



ELSEVIER

Contents lists available at SciVerse ScienceDirect

## Journal of Statistical Planning and Inference

journal homepage: [www.elsevier.com/locate/jspi](http://www.elsevier.com/locate/jspi)

## M-estimators for isotonic regression

Enrique E. Álvarez<sup>a,\*</sup>, Víctor J. Yohai<sup>b</sup><sup>a</sup> University of La Plata and CONICET, Argentina<sup>b</sup> University of Buenos Aires and CONICET, Argentina

## ARTICLE INFO

## Article history:

Received 13 March 2011

Received in revised form

20 February 2012

Accepted 21 February 2012

Available online 28 February 2012

## Keywords:

Isotonic regression

M-estimators

Robust estimates

## ABSTRACT

In this paper we propose a family of robust estimates for isotonic regression: isotonic M-estimators. We show that their asymptotic distribution is, up to a scalar factor, the same as that of Brunk's classical isotonic estimator. We also derive the influence function and the breakdown point of these estimates. Finally we perform a Monte Carlo study that shows that the proposed family includes estimators that are simultaneously highly efficient under Gaussian errors and highly robust when the error distribution has heavy tails.

© 2012 Published by Elsevier B.V.

## 1. Introduction

Let  $x_1, \dots, x_n$  be independent random variables collected along observation points  $t_1 \leq \dots \leq t_n$  according to the model:

$$x_j = \mu(t_j) + u_j, \quad (1)$$

where the  $u_j$ 's are i.i.d. random variables with distribution  $G$ . In *isotonic regression* the trend term  $\mu(t)$  is monotone non-decreasing, i.e.,  $\mu(t_1) \leq \dots \leq \mu(t_n)$ , but it is otherwise arbitrary. In this set-up, the classical estimator of  $\mu(t)$  is the function  $g$  which minimizes the  $L_2$  distance between the vector of observed and fitted responses, i.e, it minimizes

$$\sum_{j=1}^n [x_j - g(t_j)]^2 \quad (2)$$

in the class  $\mathcal{G}$  of non-decreasing piecewise continuous functions. It is trivial but noteworthy that Eq. (2) posits a finite dimensional convex constrained optimization problem. Its solution was first proposed by Brunk (1958) and has received extensive attention in the statistical literature (see, e.g., Robertson et al., 1988 for a comprehensive account). It is also worth noting that any piecewise continuous non-decreasing function which agrees with the optimizer of (2) at the  $t_j$ 's will be a solution. For that reason, in order to achieve uniqueness, it is traditional to restrict further the class  $\mathcal{G}_0$  to the subset of piecewise constant non-decreasing functions. Another valid choice consists of the interpolation at the knots with non-decreasing cubic splines or any other piecewise continuous monotone function, e.g., Meyer (1996). We will call this estimator the  $L_2$  isotonic estimator.

\* Corresponding author.

E-mail addresses: [eealvarez@mate.unlp.edu.ar](mailto:eealvarez@mate.unlp.edu.ar), [enriquealvarez@fibertel.com.ar](mailto:enriquealvarez@fibertel.com.ar) (E.E. Álvarez).

The sensitivity of this estimator to extreme observations (outliers) was noted by Wang and Huang (2002), who propose minimizing instead using the  $L_1$  norm, i.e., minimizing

$$\sum_{j=1}^n |x_j - g(t_j)|.$$

This estimator will be called here  $L_1$  isotonic estimator. Wang and Huang (2002) developed the asymptotic distribution of the trend estimator at a given observation point  $t_0$  and obtained the asymptotic relative efficiency of this estimator compared with the classical  $L_2$  estimator. Interestingly, this efficiency turned out to be  $2/\pi = 0.637$ , the same as in the i.i.d. location problem.

In this paper we will propose instead a robust isotonic M-estimator aimed at balancing robustness with efficiency. Specifically we shall seek the minimizer of

$$\sum_{j=1}^n \rho\left(\frac{x_j - g(t_j)}{\hat{\sigma}_n}\right), \quad (3)$$

where  $\hat{\sigma}_n$  is an estimator of the error scale previously obtained and  $\rho$  satisfies the following properties:

A1 (i)  $\rho(x)$  is non-decreasing in  $|x|$ , (ii)  $\rho(0) = 0$ , (iii)  $\rho$  is even, (iv)  $\rho(x)$  is strictly increasing for  $x > 0$ , (v)  $\rho$  has two continuous derivatives and  $\psi = \rho'$  is bounded and monotone non-decreasing, and (vi)  $E(\psi(u/\sigma_0)) = 0$ .

Note that in the case that  $E(\psi(u/\sigma_0)) = c \neq 0$ , Assumption A1 (vi) holds changing the trend function  $\mu$  by  $\mu + c$ . If we assume errors with symmetric distribution, A1 (vi) holds automatically when  $\psi$  is even. Clearly, the  $L_2$  choice corresponds to taking  $\rho(x) = x^2$  while the  $L_1$  option is akin to opting for  $\rho(x) = |x|$ . These two estimators do not require the scale estimator  $\hat{\sigma}_n$ . Note also that the class of M-estimators satisfying A1 does not include estimators with a redescending choice for  $\psi$ . We believe that the strict differentiability conditions on  $\rho$  required in A1 are not strictly necessary, but they make the proofs for the asymptotic theory simpler. Moreover, some functions  $\rho$  which are not twice differentiable everywhere such as  $|x|$  or the Huber's functions defined below in (7) can be approximated by functions satisfying A1.

The asymptotic distribution of the  $L_2$  isotonic estimators at a given point was found by Brunk (1970) and Wright (1981) and the one of the  $L_1$  estimator by Wang and Huang (2002). They prove that the distribution of these estimators conveniently normalized converge to the distribution of the slope at zero of the greatest convex minorant of the two-sided Brownian Motion with parabolic drift. In this paper, we prove a similar result for isotonic M-estimators. The focus of this paper is on estimation of the trend term at a single observation point  $t_0$ . We do not address the issue of distribution of the whole stochastic process  $\{\hat{\mu}_n(t), t \in \mathcal{T}\}$ . Recent research along those lines are given by Kulikova and Lopuhaä (2006) and a related result with smoothing was also obtained simultaneously in Pal and Woodroffe (2006). The isotonic regression problem has been also addressed with modern techniques of empirical processes. We refer the reader here to Van de Geer (2000, p. 58), who shows the uniform consistency of the classical isotonic regression estimator within a class of functions utilizing methods based on entropy with bracketing.

In order to place this paper within the Statistical Literature, it is worth mentioning that the literature on estimation of a monotone regression function is by now enormous. In an attempt to describe it, one can broadly distinguish between two main approaches, based on the type and extent of smoothness assumed for the underlying trend function, denoted by  $\mu(\cdot)$ . The first approach, which we will call the *raw isotonic regression* approach, only assumes that  $\mu$  is monotonic, with the result that its estimate  $\hat{\mu}$  is a non-decreasing piecewise constant function. As mentioned, the main difference within the raw approach is the choice of  $\rho$ -function in Eq. (3), which yields either classical isotonic regression, median isotonic regression or the estimators that we propose here in this paper. Those three options also share the property that when  $t_0$  is a point for which  $\mu'(t_0)$  exists and is positive, the asymptotic distribution of  $n^{-1/3}[\hat{\mu}(t_0) - \mu(t_0)]$ , is proportional to the slope at zero of the greatest convex minorant of two-sided Brownian motion with parabolic drift. Better rates could be obtained within the classical approach at the expense of stronger smoothness assumptions (e.g., Leurgans, 1982). In this paper we present estimators that are highly efficient under Gaussian errors as well as highly robust when the error distribution has heavy tails, and we measure the quantitative robustness by the a version of the influence function and by the asymptotic breakdown point.

An important advantage of the raw isotonic approach is that it facilitates graphically the detection of change-points or abrupt changes in the trend  $\mu(\cdot)$ , which is of interest in many applications, such as the Global Warming data presented in Section 7. However, the estimate  $\hat{\mu}(\cdot)$ , being piecewise constant, is less attractive visually than alternative methods that yield smooth curves as estimates. Moreover, the lack of smoothness of the raw isotonic estimate seems unnatural in such situations where abrupt changes appear physically unlikely, such as plant or children growth. Accordingly, there has been interest in the literature in combining smoothness and monotonicity. Friedman and Tibshirani (1984), Mukherjee (1988) and Mammen (1991) were early contributors, and Ramsey (1998) is a more recent method. All these methods, to which we will generically refer as the *smooth isotonic approach* need to assume that  $\mu(\cdot)$  is smooth. In this line of work there are two main choices, namely the selection of a method for smoothing and a technique to achieve a monotone estimate. The usual options for smoothing include local polynomials, splines, and kernel estimators. Each of these methods regulate the smoothness of the resulting  $\hat{\mu}(\cdot)$  in its own way, for example by choosing the kernels and the bandwidth in the kernel

method, or by choosing either the basis of splines, the number of knots, or a penalty term in the splines method. Härdle and Tsybakov (1988) and Boente and Fraiman (1988) proposed smooth curve robust estimation procedures based on M-estimators.

The other choice involved in the smooth isotonic approach is the method for monotonicization, i.e., a transformation of the smooth estimates  $\hat{\mu}_i$  so that they are monotone. Basic choices here are an application of the *min-max* algorithm and the technique proposed by Dette et al. (2006), which basically exploits a type of duality between isotonic regression and density estimation. Another necessary decision within the smooth approach consists on the order of the operations, i.e., the question of whether it is better to smooth first and isotonify later, to isotonify first and smooth later, or to achieve both simultaneously for example by incorporating monotonicity constraints in the smooth estimation method (e.g., Hall and Huang, 2001; He and Shi, 1998). When a two-step method is chosen, there seems to be consensus that it is better to smooth first and monotonicize later, because otherwise there is a danger of losing the monotonic structure (as could happen for example when smoothing is done by kernels estimator). The smooth isotonic approach differs also with the raw isotonic approach in a second way (the first difference being that  $\hat{\mu}$  is smooth). The asymptotic distributions are normal and the convergence rates depend on the smoothness of  $\mu$ , as given for instance by the number of the continuous derivatives of  $\mu(\cdot)$ , and on the smoothing parameters (such as the bandwidth in the kernel method).

From the robustness point of view, a natural way to obtain a robust, smooth and isotonic estimate of  $\mu$  could be achieved by first applying Härdle and Tsybakov (1988) or Boente and Fraiman (1988) methods to obtain a robust and smooth  $\hat{\mu}(\cdot)$ , followed by the isotonicization procedure of Dette et al. (2006). We call this proposal the robust-smooth-isotonic alternative. Our method is better suited to situations in which  $\mu(\cdot)$  could be inflicted by change points, while the robust-smooth-isotonic technique would be a better choice when one desires a smoother estimate  $\hat{\mu}(\cdot)$ .

This paper is structured as follows. In Section 2 we propose the robust isotonic M-estimator. In Section 3 we obtain the limiting distribution of the isotonic M-estimator when the error scale is known. In Section 4 we prove that under general conditions the M-estimators with estimated scale have the same asymptotic distribution than when the scale is known. In Section 5 we define an influence function which measures the sensitivity of the isotonic M-estimator to an infinitesimal amount of pointwise contamination. In Section 6 we calculate the breakdown point of the isotonic M-estimators. In Section 7 we analyze two real dataset using the  $L_2$  and the isotonic M-estimators. In Section 8 we compare by Monte Carlo simulations the finite sample variances of the estimators for two error distributions: normal and Student's with three degrees of freedom. Appendix A contains the proofs.

