freedom in shape-restricted regression. *Annals of Statistics* **28**, 1083–1104 (2000)
17. Mukerjee, H.: Monotone nonparametric regression. *Annals of Statistics* **16**, 741–75 (1988)
18. Pal, J.K.: Spiking problem in monotone regression: Penalized residual sum of squares. *Statistics and Probability Letters* **78**, 1548–1556 (2008)
19. Pal, J.K., Woodroffe, M.: Large sample properties of shape restricted regression estimators with smoothness adjustments. *Statistica Sinica* **17**, 1601–1616 (2007)
20. Robertson, T., Wright, F.T., Dykstra, R.L.: *Order Restricted Statistical Inference*. John Wiley & Sons, New York (1988)
21. Sampson, A.R., Singh, H., Whitaker, L.: Simultaneous confidence bands for isotonic functions. *Journal of Statistical Planning and Inference* **139**, 828–842 (2009)
22. Tibshirani, R.J., Hoefling, H., Tibshirani, R.: Nearly-isotonic regression. *Technometrics* **53**, 54–61 (2011)
23. Wang, X., Li, F.: Isotonic smoothing spline regression. *Journal of Computational and Graphical Statistics* **17**, 281–287 (2008)
24. Wang, Y., Huang, J.: Limiting distribution for monotone median regression. *Journal of Statistical Planning and Inference* **107**, 281–287 (2002)
25. Wright, F.T.: The asymptotic behavior of monotone regression estimates. *Annals of Statistics* **9**, 443–448 (1981)
26. Wu, W.B., Woodroffe, M., Mentz, G.: Isotonic regression: Another look at the change-point problem. *Biometrika* **88**, 793–804 (2001)
27. Yeganova, L., Wilbur, W.J.: Isotonic regression under lipschitz constraint. *Journal of Optimization Theory and Applications* **141**, 429–443 (2009)