## 2. Isotonic M-estimators

In similarity with the classical setup, we consider isotonic M-estimators that minimize the objective function (3) within the class  $\mathcal{G}_0$  of piecewise constant non-decreasing functions. As in the  $L_2$  and  $L_1$  cases, the isotonic M-estimator is a step function with knots at (some of) the  $t_j$ 's. In Robertson and Waltman (1968) it is shown that maximum-likelihood-type estimation under isotonic restrictions can be calculated via *min-max* formulae. Assume first that we know that the scale parameter (e.g., the MAD of  $u_j$ ) is  $\sigma_0$ . Since we are considering M-estimators with  $\psi$  non-decreasing (see A1), they can be viewed as the maximum likelihood estimators corresponding to errors with density:

$$g(u) = \frac{\exp\left(-\frac{1}{\sigma_0} \int_0^u \psi(v/\sigma_0) dv\right)}{\int_{-\infty}^{\infty} \left[\exp\left(-\frac{1}{\sigma_0} \int_0^u \psi(v/\sigma_0) dv\right) du\right]}.$$

Then we can compute the isotonic M-estimator at a point  $t$  using the min-max calculation formulae:

$$\hat{\mu}_n(t) = \max_{u \leq t} \min_{v \geq t} \hat{\mu}_n(u, v) = \min_{v \geq t} \max_{u \leq t} \hat{\mu}_n(u, v), \quad (4)$$

where  $\hat{\mu}_n(u, v)$  is the unrestricted M-estimator which minimizes

$$\sum_{j \in C(u, v)} \rho\left(\frac{x_j - \mu}{\sigma_0}\right), \quad (5)$$

and  $C(u, v) = \{j : 1 \leq j \leq n; u \leq t_j \leq v\}$ . Alternatively, if  $\rho$  is convex and differentiable, as we are assuming, the terms  $\hat{\mu}_n(u, v)$  in (4) can be represented uniquely as a zero of

$$S_n(u, v, \mu) = \sum_{j \in C(u, v)} \psi\left(\frac{x_j - \mu}{\sigma_0}\right). \quad (6)$$

In particular, when  $\rho(u) = -\log(g(u)) + \log(g(0))$ , where  $g$  is a probability density, the isotonic M-estimator coincides with the maximum likelihood estimator when  $u$  is assumed to have density  $g$ . In particular if  $g$  is the  $N(0, \sigma_0^2)$  density, the MLE is the M-estimator defined by  $\rho(u) = u^2$  and therefore it coincides with the classical  $L_2$  estimator. When  $g$  is the density of a double exponential distribution, the MLE is the M-estimator defined by  $\rho(u) = |u|$ , and therefore it coincides with the  $L_1$  isotonic estimator. In these two cases the estimators are independent of the value of  $\sigma_0$ . One popular family of  $\psi$  functions

to define M-estimators is the Huber family:

$$\psi_k^H(u) = \text{sign}(u)\min(|u|, k). \tag{7}$$

Clearly, when  $\sigma_0$  is replaced by  $\hat{\sigma}_n$ , Eqs. (4)–(6) still hold with  $\sigma_0$  replaced by  $\hat{\sigma}_n$ . Since  $\psi$  is non-decreasing, the function  $S_n(u, v, \mu)$  defined in Eq. (6) is non-increasing as a function of  $\mu$ . This entails the fundamental identities given below:

$$S_n(u, v, a) > 0 \text{ if and only if } \hat{\mu}_n(u, v) > a, \tag{8}$$

$$S_n(u, v, a) < 0 \text{ if and only if } \hat{\mu}_n(u, v) < a. \tag{9}$$

These identities will be very useful in the development of the asymptotic distribution.

### 3. Asymptotic distribution

In this section we derive the asymptotic distribution of the isotonic M-estimator  $\hat{\mu}_n(t_0)$  of  $\mu(t_0)$ . We first make the sample size  $n$  explicit in the formulation of the model by postulating:

$$x_{n,i} = \mu(t_{n,i}) + u_{n,i}, \tag{10}$$

where the errors  $\{u_{n,i}, 1 \leq i \leq n\}$  form a triangular array of i.i.d. random variables with distribution  $G$  and  $\{t_{n,i}, 1 \leq i \leq n\}$  is a triangular array of observation points. Their exact location is described by the function  $H_n(t) = n^{-1} \sum_{i=1}^n 1(t_{n,i} \leq t)$ . The values  $t_{n,i}$  may be fixed or random but we will assume that there exists a continuous distribution function  $H$  which has as support a finite closed interval such that

$$\sup_t |H_n(t) - H(t)| = o_p(n^{-1/3}). \tag{11}$$

Without loss of generality we shall assume in the sequel it is the interval  $[0, 1]$ .

We will study the asymptotic distribution of  $\hat{\mu}_n(t_0)$  where  $t_0$  is an interior point of  $[0, 1]$ . The classical  $L_2$  isotonic estimator  $\hat{\mu}_n(t_0)$ , with  $t_0$  at the boundary of the support of  $H$ , is known to suffer from the so-called *spiking problem* (e.g., Sun and Woodroffe, 1999), i.e.,  $\hat{\mu}_n(t_0)$  is not even consistent. We further make the following assumptions.

- A2 The function  $H$  is continuously differentiable in a neighborhood of  $t_0$  with  $h(t_0) = H'(t_0) > 0$ .
- A3 For a fixed  $t_0$ , we assume the function  $\mu(t)$  has two continuous derivatives in a neighborhood of  $t_0$ , and  $\mu'(t_0) > 0$ .
- A4 The error distribution  $G$  has a continuous density  $g$  with  $g(0) < 0$ .

We consider first the case where  $\sigma_0$  is known. Our first aim is to show that isotonic M-estimation is asymptotically a local problem. Specifically, we will see in Lemma 1 that  $\hat{\mu}_n(t_0)$  depends only on those  $x_j$  corresponding to observation points  $t_j$  lying in a neighborhood of order  $n^{1/3}$  about  $t_0$ . This result is similar to Prakasa Rao (1969, Lemma 4.1) who stated it in the context of density estimation. Our treatment here will parallel that of Wright (1981), who worked on the asymptotics of the  $L_2$  isotonic regression estimator when the smoothness of the underlying trend function  $\mu(\cdot)$  is specified via the number of its continuous derivatives.

Specifically, since  $H'(t_0) > 0$  we may choose for an arbitrary  $c$  and  $n$  sufficiently large, positive numbers  $\alpha_l(n)$  and  $\alpha_u(n)$  for which

$$H(t_0) - H(t_0 - \alpha_l(n)) = H(t_0 + \alpha_u(n)) - H(t_0) = 2cn^{-1/3}.$$

With this, define the *localized version* of the isotonic M-estimator as

$$\mu_n^*(t_0) = \max_{t_0 - \alpha_l(n) < u \leq t_0} \min_{t_0 \leq v < t_0 + \alpha_u(n)} \hat{\mu}_n(u, v). \tag{12}$$

Then we have the following Lemma

**Lemma 1.** Assume A1–A4 and (11). Then if  $\hat{\mu}_n(t_0)$  is defined by (4), we have,

$$\lim_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} P[\hat{\mu}_n(t_0) \neq \mu_n^*(t_0)] = 0. \tag{13}$$

It is also noteworthy that the estimator in Eq. (12) is not computable, for  $\alpha_l$  and  $\alpha_u$  depend on the distribution  $H$  which is generally unknown. For computational purposes this implies that the calculation of these estimators will indeed be global for fixed sample sized. Lemma 1 is, however, crucial to study the asymptotic properties of  $\hat{\mu}_n(t)$ .

Given an stochastic process  $\{Z(v), -\infty < v < \infty\}$ , we denote by “slogcm $[Z(t)]$ ” the random variable that corresponds to the slope at zero of the greatest convex minorant of  $Z(t)$ . The following theorem gives the asymptotic distribution of  $\hat{\mu}_n(t_0)$ .

**Theorem 1.** Assume A1–A4 and (11), Let  $\hat{\mu}_n(t_0)$  be given by (4), then

$$\left[ \frac{1}{2} \mu'(t_0) H'(t_0) \sigma_0^2 \frac{E_G(\psi^2(u/\sigma))}{[E_G(\psi'(u/\sigma))]^2} \right]^{-1/3} n^{1/3} (\hat{\mu}_n(t_0) - \mu(t_0)) \Rightarrow \text{slogcm}(\mathbb{W}(v) + v^2), \tag{14}$$

where  $\mathbb{W}(v)$  is a two-sided standard Brownian motion.

**Remark 1.** Notice that in the case of the  $L_2$  isotonic estimator the function  $\rho(x) = x^2$ , so  $\psi(x) = 2x$  and  $\psi'(x) = 2$  so that  $E_G(\psi'(u)) = 2$  and  $\sigma_\psi^2 = \sigma_{2u}^2 = 4\sigma^2$ . Then the standardizing constant is given by

$$\frac{1}{2} \mu'(t_0) H'(t_0) \frac{E_G(\psi^2(u))}{(E_G(\psi'(u)))^2} = \frac{1}{2} \mu'(t_0) H'(t_0) \sigma^2,$$

as it is known for the  $L_2$  isotonic estimator.

**Remark 2.** In the case of  $L_1$  isotonic regression notice that in the function  $\rho(x) = |x|$ , so  $\psi(x) = \text{sign}(x)$  for  $x \neq 0$  or else is left undefined. Our method is thus not applicable as the assumptions on  $\psi$  do not hold. However, consider a sequence of functions  $\psi_m(x)$  for which

$$\psi_m(x) = \begin{cases} -1, & x \leq -1/m \\ mx, & -1/m + 1/m^2 < x < 1/m - 1/m^2 \\ 1, & x \geq 1/m \end{cases}, \tag{15}$$

and so that there is a continuity of the first 3 derivatives everywhere; for a construction of such type of functions it is enough to consider quartic splines (e.g., De Boor, 2001). In this setup we get

$$\lim_{m \rightarrow \infty} E_G(\psi_m^2(u)) = 1, \\ \lim_{m \rightarrow \infty} E_G \psi_m'(u) = \lim_{m \rightarrow \infty} m[G(1/m - 1/m^2) - G(-1/m + 1/m^2)] = 2G'(0).$$

Letting  $m \rightarrow \infty$  and  $n \rightarrow \infty$  so that  $m/n \rightarrow \infty$ , we obtain

$$\lim_{n \rightarrow \infty} \frac{1}{2} \mu'(t_0) H'(t_0) \frac{E_G(\psi_n^2(u))}{(E_G(\psi_n'(u)))^2} = \frac{1}{8} \frac{\mu'(t_0)}{[G'(0)]^2} H'(t_0),$$

as it is known in the case of  $L_1$  isotonic regression (see Wang and Huang, 2002).

**Remark 3.** A similar construction to Eq. (15) may be applied to the functions  $\psi_k^H$  in the Huber's family.

#### 4. Robust isotonic M-estimators with a previous scale estimator

We will consider now the more realistic case where  $\sigma_0$  is not known and it is replaced by an estimator  $\hat{\sigma}_n$  previously calculated. Then, in order to obtain an scale equivariant estimator we should replace  $\sigma_0$  in (5) and (6) by a robust scale equivariant estimator  $\hat{\sigma}_n$ . In Remarks 4 and 5 we give some possible choices for  $\hat{\sigma}_n$ .

In the next theorem it is shown that under suitable regularity conditions, it can be proved that if  $\hat{\sigma}_n$  converges to  $\sigma_0$  fast enough, both isotonic M-estimators, the one using the fixed scale  $\sigma_0$  and the one using the scale  $\hat{\sigma}_n$ , have the same asymptotic distribution. Making explicit the scale in the notation, denote the isotonic M-estimator of  $\mu(t)$  based on a fixed scale  $\sigma$  by  $\hat{\mu}_n(t, \sigma)$ . Then

$$\hat{\mu}_n(t, \sigma) = \min_{u \leq t} \max_{v \geq t} \hat{\mu}_n(u, v, \sigma) = \max_{u \leq t} \min_{v \geq t} \hat{\mu}_n(u, v, \sigma),$$

where  $\hat{\mu}_n(u, v, \sigma)$  solves

$$S_n(u, v, \sigma) := \sum_{j \in C(u, v)} \psi \left( \frac{x_j - \hat{\mu}_n(u, v, \sigma)}{\sigma} \right) = 0, \tag{16}$$

over  $C(u, v) := \{j : 1 \leq j \leq n; u \leq t_j \leq v\}$ .

We need the following Additional Assumptions:

- A5 There exists  $k > 0$  such that  $\psi'(u) > 0$  for  $|u| < k$  and  $\psi'(u) = 0$  if  $|u| > k$ .
- A6 The estimator  $\hat{\sigma}_n$  satisfies  $n^{1/3}(\hat{\sigma}_n - \sigma_0) = o_p(1)$ . (.)

Then we have the following theorem:

**Theorem 2.** Assume A1–A6. Then

$$n^{1/3} |\hat{\mu}_n(t, \sigma_0) - \hat{\mu}_n(t, \hat{\sigma}_n)| = o_p(1).$$

Assume also that (11) holds, then both estimators have the same asymptotic distribution.

**Remark 4.** In the context of nonparametric regression, Ghement et al. (2008) propose to use as scale estimator  $\hat{\sigma}_n$  given by

$$\hat{\sigma}_n = \frac{1}{\sqrt{2}}s(x_2 - x_1, \dots, x_n - x_{n-1}),$$

where  $s$  is an M-estimator of scale, i.e.,  $s(u_1, \dots, u_n)$  is defined as the value  $s$  satisfying

$$\frac{1}{n} \sum_{i=1}^n \chi\left(\frac{u_i}{s}\right) = b, \tag{17}$$

where  $\chi(u)$  is a function which is even, non-decreasing for  $u \geq 0$ , bounded and continuous. The right hand side is generally taken so that if  $u$  is  $N(0,1)$ ,  $E\chi(u) = b$ . This condition makes the estimator converging to the standard deviation when applied to a random sample of the  $N(0,1)$  distribution. A popular family of functions  $\chi$  to compute scale M-estimators is the bisquare family given by

$$\chi_c(u) = \begin{cases} 1 - \left(1 - \left(\frac{u}{c}\right)^2\right)^3 & \text{if } |u| \leq c, \\ 1 & \text{if } |u| > c. \end{cases} \tag{18}$$

Ghement et al. (2008) prove that if  $\mu(t)$  is continuous under general conditions on  $\chi$  Assumption A6 is satisfied with  $\sigma_0$  defined by

$$E_G \chi_c\left(\frac{u}{\sigma_0}\right) = b. \tag{19}$$

**Remark 5.** An alternative scale estimator, which does not require the continuity of  $\mu$ , is provided by

$$\hat{\sigma}_n = \frac{1}{\Phi^{-1}(3/4)} \text{median}(|\hat{u}_1|, \dots, |\hat{u}_n|)$$

where  $\hat{u}_1, \dots, \hat{u}_n$  are the residuals corresponding to the  $L_1$  isotonic estimator. We conjecture but we do not have a proof that this estimator also satisfies Assumption A6 with  $\sigma_0 = \text{median}_G(|u|)/\Phi^{-1}(3/4)$ .

**5. Influence function**

In order to obtain the influence function of the isotonic M-estimator at a given point  $t$  we need to assume that the pair  $(x,t)$  is random. In this case the isotonic regression model assumes that  $x = \mu(t) + u$ , where  $u$  is independent of  $t$  and  $\mu(t)$  is non-decreasing. We assume that the error term  $u$  has a symmetric density  $g$ , and that the observation point  $t$  has a distribution with density  $h$ .

We start assuming that  $\sigma_0$  is known and suppose that we want to estimate  $\mu(t_0)$ . Given an arbitrary distribution  $A$  of  $(t,x)$ , the isotonic M-estimating functional of  $\mu(t_0)$  which we henceforth denote by  $T_{t_0}(A)$  is defined in three steps as follows. First for  $r,s \geq 0$  let  $m(t_0,r,s,A)$  be defined as the value  $m$  satisfying

$$\int_{-\infty}^{\infty} \int_{t_0-r}^{t_0+s} \psi\left(\frac{x-m}{\sigma_0}\right) dA = 0.$$

Let

$$m^-(t_0,r,A) = \min_{s \geq 0} m(t_0,r,s,A)$$

and then  $T_{t_0}(A)$  is defined by

$$T_{t_0}(A) = \max_{r \geq 0} m^-(t_0,r,A).$$

Let  $A_n$  be the empirical distribution of  $\{(x_{nj}, t_{nj}), 1 \leq j \leq n\}$ , then if  $\hat{\mu}_n(t)$  is the estimator defined in (4), we have

$$\hat{\mu}_n(t) = T_t(A_n).$$

It is immediate that if  $A_0$  is the joint distribution corresponding to model (1) we have  $T_{t_0}(A_0) = \mu(t_0)$ , so that the isotonic M-estimator is Fisher-consistent. Consider now the contaminated distribution:

$$A_{\varepsilon,t^*,x^*} = (1-\varepsilon)A_0 + \varepsilon\delta_{(t^*,x^*)},$$

where  $\delta_{(t^*,x^*)}$  represents a point mass at  $(t^*,x^*)$ . In this case we define the influence function of  $T_{t_0}$  by

$$IF^*(T_{t_0}, t^*, x^*) = \lim_{\varepsilon \rightarrow 0} \frac{(T_{t_0}(A_{\varepsilon,t^*,x^*}) - T_{t_0}(A_0))^2}{\varepsilon}. \tag{20}$$

Then, we have the following theorem:

**Theorem 3.** Consider the isotonic regression model given in (1) and let  $T_{t_0}$  be an isotonic M-estimating functional, where  $t_0$  is an interior observation point. Then, under Assumptions A1–A4 we have

$$IF^*(T_{t_0}, t^*, x^*) = \begin{cases} \frac{2\mu'(t_0)\sigma_0|\psi((x-\mu(t_0))/\sigma_0)|}{h(t_0)E_G(\psi'(u/\sigma_0))} & \text{if } t^* = t_0, \\ 0 & \text{if } t^* \neq t_0. \end{cases} \tag{21}$$

Notice that in the numerator of (20) appears the square of the bias instead of the plain bias as in the classical definition of Hampel (1974). Therefore for the isotonic M-estimator  $T_{t_0}$  the bias caused by a point mass contamination  $(t_0, x^*)$  is of order  $\varepsilon^{1/2}$  instead of the usual order of  $\varepsilon$ .

Alternatively, it is also of interest to know what happens when we are estimating  $\mu(t_0)$  and contamination takes place at a point  $t^* \neq t_0$ . According to (21), the influence function in this case is zero. This occurs because in this case for  $\varepsilon$  sufficiently small  $T_{t_0}(A_{\varepsilon, t^*, x^*}) = T_{t_0}(A_0)$ .

It is easy to show that when we use a scale  $\hat{\sigma}_n \rightarrow \sigma_0$  defined by a continuous functional, the influence function of the isotonic M-estimator is still given by (21).

### 6. Breakdown point

Roughly speaking the breakdown point of an estimating functional  $T_{t_0}$  of  $\mu(t_0)$  is the smallest fraction of outliers which suffices to drive  $|T_{t_0}|$  to infinity. More precisely, consider the contamination neighborhood  $\mathcal{V}_{A_0, \varepsilon}$  of the distribution  $A_0$  of size  $\varepsilon$  defined as

$$\mathcal{V}_{A_0, \varepsilon} = \{A : A = (1-\varepsilon)A_0 + \varepsilon A^*\},$$

where  $A^*$  is an arbitrary distribution of  $(x, t)$  such that  $t$  takes values in  $[0, 1]$  and  $x$  in  $\mathbb{R}$ . The asymptotic breakdown point of  $T_{t_0}$  at  $A_0$  is defined by

$$\varepsilon^*(T_{t_0}, A_0) = \inf \left\{ \varepsilon : \sup_{A \in \mathcal{V}_{A_0, \varepsilon}} |T_{t_0}(A)| = \infty \right\}.$$

We start considering the case that  $\sigma_0$  is known. Then we have the following theorem.

**Theorem 4.** Consider the isotonic regression model given in (1) and let  $T_{t_0}$  be an isotonic M-estimating functional where  $t_0$  is an interior observation point. Then under Assumptions A1–A4 we have

$$\varepsilon^*(T_{t_0}, A_0) \geq \min \left\{ \frac{H(t_0)}{1+H(t_0)}, \frac{1-H(t_0)}{2-H(t_0)} \right\}.$$

In the special case when  $H$  is uniform, this becomes

$$\varepsilon^*(T_{t_0}, A_0) \geq \min \left\{ \frac{t_0}{1+t_0}, \frac{1-t_0}{2-t_0} \right\}, \tag{22}$$

which takes a maximum value of  $1/3$  at  $t_0 = 1/2$ .

In the case that  $\sigma_0$  is replaced by an estimator  $\hat{\sigma}_n$  derived from a continuous functional  $S$ , it can be proved that the breakdown point of  $T_{t_0}$  satisfies

$$\varepsilon^*(T_{t_0}, A_0) \geq \min \left\{ \frac{H(t_0)}{1+H(t_0)}, \frac{1-H(t_0)}{2-H(t_0)}, \varepsilon^*(A_0) \right\},$$

where  $\varepsilon^*(A_0)$  is the asymptotic breakdown point of the scale estimator.

Ghement et al. (2008) showed that if  $\hat{\sigma}_n$  is defined as in Remark 4, where  $s$  is defined by (17)–(19) with  $c=0.7094$  and  $b=3/4$ , then  $\varepsilon^*(A_0) = 0.5$ . Moreover in this case  $\sigma_0$  coincides with the standard deviation when the error has a normal distribution.

### 7. Examples

**Example 1.** In this section we consider data on Infant Mortality across Countries. The dependent variable, the number of infant deaths per each thousand births is assumed decreasing in the country's per capita income. These data are part of the R package “faraway” and was used in Faraway (2004). The manual of this package only mentions that the data are not recent but it does not give information on the year and source. In Fig. 1 we compare the  $L_2$  isotonic regression estimator with the isotonic M-estimator computed with the Huber's function with  $k=0.98$  and  $\hat{\sigma}_n$  as in Remark 2, where  $s$  is defined by (17)–(19) with  $c=0.7094$  and  $b=3/4$ . There are four countries with mortality above 250: Saudi Arabia (650), Afghanistan (400), Libya (300) and Zambia (259). These countries, specially Saudi Arabia and Libya due to their higher



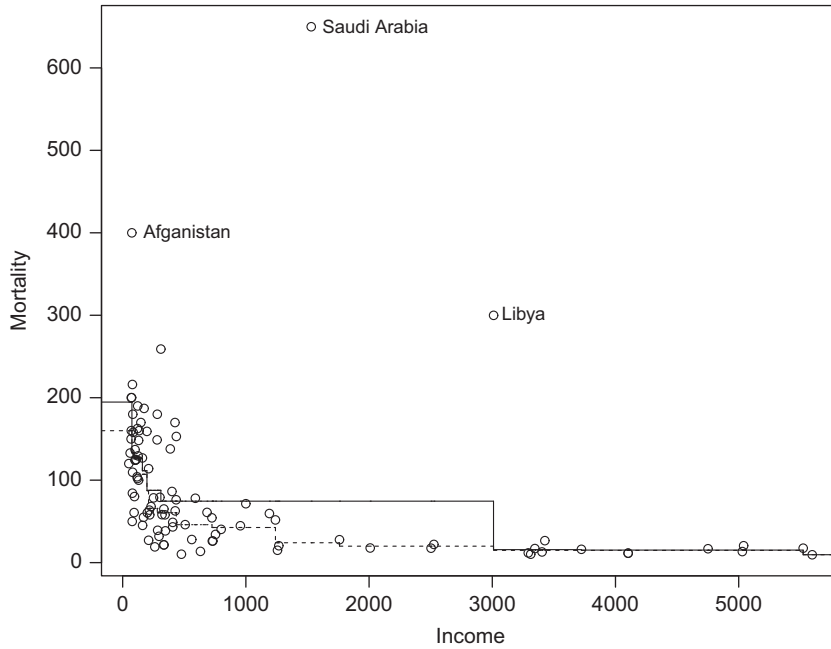


Fig. 1. Infant Mortality Data. The solid line corresponds to the classical isotonic regression and the dashed line to the isotonic M-estimate.

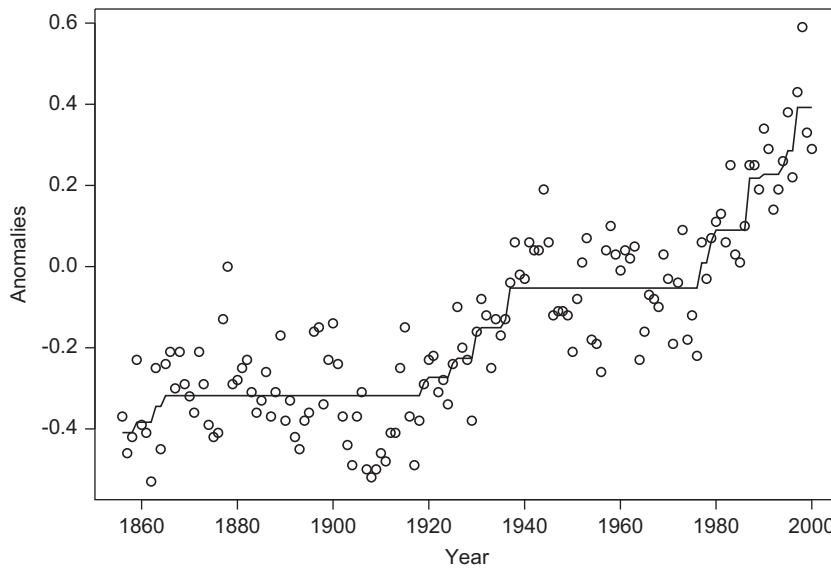


Fig. 2. Word annual weather anomalies 1958–2000. Classical vs. robust estimators.

relative income per capita, exert a large impact on the  $L_2$  estimator. The robust choice, on the other hand, appears to be resistant to these outliers and provides a good fit.

**Example 2.** We reconsider the Global Warming dataset first analyzed in the context of isotonic regression by Wu et al. (2001) from a classical perspective and subsequently analyzed from a Bayesian perspective in Alvarez and Dey (2009). The original data is provided by Jones et al. (see <http://cdiac.esd.ornl.gov/trends/temp/jonescru/jones.html>) containing annual temperature anomalies from 1858 to 2000, expressed in degrees Celsius and are relative to the 1961–1990 mean. Even though the global warming data, being a time series, might be affected by serial correlation, e.g., Fomby and Vogelsang (2002), we opted for simplicity as an illustration, to ignore that aspect of the data and model it as a sequence of i.i.d. observations.

In Fig. 2 we compare the  $L_2$  isotonic estimator with the  $L_1$  alternative. We also computed the isotonic M-estimator used in Example 1. This estimator cannot be distinguished graphically from the  $L_2$  estimator. Inspection of the QQ-plot in the



**Table 1**  
Sample MSE and avar for isotonic regression estimators.

Estimator	n=100		n=500		Avar	
	Normal	Student <sub>3</sub>	Normal	Student <sub>3</sub>	Normal	Student <sub>3</sub>
L <sub>2</sub>	1.93	3.78	1.85	3.65	1.92	3.98
L <sub>1</sub>	2.38	2.89	2.67	2.76	2.59	2.89
M	2.04	2.86	2.11	2.51	2.06	2.53

right hand side of the figure reveals a lightly tailed distribution for the residuals. This seems to indicate that the phenomenon of Global Warming is not due to isolated outlying anomalies, but it is due instead to a steady increasing trend phenomenon. In our view that validates from the point of view of robustness, the conclusions of other authors on the same data (e.g., Wu et al., 2001; Alvarez and Dey, 2009) who have rejected the hypothesis of constancy in series of the worlds annual temperatures in favor of an increasing trend.

**8. Monte carlo results**

Interestingly the limiting distribution of the Isotonic M-estimator is based on the ratio

$$\frac{E_G(\psi'(u/\sigma_0))^2}{E_G(\psi^2(u/\sigma_0))}$$

as in the i.i.d. location problem (e.g., Maronna et al., 2006). The slower convergence rate, however, entails that the respective asymptotic relative efficiencies are those of the location situation taken to the power 2/3. Specifically, note that from Theorem 1 for any isotonic M-estimator:

$$\text{avar}\{n^{1/3}[\hat{\mu}_n(t_0) - \mu(t_0)]\} \tag{23}$$

$$= \left[ \frac{1}{2} \mu'(t_0) H'(t_0) \frac{E_G[\psi(u)^2]}{[E_G\psi'(u)]^2} \right]^{2/3} \text{var}[\text{slogcm}(\mathbb{W}(v) + v^2)], \tag{24}$$

where avar stands for asymptotic variance and var for variance.

In order to determine the finite sample behavior of the isotonic M-estimators we have performed a Monte Carlo study. We took i.i.d. samples from the model (1) with trend term  $\mu(t) = 10 + 5t^2$  and where the distribution G is N(0,1) and Student's t distribution with three degrees of freedom. The values  $\{t_i = i/(n+1), 1 \leq i \leq n\}$  correspond to a uniform limiting distribution  $H(t) = t$  for  $0 < t < 1$ .

We estimated  $\mu(t_0)$  at  $t_0 = 1/2$ , the true value of which is  $\mu(t_0) = 11.25$  using three isotonic estimators: the  $L_2$  isotonic estimator, the  $L_1$  isotonic estimator and the same isotonic M-estimator that was used in the examples. We performed  $N=500$  replicates at two sample sizes,  $n=100$  and 500. Dykstra and Carolan (1998) have established that the variance of the random variable “slogcm( $\mathbb{W}(v) + v^2$ )” is approximately 1.04. Using this value, we present in Table 1 sample mean square errors (MSE) times  $n^{2/3}$  as well as the corresponding asymptotic variances.

We note that for both distributions, the empirical MSEs for  $n=500$  are close to the avar values. We also see that under both distributions the M-estimator is more efficient than the  $L_1$  one, that the M-estimator is more efficient than the  $L_1$  one for both distributions and that the  $L_1$  estimator is slightly less efficient than the  $L_2$  estimator for the normal case but much more efficient for the Student distribution. In summary, the isotonic M-estimate seems to have a good behavior under both distributions.

**Acknowledgments**

We thank two anonymous Referees and the Editor for valuable insights and suggestions that contributed to the improvement of the paper. This research was supported in part by Grants 20020100100276 from Universidad of Buenos Aires, PID 112-200801-00216 from CONICET and PICT 00899 from ANPCYT, Argentina.

**Appendix A**

*A.1. Proof of Lemma 1*

Without loss of generality we can assume that  $\sigma_0 = 1$ . Given  $c > 0$ , for sufficiently large  $n$  there exist positive numbers  $\beta_l(n)$  and  $\beta_u(n)$  for which

$$H(t_0) - H(t_0 - \beta_l(n)) = H(t_0 + \beta_u(n)) - H(t_0) = cn^{-1/3}.$$

As in Wright (1981), we first argue that

$$P[\hat{\mu}_n(t_0) \neq \hat{\mu}_n^*(t_0)] \leq P(\Omega_{1n}) + P(\Omega_{2n}),$$

where

$$\Omega_{1n} = \left\{ \min_{v \geq t_0} \hat{\mu}_n(t_0 - \beta_l(n), v) < \max_{u \leq t_0 - \alpha_l(n)} \hat{\mu}_n[u, t_0 - \beta_l(n)] \right\}, \tag{25}$$

$$\Omega_{2n} = \left\{ \max_{u \leq t_0} \hat{\mu}_n[u, t_0 + \beta_u(n)] > \min_{v \geq t_0 + \alpha_u(n)} \hat{\mu}_n[t_0 + \beta_u(n), v] \right\}. \tag{26}$$

To see this, note that the complement of  $\Omega_{2n}$  is the set in which, for all  $u \leq t_0$  and all  $v \geq t_0 + \beta_u(n)$  we have that  $\{\hat{\mu}_n[u, t_0 + \beta_u(n)] \leq \hat{\mu}_n[t_0 + \beta_u(n), v]\}$ . Since  $\psi$  is non-decreasing we can write

$$\hat{\mu}_n[u, t_0 + \beta_u(n)] \leq \hat{\mu}_n[u, v].$$

This in turn entails that in  $\Omega_{2n}^c$

$$\mu_n^*(t_0) = \max_{u \leq t_0} \min_{t_0 \leq v < t_0 + \alpha_u(n)} \hat{\mu}_n[u, v].$$

Using the fact that the maximum and the minimum may be reversed in computing these estimators (e.g., Robertson and Waltman, 1968) and a similar argument for  $\Omega_{1n}$  in Eq. (25) one can show that

$$P(\Omega_{1n}^c \cap \Omega_{2n}^c) \leq P\{\mu_n^*(t_0) = \hat{\mu}_n(t_0)\}.$$

So we need to prove that

$$\lim_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\Omega_{1n}) = \lim_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\Omega_{2n}) = 0.$$

We will prove  $\lim_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\Omega_{1n}) = 0$ . The result for  $\Omega_{2n}$  can be obtained in a similar manner.

Let

$$A_{1n} = \left\{ \min_{v \geq t_0} \hat{\mu}_n(t_0 - \beta_l(n), v) < \mu(t_0 - \beta_l(n)) \right\}, \tag{27}$$

$$A_{2n} = \left\{ \max_{u \leq t_0 - \alpha_l(n)} \hat{\mu}_n[u, t_0 - \beta_l(n)] > \mu(t_0 - \beta_l(n)) \right\}. \tag{28}$$

Since

$$P(\Omega_{1n}) \leq P(A_{1n}) + P(A_{2n}),$$

it will be enough to prove that

$$\lim_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} P(A_{in}) = 0, i = 1, 2. \tag{29}$$

Since the proofs of (29) for  $i = 1$  and  $2$  are similar, (29) will be only proved for  $i = 1$ . By the fundamental identity (9) we have

$$A_{1n} = \left\{ \min_{v \geq t_0} S_n(t_0 - \beta_l(n), v, \mu(t_0 - \beta_l(n))) < 0 \right\}. \tag{30}$$

In the sequel in order to simplify notation we will omit the subindex  $n$  writing  $x_{nj} = x_j$ ,  $t_{nj} = t_j$  and  $u_{nj} = u_j$  making it explicit only when there is a risk of confusion. We can write

$$\begin{aligned} S_n(t_0 - \beta_l(n), v, \mu(t_0 - \beta_l(n))) &= \sum_{j \in C(t_0 - \beta_l(n), v)} \psi(x_j - \mu(t_0 - \beta_l(n))) = \sum_{j \in C(t_0 - \beta_l(n), v)} \psi(x_j - \mu(t_j) + \mu(t_j) - \mu(t_0 - \beta_l(n))) \\ &= \sum_{j \in C(t_0 - \beta_l(n), v)} \psi(u_j + (\mu(t_j) - \mu(t_0 - \beta_l(n))), \end{aligned}$$

and by a Taylor expansion we get

$$S_n(t_0 - \beta_l(n), v, \mu(t_0 - \beta_l(n))) = \sum_{j \in C(t_0 - \beta_l(n), v)} [\psi(u_j) + \psi'(u_j + a_j^*)(\mu(t_j) - \mu(t_0 - \beta_l(n))),$$

where  $0 \leq a_j^* \leq \mu(t_j) - \mu(t_0 - \beta_l(n))$ . Put  $\tau = \sup \psi'$ , then

$$S_n(t_0 - \beta_l(n), v, \mu(t_0 - \beta_l(n))) \leq \sum_{j \in C(t_0 - \beta_l(n), v)} \psi(u_j) + \tau \sum_{j \in C(t_0 - \beta_l(n), v)} (\mu(t_j) - \mu(t_0 - \beta_l(n))).$$

Thus, since  $\mu(t)$  is increasing we get

$$\min_{v \geq t_0} S_n(t_0 - \beta_1(n), v, \mu(t_0 - \beta_1(n))) \leq \min_{v \geq t_0} \sum_{j \in C(t_0 - \beta_1(n), v)} \psi(u_j) + \tau \sum_{j \in C(t_0 - \beta_1(n), t_0)} (\mu(t_j) - \mu(t_0 - \beta_1(n))). \tag{31}$$

Put  $n_l(v) := \#\{j : t_0 - \beta_1(n) \leq t_{nj} \leq v\}$ . As  $n_l(v) \geq n_l(t_0)$ , we obtain

$$\min_{v \geq t_0} S_n(t_0 - \beta_1(n), v, \mu(t_0 - \beta_1(n))) \leq \min_{v \geq t_0} \frac{1}{n_l(v)} \sum_{j \in C(t_0 - \beta_1(n), v)} \psi(u_j) + \frac{\tau}{n_l(t_0)} \sum_{j \in C(t_0 - \beta_1(n), t_0)} (\mu(t_j) - \mu(t_0 - \beta_1(n))). \tag{32}$$

Therefore the event  $A_{1n}$  defined in (30) is included in the event  $\Delta_n$  defined by

$$\Delta_n = \left\{ \max_{v \geq t_0} \frac{1}{n_l(v)} \sum_{j \in C(t_0 - \beta_1(n), v)} -\psi(u_j) > \frac{\tau}{n_l(t_0)} \sum_{j \in C(t_0 - \beta_1(n), t_0)} (\mu(t_j) - \mu(t_0 - \beta_1(n))) \right\}.$$

The equation above can be rewritten in terms of integrals with respect to the empirical distribution of the  $t$ 's as

$$\max_{v \geq t_0} \frac{1}{n_l(v)} \sum_{j \in C(t_0 - \beta_1(n), v)} -\psi(u_j) > \tau \int_{t_0 - \beta_1(n)}^{t_0} |\mu(s) - \mu(t_0 - \beta_1(n))| dH_n(s). \tag{33}$$

Since  $u_1, \dots, u_n$  are i.i.d., relabelling the  $u_j$ 's on the left hand side we get that

$$P(A_{1n}) \leq P(\Delta_n^*), \tag{34}$$

where

$$\Delta_n^* = \left\{ \max_{n_l(t_0) \leq k \leq n} \frac{1}{k} \sum_{n_l(t_0) \leq j \leq k} -\psi(u_j) > 2\tau \int_{t_0 - \beta_1(n)}^{t_0} |\mu(s) - \mu(t_0 - \beta_1(n))| dH_n(s) \right\}. \tag{35}$$

Adding and subtracting  $dH(s)$  we can write

$$\int_{t_0 - \beta_1(n)}^{t_0} [\mu(s) - \mu(t_0 - \beta_1(n))] dH_n(s) = \int_{t_0 - \beta_1(n)}^{t_0} [\mu(s) - \mu(t_0 - \beta_1(n))] dH(s) + \int_{t_0 - \beta_1(n)}^{t_0} [\mu(s) - \mu(t_0 - \beta_1(n))] d(H_n(s) - H(s)). \tag{36}$$

Using (11), for  $n$  large enough, the second term in the above equation is bounded by

$$\left| \int_{t_0 - \beta_1(n)}^{t_0} [\mu(s) - \mu(t_0 - \beta_1(n))] d(H_n(s) - H(s)) \right| \leq 2(\mu(t_0) - \mu(t_0 - \beta_1(n))) \sup_t |H_n(t) - H(t)| \leq 2\mu'(t_0)\beta_1(n)n^{-1/3}o(1),$$

and since by the inverse function theorem  $\beta_1(n) = c[H'(t_0)]^{-1}n^{-1/3}[1 + o(1)]$ , we obtain that for some constant  $A$  which does not depend on  $c$  we can write

$$\left| \int_{t_0 - \beta_1(n)}^{t_0} [\mu(s) - \mu(t_0 - \beta_1(n))] d(H_n(s) - H(s)) \right| \leq Acn^{-2/3}o(1). \tag{37}$$

Consider now the first term in the right hand side of Eq. (36). Using (11) we have

$$\begin{aligned} \int_{(t_0 - \beta_1(n), t_0)} [\mu(s) - \mu(t_0 - \beta_1(n))] dH(s) &= \int_{t_0 - \beta_1(n)}^{t_0} [\mu(t_0) - \mu(t_0 - \beta_1(n))] dH(s) - \int_{t_0 - \beta_1(n)}^{t_0} [\mu(t_0) - \mu(s)] dH(s) \\ &= [\mu(t_0) - \mu(t_0 - \beta_1(n))] [H(t_0) - H(t_0 - \beta_1(n))] - \int_{t_0 - \beta_1(n)}^{t_0} [\mu(t_0) - \mu(s)] dH(s) \\ &\leq \left( \frac{\mu(t_0) - \mu(t_0 - \beta_1(n))}{\beta_1(n)} \right) \left( \frac{H(t_0) - H(t_0 - \beta_1(n))}{\beta_1(n)} \right) \beta_1(n)^2 \\ &= \mu'(t_0)[1 + o(1)]H'(t_0)[1 + o(1)]\beta_1(n)^2 = \mu'(t_0)[H'(t_0)]^{-1}c^2n^{-2/3}[1 + o(1)]. \end{aligned}$$

Therefore

$$\begin{aligned} \int_{t_0 - \beta_1(n)}^{t_0} [\mu(s) - \mu(t_0 - \beta_1(n))] dH_n(s) &\leq \mu'(t_0)c^2[H'(t_0)]^{-1}n^{-2/3}(1 + o(1)) + 2\mu'(t_0)c[H'(t_0)]^{-1}n^{-2/3}o(1) \\ &\leq \mu'(t_0)c^* [H'(t_0)]^{-1}n^{-1/3}\{n^{-1/3}[1 + o(1)] + 2n^{-1/3}o(1)\} \end{aligned}$$

with  $c^* = \max(c, c^2)$ . Then, for some constant  $B$  which does not depend on  $c$  we can write

$$\int_{t_0 - \beta_1(n)}^{t_0} [\mu(s) - \mu(t_0 - \beta_1(n))] dH_n(s) \leq Bc^* n^{-1/3}. \tag{38}$$

Using (30), (33)–(38) we derive that there exists a constant  $D$  independent of  $c$  such that for  $n$  large enough and  $c > 1$

$$P(A_{1n}) \leq P \left\{ \max_{n_l(t_0) \leq k \leq n} \frac{1}{k} \sum_{1 \leq j \leq k} -\psi(u_j) > Dc^2n^{-1/3} \right\}. \tag{39}$$

At this point, we use the Hájek-Rényi Maximal Inequality (e.g., [Shorack, 2000](#)) which asserts that for a sequence  $y_1, \dots, y_n$  of independent random variables with mean 0 and finite variances and for a positive non-decreasing real sequence  $\{b_k, k \in N\}$ :

$$P\left\{\max_{m \leq k \leq n} \left| \frac{\sum_{j=1}^k Y_j}{b_k} \right| \geq \lambda\right\} \leq \frac{1}{\lambda^2} \left\{ \sum_{k=1}^m \frac{E(Y_k^2)}{b_m^2} + \sum_{k=m+1}^n \frac{E(Y_k^2)}{b_k^2} \right\}. \tag{40}$$

Using this inequality from (39) we get that

$$P(A_{1n}) \leq P\left\{\max_{n_l(t_0) \leq k \leq n_1} \sum_{1 \leq j \leq k} \frac{-\psi(u_j)}{k} > Dc^2 n^{-1/3}\right\} \leq \frac{E_G(\psi^2(u)) \left( \frac{1}{n_l(t_0)} + \sum_{k=n_l(t_0)}^n \frac{1}{k^2} \right)}{D^2 c^4 n^{-2/3}}. \tag{41}$$

Approximating the Riemann sum we obtain

$$\sum_{k=n_l(t_0)}^n k^{-2} \leq \frac{1}{n_l(t_0)} \tag{42}$$

and since by (11)  $n_l(t_0) = cn^{2/3}(1 + o(1))$ , for  $n$  large enough we have

$$n_l(t_0)^{-1} \leq 2c^{-1} n^{-2/3}. \tag{43}$$

Using (41)–(43) we derive that for  $n$  large enough

$$P(A_{1n}) \leq \frac{2E_G(\psi^2(u))}{D^2 c^4 n^{-2/3} n_l(t_0)} \leq \frac{4E_G(\psi^2(u))}{D^2 c^5}.$$

Then the Lemma follows immediately.

### A.2. Proof of Theorem 1

Without loss of generality we can assume that  $\sigma_0 = 1$ . Since  $\alpha_l(n) = \alpha_u(n) = 2c[H'(t_0)]^{-1}n^{-1/3}[1 + o(1)]$ , and  $c$  is arbitrary, we will consider the localized estimator:

$$\hat{\mu}_n^c(t_0) = \max_{t_0 - cn^{-1/3} < u \leq t_0} \min_{t_0 \leq v < t_0 + cn^{-1/3}} \hat{\mu}_n(u, v) = \max_{u \leq t_0} \min_{v \geq t_0} \hat{\mu}_n^c(u, v), \tag{44}$$

where  $\hat{\mu}_n^c(u, v)$  is defined as the root of

$$S_n^c(u, v, \mu) = \sum_{j \in D(u, v)} \psi(x_j - \mu) \tag{45}$$

over  $D(u, v) = \{j : 1 \leq j \leq n; t_0 - cn^{-1/3} < u \leq t_j \leq v < t_0 + cn^{-1/3}\}$ . Note that the localized estimator depends on the  $t_j$ 's that lie on a neighborhood about  $t_0$  which shrinks at a rate  $n^{-1/3}$ . To proceed with the development of the asymptotic distribution let now  $w_j = n^{1/3}(t_j - t_0)$ ,  $r = n^{1/3}(u - t_0)$  and  $s = n^{1/3}(v - t_0)$ . With this notation,  $\hat{\mu}_n^c(u, v)$  is a root of the partial sums in the parametrization:

$$\check{S}_n^c(r, s, \mu) = \sum_{j \in B(r, s)} \psi(x_j - \mu) = 0, \tag{46}$$

where  $B(r, s) = \{j : 1 \leq j \leq n; r \leq w_j \leq s; r, s \in [-c, c]\}$ . So that the relabelling implies  $\hat{\mu}_n^c(u, v) \equiv \hat{\mu}_n^c(r, s)$ . Consequently,

$$\hat{\mu}_n^c(t_0) = \max_{r \leq 0} \min_{v \geq 0} \hat{\mu}_n^c(r, s).$$

Now a Taylor expansion of  $\mu(t_j)$  around  $t_0$  for any  $j \in B(r, s)$  gives

$$\mu(t_j) = \mu(t_0) + \mu'(t_0)(t_j - t_0) + o(|t_j - t_0|) = \mu(t_0) + \mu'(t_0)n^{-1/3}w_j + o_j(n^{-1/3})$$

which entails that

$$x_j = \mu(t_0) + \mu'(t_0)n^{-1/3}w_j + u_j + o_j(n^{-1/3}),$$

where

$$|o_j(n^{-1/3})| \leq K_1 c^2 n^{-2/3},$$

and  $K_1 = \max \psi''$ .

Using the equivariance of M-estimators, the monotonicity of  $\psi$  and the fact that  $\psi''$  is bounded, it can be proved that

$$\hat{\mu}_n^c(r, s) = \mu(t_0) + \check{\mu}_n^c(r, s) + o_{rs}(n^{-1/3}),$$

where  $\tilde{\mu}_n^c(r,s)$  solves

$$\tilde{S}_n^c(r,s,\mu) = \sum_{j \in B(r,s)} \psi(n^{-1/3} \mu'(t_0) w_j + u_j - \mu) = 0 \tag{47}$$

and

$$|O_{rs}(n^{-1/3})| \leq K_1 c^2 n^{-2/3}.$$

Thus, using that the  $w_j$  are bounded over  $-c \leq r \leq w_i \leq s \leq c$  we have

$$\hat{\mu}_n^c(t_0) = \mu(t_0) + \max_{-c \leq r \leq 0} \min_{0 \leq v \leq c} \tilde{\mu}_n^c(r,s) + o(n^{-1/3}).$$

This entails that

$$n^{1/3} [\hat{\mu}_n^c(t_0) - \mu(t_0)] = \max_{r \leq 0} \min_{v \geq 0} n^{1/3} \tilde{\mu}_n^c(r,s) + O_{rs}^*(1),$$

where

$$|O_{rs}^*(n^{-1/3})| \leq K_2 c^2.$$

Then, we only need to obtain the asymptotic distribution of

$$\Delta_{n,c} = n^{1/3} \max_{r \leq 0} \min_{s \geq 0} \tilde{\mu}_n^c(r,s).$$

Let  $\tilde{\mu}_n^{c*}(r,s)$  be the solution of

$$\sum_{j \in B(r,s)} \psi(u_j - \mu) = 0.$$

Since  $|n^{-1/3} \mu'(t_0) w_j| \leq n^{-1/3} \mu'(t_0) c$  we have that

$$\tilde{\mu}_n^c(r,s) = \tilde{\mu}_n^{c*}(r,s) + d_{nrs}, \tag{48}$$

where

$$|d_{nrs}| \leq K_3 c n^{-1/3}. \tag{49}$$

We will approximate now  $n_{rs}$  as follows:

$$\begin{aligned} \frac{n_{rs}}{n} &= \frac{1}{n} \#\{1 \leq j \leq n : r \leq 0 \leq w_j \leq s\} = \frac{1}{n} \#\{1 \leq j \leq n : t_0 - r n^{-1/3} \leq t_0 \leq t_j \leq t_0 + s n^{-1/3}\} \\ &= H_n(t_0 + s n^{-1/3}) - H_n(t_0 - r n^{-1/3}) = [H_n(t_0 + s n^{-1/3}) - H(t_0 + s n^{-1/3})] \\ &\quad + [H(t_0 + s n^{-1/3}) - H(t_0 - r n^{-1/3})] - [H_n(t_0) - H(t_0)] = H'(t_0)(s-r)n^{-1/3} + o(n^{-1/3}) = n^{-1/3} H'(t_0)(s-r)[1 + o(1)], \end{aligned} \tag{50}$$

and therefore

$$n_{rs} = n^{2/3} H'(t_0)(s-r)[1 + o(1)], \tag{51}$$

$$n_{rs}^{1/2} = n^{1/3} H'(t_0)^{1/2} (s-r)^{1/2} (1 + o(1)), \tag{52}$$

and

$$\frac{n^{1/3}}{n_{rs}} = \frac{1}{n_{rs}^{1/2} H'(t_0)^{1/2} (s-r)^{1/2} (1 + o(1))}. \tag{53}$$

Then, taking  $n_{rs} \rightarrow \infty$  and applying the law of large numbers is easy to show that  $\tilde{\mu}_n^{c*}(r,s) \rightarrow \mu_0$  a.s. and therefore by (48) and (49)  $\tilde{\mu}_n^{c*}(r,s) \rightarrow_p \mu_0$  too. Since  $\tilde{\mu}_n^c(r,s)$  satisfies (47), by a Taylor expansion of  $\tilde{S}_n^c(r,s)$  we get

$$\sum_{B(r,s)} \psi(u_j) - \sum_{B(r,s)} \psi'(u_j) (\tilde{\mu}_n^c(r,s) - n^{-1/3} \mu'(t_0) w_j) + \sum_{B(r,s)} \psi''(\varepsilon_j^*) (\tilde{\mu}_n^c(r,s) - n^{-1/3} \mu'(t_0) w_j)^2 = 0.$$

Then, we obtain

$$\tilde{\mu}_n^c(r,s) = \frac{\sum_{j \in B(r,s)} \psi(u_j) + \mu'(t_0) n^{-1/3} \sum_{j \in B(r,s)} w_j \psi'(u_j) + n^{-2/3} \mu'(t_0)^2 \sum_{j \in B(r,s)} w_j^2 \psi''(\varepsilon_j^*)}{\sum_{j \in B(r,s)} \psi'(u_j) - \tilde{\mu}_n^c(r,s) \sum_{j \in B(r,s)} \psi''(\varepsilon_j^*)^2 - 2n^{-1/3} \mu'(t_0) \sum_{j \in B(r,s)} w_j \psi''(\varepsilon_j^*)}$$

and then

$$n^{1/3} \tilde{\mu}_n^c(r,s) = \frac{\frac{n^{1/3}}{n^{2/3}} \sum_{j \in B(r,s)} \psi(u_j) + \mu'(t_0) \frac{1}{n^{2/3}} \sum_{j \in B(r,s)} w_j \psi'(u_j) + n^{-1/3} \mu'(t_0)^2 \frac{1}{n^{2/3}} \sum_{j \in B(r,s)} w_j^2 \psi''(\varepsilon_j^*)}{\frac{1}{n^{2/3}} \sum_{j \in B(r,s)} \psi'(u_j) - \tilde{\mu}_n^c(r,s) \frac{1}{n^{2/3}} \sum_{j \in B(r,s)} \psi''(\varepsilon_j^*)^2 - 2n^{-1/3} \mu'(t_0) \frac{1}{n^{2/3}} \sum_{j \in B(r,s)} w_j \psi''(\varepsilon_j^*)}. \tag{54}$$

By (47), the Law of the Large Numbers,  $|w_j| \leq c$  and  $\psi''$  bounded we have

$$n^{-1/3} \mu'(t_0)^2 \frac{1}{n^{2/3}} \sum_{j \in B(r,s)} w_j^2 \psi''(\varepsilon_j^*) \rightarrow 0,$$

$$2n^{-1/3} \mu'(t_0) \frac{1}{n^{2/3}} \sum_{j \in B(r,s)} w_j \psi''(\varepsilon_j^*) \rightarrow 0,$$

$$\tilde{\mu}_n^c(r,s) \frac{1}{n^{2/3}} \sum_{j \in B(r,s)} \psi''(\varepsilon_j^*)^2 \rightarrow 0$$

and

$$\frac{1}{n^{2/3}} \sum_{j \in B(r,s)} \psi'(u_j) \rightarrow (s-r)H'(t_0)E_G(\psi'(u)) \text{ a.s.}$$

Then, (54) entails

$$(s-r)E_G(\psi'(u))H'(t_0)n^{1/3} \tilde{\mu}_n^c(r,s) = \frac{1}{n^{1/3}} \sum_{j \in B(r,s)} \psi(u_j) + \mu'(t_0) \frac{1}{n^{2/3}} \sum_{j \in B(r,s)} w_j \psi'(u_j) + o_{rs}(1). \tag{55}$$

Let

$$B_n(s) = \begin{cases} \frac{\mu'(t_0)}{n^{1/3} E_G(\psi^2(u))^{1/2} H'(t_0)^{1/2}} \sum_{j \in B(0,s)} \psi(u_j) & \text{if } s > 0 \\ \frac{\mu'(t_0)}{n^{1/3} E_G(\psi^2(u))^{1/2} H'(t_0)^{1/2}} \sum_{j \in B(s,0)} -\psi(u_j) & \text{if } s < 0, \end{cases} \tag{56}$$

By (52) and the Central Limit Theorem we have that for any set of finite numbers  $s_1, s_2, \dots, s_r, -c \leq s_i \leq c$ , the random vector  $(B_{n,s_1}, \dots, B_{n,s_r})$  converges in distribution to  $N(0, \Sigma)$  where  $\Sigma = (\sigma_{ij})$  with

$$\sigma_{ij} = \begin{cases} s_i \wedge s_j & \text{if } s_i \geq 0, s_j \geq 0, \\ -s_i \wedge -s_j & \text{if } s_i \leq 0, s_j \leq 0, \\ 0 & \text{if } s_i \geq 0, s_j \leq 0. \end{cases}$$

Moreover, using standard arguments, it can be proved that  $B_n(s)$  is tight. Then, we have

$$B_n(s) \xrightarrow{D} B(s), \tag{57}$$

where  $B$  is a two sided Brownian motion

As for the second term in the right hand side of (55) define

$$A_n(s) = \begin{cases} \frac{1}{n^{2/3}} \sum_{0 \leq w_j \leq s} \psi'(u_j) w_j & \text{if } s > 0, \\ \frac{1}{n^{2/3}} \sum_{s \leq w_j \leq 0} -\psi'(u_j) w_j & \text{if } s < 0. \end{cases} \tag{58}$$

For  $s > 0$  we can write

$$A_n(s) = \frac{1}{n} \sum_{j=1}^n n^{2/3} \psi'(u_j) (t_j - t_0) 1(t_0 \leq t_j \leq t_0 + sn^{-1/3}) \tag{59}$$

and then

$$E(A_n(s)) = E_G(\psi'(u_j)) \int_{t_0}^{t_0 + sn^{-1/3}} n^{2/3} (t - t_0) dH_n. \tag{60}$$

Integrating by parts we get

$$\int_{t_0}^{t_0 + sn^{-1/3}} n^{2/3} (t - t_0) dH_n = n^{1/3} s^2 H_n(t_0 + sn^{-1/3}) - \int_{t_0}^{t_0 + sn^{-1/3}} n^{2/3} H_n(t) dt \tag{61}$$

and by (11) we have

$$n^{1/3} s^2 H_n(t_0 + sn^{-1/3}) = n^{1/3} s^2 H(t_0 + sn^{-1/3}) + o(1). \tag{62}$$

We can write

$$\int_{t_0}^{t_0 + sn^{-1/3}} n^{2/3} H_n(t) dt = \int_{t_0}^{t_0 + sn^{-1/3}} n^{2/3} H(t) dt + \int_{t_0}^{t_0 + sn^{-1/3}} n^{2/3} (H_n(t) - H(t)) dt,$$

and by (11) we get

$$\left| \int_{t_0}^{t_0+sn^{-1/3}} n^{2/3}(H_n(t)-H(t)) dt \right| \leq sn^{-1/3} n^{1/3} \sup_t n^{1/3} |H_n(t)-H(t)| = o(1). \tag{63}$$

Therefore by (61)–(63) we get

$$\begin{aligned} \int_{t_0}^{t_0+sn^{-1/3}} n^{2/3}(t-t_0)dH_n &= n^{1/3}s^2H_n(t_0+sn^{-1/3}) - \int_{t_0}^{t_0+sn^{-1/3}} n^{2/3}H_n(t) dt + o(1) = \int_{t_0}^{t_0+sn^{-1/3}} n^{2/3}(t-t_0)dH + o(1) \\ &= \int_{t_0}^{t_0+sn^{-1/3}} n^{2/3}(t-t_0)H'(t) dt + o(1) = H'(t_0) \int_{t_0}^{t_0+sn^{-1/3}} n^{2/3}(t-t_0) dt + o(1) = H'(t_0) \frac{s^2}{2} + o(1) \end{aligned}$$

and for (60) we get that for  $s > 0$

$$E(A_n(s)) = E_G(\psi'(u_j))H'(t_0) \frac{s^2}{2} + o(1). \tag{64}$$

Now we compute the variance of  $A_n(s)$ . Using (59) we have

$$A_n(s) = \frac{1}{n} \sum_{j=1}^n n^{2/3} \psi'(u_j)(t_j-t_0) 1(t_0 \leq t_j \leq t_0+sn^{-1/3})$$

$$\begin{aligned} \text{var}(A_n(s)) &= \frac{\text{var}(\psi'(u))}{n} \sum_{j=1}^n n^{1/3}(t_j-t_0) 1(t_0 \leq t_j \leq t_0+sn^{-1/3}) = n^{1/3} \text{var}(\psi'(u)) \int_{t_0}^{t_0+sn^{-1/3}} (t_j-t_0)^2 dH_n \\ &\leq \text{var}(\psi'(u)) n^{1/3} s^2 n^{-2/3} sn^{-1/3} = \text{var}(\psi'(u)) s^3 n^{-2/3} = o(1). \end{aligned}$$

Then by (64) we obtain

$$A_n(s) \xrightarrow{P} E_G(\psi'(u_j))\mu'(t_0)H'(t_0) \frac{s^2}{2} \quad \text{for } s > 0. \tag{65}$$

Similarly we can prove that

$$A_n(s) \xrightarrow{P} E_G(\psi'(u_j))\mu'(t_0)H'(t_0) \frac{s^2}{2} \quad \text{for } s < 0. \tag{66}$$

Therefore from (55)–(58), (65) and (66) we get that

$$(s-r) \frac{E_G(\psi'(u))}{E_G(\psi^2(u))^{1/2}} H'(t_0)^{1/2} n^{1/3} \tilde{\mu}_n^c(r,s) \xrightarrow{D} (B(s)-B(r)) + \frac{E_G(\psi'(u_j))}{E_G(\psi^2(u))^{1/2}} \mu'(t_0)H'(t_0)^{1/2} \frac{s^2-r^2}{2}.$$

Now the rest of the proof is as in Wright (1981).

### A.3. Proof of Theorem 2

We require the following Lemma

**Lemma 2.** Assume A1–A5. Then,

$$\left| \frac{\partial}{\partial \sigma} \hat{\mu}_n(u, v, \sigma) \right| \leq k, \quad \text{for all } u \leq v. \tag{67}$$

**Proof.** Taking the first derivative of Eq. (16) with respect to  $\sigma$  yields

$$\sum_{j \in C(u, v)} \psi' \left( \frac{x_j - \hat{\mu}_n(u, v, \sigma)}{\sigma} \right) \left\{ -\frac{1}{\sigma^2} (x_j - \hat{\mu}_n(u, v, \sigma)) - \frac{1}{\sigma} \frac{\partial \hat{\mu}_n(u, v, \sigma)}{\partial \sigma} \right\} = 0$$

and then

$$\frac{\partial}{\partial \sigma} \hat{\mu}_n(u, v, \sigma) = - \frac{\sum_{j \in C(u, v)} \psi' \left( \frac{x_j - \hat{\mu}_n(u, v, \sigma)}{\sigma} \right) \frac{x_j - \hat{\mu}_n(u, v, \sigma)}{\sigma}}{\sum_{j \in C(u, v)} \psi' \left( \frac{x_j - \hat{\mu}_n(u, v, \sigma)}{\sigma} \right)}.$$

Let  $D(u, v) = C(u, v) \cap \{j : |x_j - \hat{\mu}_n(u, v, \sigma)| / \sigma \leq k\}$ . Then by A5 we obtain

$$\left| \frac{\partial}{\partial \sigma} \hat{\mu}_n(u, v, \sigma) \right| \leq \frac{\sum_D \psi' \left( \frac{x_j - \hat{\mu}_n(u, v, \sigma)}{\sigma} \right) \left| \frac{x_j - \hat{\mu}_n(u, v, \sigma)}{\sigma} \right|}{\sum_D \psi' \left( \frac{x_j - \hat{\mu}_n(u, v, \sigma)}{\sigma} \right)} \leq k.$$



Therefore

$$\left| \frac{\partial}{\partial \sigma} \hat{\mu}_n(u, \nu, \sigma) \right| \leq k, \quad \square$$

**Proof of Theorem 2.** By the mean value theorem

$$\hat{\mu}_n(u, \nu, \hat{\sigma}_n) = \hat{\mu}_n(u, \nu, \sigma_0) + \frac{\partial}{\partial \sigma} \hat{\mu}_n(u, \nu, \sigma_n^*)(\hat{\sigma}_n - \sigma_0),$$

where  $\sigma_n^*$  is some intermediate point between  $\sigma$  and  $\hat{\sigma}_n$ . Hence, by Lemma 2 we have

$$\max_{u \leq t} \min_{\nu \geq t} \hat{\mu}_n(u, \nu, \hat{\sigma}_n) - k |\hat{\sigma}_n - \sigma_0| \leq \max_{u \leq t} \min_{\nu \geq t} \hat{\mu}_n(u, \nu, \sigma) \leq \max_{u \leq t} \min_{\nu \geq t} \hat{\mu}_n(u, \nu, \hat{\sigma}_n) + k |\hat{\sigma}_n - \sigma_0|$$

and A6 implies

$$n^{1/3} |\hat{\mu}_n(t, \hat{\sigma}_n) - \hat{\mu}_n(t, \sigma_0)| \leq kn^{1/3} |\hat{\sigma}_n - \sigma_0| = o_p(1). \quad \square$$

#### A.4. Proof of Theorem 3

Without loss of generality we can assume that  $\sigma_0 = 1$ . We consider first the case  $t^* = t_0$ . Assume that  $x^* < \mu(t_0)$ . Then  $\delta_{(t_0, x^*)}$  represents a contamination model where an outlier is placed at the observation point  $t_0$  with value  $x^*$  which is below the trend  $\mu(t_0)$  at the point. Because of that,

$$m^-(t_0, r, A_{\varepsilon, t_0, x^*}) = m(t_0, r, t_0, A_0).$$

Let  $k = k(\varepsilon)$  be the value such that  $T_{t_0}(A_{\varepsilon, t_0, x^*}) = m^-(t_0, t_0 - k, A_0) = m(t_0, t_0 - k, t_0, A_0)$ . It is immediate that

$$\mu(t_0 - k) = T_{t_0}(A_{\varepsilon}) = m(t_0, t_0 - k, t_0, A_0). \tag{68}$$

Then  $m(t_0 - k, t_0, A_0)$  should be the value of  $m$  satisfying

$$\varepsilon \psi(x_0 - m) + (1 - \varepsilon) \int_{t_0 - k(\varepsilon)}^{t_0} \int_{-\infty}^{\infty} \psi(\mu(t) + u - m) h(t) g(u) dt du = 0, \tag{69}$$

and, since by (68)  $m = \mu(t_0 - k)$ , we have

$$\varepsilon \psi(x_0 - \mu(t_0 - k(\varepsilon))) + (1 - \varepsilon) \int_{t_0 - k(\varepsilon)}^{t_0} \int_{-\infty}^{\infty} \psi(\mu(t) + u - \mu(t_0 - k(\varepsilon))) h(t) g(u) dt du = 0. \tag{70}$$

Applying the Mean Value Theorem to the first term of (70) we can find  $0 \leq \varepsilon^* < \varepsilon$  such that

$$\psi(x^* - \mu(t_0 - k(\varepsilon))) = \psi(x^* - \mu(t_0)) - \psi'(x^* - \mu(t_0) - k(\varepsilon^*)) \mu'(t_0 - k(\varepsilon^*)) k(\varepsilon). \tag{71}$$

As for the second term in (70), we also have that

$$\begin{aligned} \int_{t_0 - k(\varepsilon)}^{t_0} \int_{-\infty}^{\infty} \psi(\mu(t) + u - \mu(t_0 - k(\varepsilon))) h(t) g(u) dt du &= \int_{t_0 - k(\varepsilon)}^{x^*} \int_{-\infty}^{\infty} \psi(u) h(t) g(u) dt du \\ &+ \int_{t_0 - k(\varepsilon)}^{t_0} \int_{-\infty}^{\infty} (\mu(t) - \mu(t_0 - k(\varepsilon))) g(u) h(t) \psi'(u + \gamma) du dt, \end{aligned}$$

where  $0 \leq \gamma \leq \mu(t_0) - \mu(t_0 - k(\varepsilon))$ . By A1 (vi) we have  $\int_{-\infty}^{\infty} \psi(u) g(u) du = 0$ , so that the first term above vanishes. As for the second term, notice that

$$\int_{t_0 - k(\varepsilon)}^{t_0} \int_{-\infty}^{\infty} (\mu(t) - \mu(t_0 - k(\varepsilon))) g(u) h(t) \psi'(u) du dt = \left[ \int_{t_0 - k(\varepsilon)}^{t_0} (\mu(t) - \mu(t_0 - k(\varepsilon))) h(t) dt \right] \left[ \int_{-\infty}^{\infty} \psi'(u + \gamma) g(u) du \right]. \tag{72}$$

The first integral factor in the right hand side of the above display can be further approximated. By the Mean Value Theorem, there exists  $\xi(t)$  such that  $t_0 - k(\varepsilon) \leq \xi(t) \leq t_0$  and

$$\begin{aligned} \int_{t_0 - k(\varepsilon)}^{t_0} (\mu(t) - \mu(t_0 - k(\varepsilon))) h(t) dt &= \int_{t_0 - k(\varepsilon)}^{t_0} \mu'(\xi(t)) (t - t_0 + k(\varepsilon)) h(t) dt \approx \int_{t_0 - k(\varepsilon)}^{t_0} \mu'(t_0) (t - t_0 + k(\varepsilon)) h(t) dt \\ &= \frac{\mu'(t_0) h(t_0)}{2} [(t - t_0 + k(\varepsilon))^2]_{t_0 - k(\varepsilon)}^{t_0} = \frac{1}{2} \mu'(t_0) h(t_0) k^2(\varepsilon). \end{aligned} \tag{73}$$

Using expressions (71)–(73) we obtain that Eq. (70), can be written as

$$\varepsilon [\psi(x^* - \mu(t_0)) - \psi'(x^* - \mu(t_0) - k(\varepsilon^*)) \mu'(t_0 - k(\varepsilon^*)) k(\varepsilon)] + (1 - \varepsilon) \frac{1}{2} \mu'(t_0) h(t_0) k^2(\varepsilon) \int_{-\infty}^{\infty} \psi'(u) g(u + \gamma) du = 0.$$

Dividing both sides of this equation by  $\varepsilon$  and using that  $k(\varepsilon) \rightarrow 0$  and  $\gamma \rightarrow 0$  when  $\varepsilon \rightarrow 0$  we obtain

$$\lim_{\varepsilon \rightarrow 0} \frac{k^2(\varepsilon)}{\varepsilon} = - \frac{2\psi(x^* - \mu(t_0))}{h(t_0) \mu'(t_0) \int_{-\infty}^{\infty} \psi'(u) g(u) du}. \tag{74}$$

Finally, according to (68) and using the Mean Value Theorem, we can write

$$\lim_{\varepsilon \rightarrow 0} \frac{(T_{t_0}(A_{\varepsilon, t_0, x^*}) - T_{t_0}(A_0))^2}{\varepsilon} = \lim_{\varepsilon \rightarrow 0} \frac{(\mu(t_0 - k(\varepsilon)) - \mu(t_0))^2}{\varepsilon} = \lim_{\varepsilon \rightarrow 0} \frac{\mu'(2(t^*(\varepsilon))k^2(\varepsilon))}{\varepsilon},$$

where  $t^*(\varepsilon) \rightarrow t_0$ . Then using Eq. (74) we obtain that

$$\text{IF}^*(T_{t_0}, t_0, x^*) = \lim_{\varepsilon \rightarrow 0} \frac{(T_{t_0}(A_{\varepsilon, t_0, x^*}) - T_{t_0}(A_0))^2}{\varepsilon} = -\frac{2\mu'(t_0)\psi(x_0 - \mu(t_0))}{h(t_0)E_G(\psi'(u))} = \frac{2\mu'(t_0)|\psi(x_0 - \mu(t_0))|}{h(t_0)E_G(\psi'(u))}.$$

The proof in the case that  $x^* < \mu(t_0)$  is similar.

We consider now the case  $t^* > t_0$ . To prove this part of the theorem is enough to show that when the contamination occurs at a point  $t^* > t_0$ , there exists  $\varepsilon^* > 0$ , so that  $\varepsilon \leq \varepsilon^*$  implies

$$T_{t_0}(A_{\varepsilon, t^*, x^*}) = T_{t_0}(A_0) = \mu(t_0)$$

and to prove this is enough to show that

$$\min_{s \geq 0} m(t_0, r, s, A_{\varepsilon, t^*, x^*}) = m(t_0, r, 0, A_{\varepsilon, t^*, x^*}) = m(t_0, r, 0, A_0). \quad (75)$$

When  $x^* \geq \mu(t_0)$ , this is immediate. Consider the case that  $x^* < \mu(t_0)$ .

Clearly for  $0 \leq s < t^*$

$$m(t_0, r, s, A_{\varepsilon, t^*, x^*}) = m(t_0, r, s, A_0) > m(t_0, r, 0, A_0). \quad (76)$$

It is also easy to show that  $s > t^*$  implies

$$m(t_0, r, s, A_{\varepsilon, t^*, x^*}) \geq m(t_0, r, t^*, A_{\varepsilon, t^*, x^*}) \quad (77)$$

and for  $r < 0$  and all  $s$

$$m(t_0, r, s, A_{\varepsilon, t^*, x^*}) < m(t_0, 0, s, A_{\varepsilon, t^*, x^*}). \quad (78)$$

Then, using (76)–(78) and the fact that  $m(t_0, 0, 0, A_0) = m(t_0, 0, 0, A_{\varepsilon, t^*, x^*})$ , in order to prove (75), it is enough to show that

$$m(t_0, 0, t^*, A_{\varepsilon, t^*, x^*}) > m(t_0, 0, 0, A_0). \quad (79)$$

Recall that  $m(t_0, 0, s, A_{\varepsilon, t^*, x^*})$  is the solution of

$$\varepsilon\psi(x^* - m)\mathbf{1}(t_0 \leq t^* \leq t_0 + s) + (1 - \varepsilon)V(s, m) = 0,$$

where

$$V(s, m) = \int_{-\infty}^{\infty} \int_{t_0}^{t_0 + s} \psi(\mu(t) + u - m) dA_0(t, u).$$

Clearly  $V(t^*, m(0, t^*, A_0)) = 0$  and since  $m(r, t, A_0)$  and  $V(r, t, m)$  are both increasing in  $t$  we get  $V(r, t^*, m(0, 0, A_0)) < 0$ . Then, since  $\psi$  is bounded, we can find  $\varepsilon^*$ , so that for  $\varepsilon < \varepsilon^*$  we have

$$\varepsilon\psi(x^* - m)\mathbf{1}(t_0 \leq t^* \leq t_0 + s) + (1 - \varepsilon)V(s, m(0, 0, A_0)) < 0,$$

and therefore  $m(0, t^*, A_{\varepsilon, t^*, x^*}) > m(0, 0, A_0)$ . Then (79) holds and this proves the Theorem for the case  $t^* > t_0$ . The proof for the case  $t^* < t_0$  is similar.

## Appendix B. Proof of Theorem 4

Without loss of generality we can assume that  $\sigma_0 = 1$ . It is easy to see that the least favorable contaminating distribution is  $A^*$  concentrated at  $\delta_{t_0, x_0}$  where  $x_0$  tends to  $-\infty$  or to  $\infty$ .

A necessary and sufficient condition for  $\varepsilon < \varepsilon^*$  is that the equation:

$$\varepsilon\psi(x_0 - m) + (1 - \varepsilon) \int_0^{t_0} \int_{-\infty}^{\infty} \psi(\mu(t) + u - m)h(t)g(u) dt du = 0 \quad (80)$$

have a bounded solution  $m$  solution for all  $x_0 < \mu(t_0)$  and that the equation:

$$\varepsilon\psi(x_0 - m) + (1 - \varepsilon) \int_{t_0}^1 \int_{-\infty}^{\infty} \psi(\mu(t) + u - m)h(t)g(u) dt du = 0 \quad (81)$$

have a solution for all  $x_0 > \mu(t_0)$ .

Taking  $x_0 \rightarrow -\infty$  we find that a sufficient condition for the existence of a bounded solution of (80) for all  $x_0 < \mu(t_0)$  is that

$$-\varepsilon k + (1 - \varepsilon)kH(t_0) \geq 0$$

and this is equivalent to

$$\varepsilon \leq \frac{H(t_0)}{1+H(t_0)}. \quad (82)$$

Taking  $x_0 \rightarrow \infty$  we obtain that a sufficient condition for the existence of solution of (81) for all  $x_0 > \mu(t_0)$  is that

$$\varepsilon k - (1-\varepsilon)k(1-H(t_0)) \leq 0$$

and this equivalent to

$$\varepsilon \leq \frac{1-H(t_0)}{2-H(t_0)}. \quad (83)$$

The theorem follows from (80) and (81).

## References

- Álvarez, E.E., Dey, D.K., 2009. Bayesian isotonic changepoint analysis. *Annals of the Institute of Statistical Mathematics* 61, 355–370.
- Boente, G., Fraiman, R., 1988. Robust nonparametric regression estimation. *Journal of Multivariate Analysis* 29, 180–198.
- Brunk, H.D., 1958. On estimation of parameters restricted by inequalities. *Annals of Mathematical Statistics* 29, 437–454.
- Brunk, H.D., 1970. Estimation of Isotonic Regression, Nonparametric Techniques in Statistical Inference. Cambridge University Press, pp. 177–195.
- De Boor, C., 2001. A Practical Guide to Splines. Springer, New York.
- Dette, H., Neumeier, N., Pilz, K.F., 2006. A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* 12 (3), 469–490.
- Dykstra, R., Carolan, C., 1998. The distribution of the arg max of two-sided Brownian motion with quadratic drift. *Journal of Statistical Computation and Simulation* 63, 47–58.
- Faraway, J.J., 2004. Linear Models with R. Chapman & Hall/CRC, Boca Raton, FL.
- Fomby, T.B., Vogelsang, T.J., 2002. The application of size-robust trend statistics to global-warming temperature series. *Journal of Climate* 15, 117–123.
- Friedman, J., Tibshirani, R., 1984. The monotone smoothing of scatterplots. *Technometrics* 26, 243–250.
- Ghement, I.R., Ruiz, M., Zamar, R.H., 2008. Journal of Statistical Planning and Inference 138, 3200–3216.
- Hall, P., Huang, L.S., 2001. Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics* 29, 624–647.
- Hampel, F.R., 1974. The influence curve and its role in robust estimation. *Journal of the American Statistical Association* 69, 383–393.
- Härdle, W., Tsybakov, A.B., 1988. Robust nonparametric regression with simultaneous scale curve estimation. *Annals of Statistics* 16 (1), 120–135.
- He, X., Shi, P., 1998. Monotone B-spline smoothing. *Journal of the American Statistical Association* 93 (442), 643–650.
- Kulikova, V.N., Lopuhaä, H.P., 2006. The limit process of the difference between the empirical distribution function and its concave majorant. *Statistics & Probability Letters* 76, 1781–1786.
- Leurgans, S., 1982. Asymptotic distributions of slope-of-greatest-convex-minorant estimators. *Annals of Statistics* 10 (1), 287–296.
- Mammen, E., 1991. Nonparametric regression under qualitative smoothness. *Annals of Statistics* 19 (2), 741–759.
- Maronna, R.A., Martin, R.D., Yohai, V.J., 2006. Robust Statistics Theory and Methods. John Wiley, New York.
- Meyer, M., 1996. Shape Restricted Inference with Applications to Nonparametric Regression, Smooth Nonparametric Regression, and Density Estimation. Ph.D. Thesis, Statistics, University of Michigan.
- Mukherjee, H., 1988. Monotone non-parametric regression. *Annals of Statistics* 16, 741–750.
- Pal, J.K., Woodroffe, M., 2006. On the distance between cumulative sum diagram and its greatest convex minorant for unequally spaced design points. *Scandinavian Journal of Statistics* 33, 279–291.
- Prakasa Rao, B.L.S., 1969. Estimation of a unimodal density. *Sankhyā A* 31, 23–36.
- Ramsey, J.O., 1998. Estimating smooth monotone functions. *Journal of the Royal Statistical Society Series B* 60, 365–375.
- Robertson, T., Waltman, P., 1968. On estimating monotone parameters. *Annals of Mathematical Statistics* 39, 1030–1039.
- Robertson, T., Wright, F.T., Dykstra, R.L., 1988. Order Restricted Statistical Inference. John Wiley, New York.
- Shorack, G.R., 2000. Probability for Statisticians. Springer, New York.
- Sun, J., Woodroffe, M., 1999. Testing uniformity versus a monotone density. *Annals of Statistics* 27 (1), 338–360.
- Van de Geer, S., 2000. Empirical Processes in M-Estimation. Cambridge University Press, Cambridge, UK.
- Wang, Y., Huang, J., 2002. Limiting distribution for monotone median regression. *Journal of Statistical Planning and Inference* 108, 281–287.
- Wright, F.T., 1981. The asymptotic behavior of monotone regression estimates. *Annals of Statistics* 9, 443–448.
- Wu, W.B., Woodroffe, M., Mentz, G.B., 2001. Isotonic regression: another look at the change-point problem. *Biometrika* 88, 793–804.