On a robust local estimator for the scale function in heteroscedastic nonparametric regression *

Graciela Boente[1], Marcelo Ruiz[2] and Ruben H. Zamar[3]

[1] *Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina*

[2] *Universidad Nacional de Río IV, Argentina*

[3] *University of British Columbia, Canada*

Abstract

When the data used to fit an heteroscedastic nonparametric regression model are contaminated with outliers, robust estimators of the scale function are needed in order to obtain robust estimators of the regression function and to construct robust confidence bands. In this paper, local $M-$estimators of the scale function based on consecutive differences of the responses, for fixed designs are considered. Under mild regularity conditions, the asymptotic behavior of the local $M-$estimators for general weight functions is derived.

Some key words: Heteroscedasticity; Local $M-$estimators; Nonparametric regression; Robust estimation.

Corresponding Author
Graciela Boente
Instituto de Cálculo, Ciudad Universitaria, Pabellón 2
Buenos Aires, C1428EHA
Argentina
email: gboente@dm.uba.ar
FAX 54-11-45763375

Running Head: Local $M-$scale estimation.

*This research was partially supported by Grants X-018 from the Universidad de Buenos Aires, PIP 0216 from CONICET and PICT 00821 from ANPCYT, Argentina and Discovery Grant of the Natural Sciences and Engineering Research Council of Canada.

1 Introduction

Consider the nonparametric regression model

$$Y_i = g(x_i) + U_i \sigma(x_i), \qquad 1 \le i \le n, \tag{1}$$

The estimation of the scale function, both in homoscedastic and heteroscedastic models, has become an essential problem, nearly as important as the estimation of g itself, for direct applications and also because the performance of the estimators of the regression function depends of the behavior of those of the scale function (see, Dette *et al.*, 1998).

Examples of scale estimation appear in diverse fields such as economy and engineering. Ruppert *et al.* (1997)reports on a study where the main interest is the analysis of data from a Monte Carlo simulation of turbulence. The estimation of the conditional variance of the particle speed given the position and its derivatives are essential. Ullah (1985) discuss data consisting of observations of individuals' annual income versus age, taken from the 1971 Canadian Population Census. Levine (2003) suggests that "variance estimation for such a data set is of some economic interest. It is a well known in labor economics that the discrepancy in individuals incomes depends primarily on educational level. Moreover, this difference tends to increase with age ".

In homoscedatic nonparametric regression, scale estimators based on differences are widely used (Hall *et al.*, 1990). These scale estimates are defined as

$$\widehat{\sigma}^2_{r,n} = \frac{1}{(n-r)} \sum_{i=m_1+1}^{n-m_2} \left(\sum_{k=-m_1}^{m_2} d_k Y_{i+k} \right)^2 ,$$

where $\{d_i\}_{i=-m_1}^{m_2}$ is a difference sequence of real numbers satisfying $\sum_{i=-m_1}^{m_2} d_i = 0$ and $\sum_{i=-m_1}^{m_2} d_i^2 = 1$, with $d_{-m_1} \neq 0$, $d_{m_2} \neq 0$ for m_1 and m_2 non–negative integers. The integer $r = m_1 + m_2$ is the estimator order. When $r = 1$, $\widehat{\sigma}^2_{r,n} = \widehat{\sigma}^2_{\text{RICE},n}$ is simply the well–known estimator proposed by Rice (1984)

$$\widehat{\sigma}^2_{\text{RICE},n} = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{i+1} - Y_i)^2 .$$

This class of scale estimators has been extended to heteroscedastic nonparametric models. See, for instance, Müller and Stadtmüller (1987), and Brown and Levine (2007), who considered local estimators based on kernel weights.

It is well known that scale estimators based on squared differences are not robust against outliers and inliers. Robust estimators of scale are needed, for instance, to detect outliers (Hannig and Lee, 2006), to provide robust estimators of the regression function (see Härdle and Gasser, 1984; Härdle and Tsybakov, 1988; Boente and Fraiman, 1989), or to improve the accuracy of bandwidth selectors when estimating g (see, among others, Boente *et al.*, 1997; Cantoni and Ronchetti, 2001; Leung *et al.*, 1993; Leung, 2005).

When the scale function is constant, Boente *et al.* (1997) proposed the robust scale estimator $\widehat{\sigma}_{\text{MSD},n} = q_{1/2} / \{\sqrt{2} \Phi^{-1}(3/4)\}$, where $q_{1/2}$ is the median of the absolute differences $|Y_{i+1} - Y_i|$, $1 \le i \le$

$n-1$. Also, for homoscedastic nonparametric regression models, Ghement *et al.* (2008) generalized the above estimators using a robust $M-$estimator based on differences defined as a solution $\widehat{\sigma}_0$ of

$$\frac{1}{n-1}\sum_{i=1}^{n-1}\chi\left(\frac{Y_{i+1}-Y_i}{a\,\widehat{\sigma}_0}\right)=b \tag{2}$$

where χ is a score function, a is a positive constant chosen to attain Fisher–consistency at the central model and b is a positive tuning constants that gives the robustness level of the estimator.

We consider the situation where the scale function is not necessarily constant and define local $M-$estimates of the scale function based on differences. Our estimators can be seen as the robust counterpart of the variance estimators of order 1 considered by Levine (2003) and Brown and Levine (2007) and are regression free, in the sense that they do not require previous nor simultaneous estimation of the regression function. Besides, their asymptotic distribution does not depend of the regression function. However, for small sample sizes, the performance of the estimators can be affected by the shape of the regression function g. As mentioned by Rousseeuw and Hubert (1986), similar situations exist in other models such as location–scale and linear regression models, in the sense that robust scale estimators are typically based on an initial estimator of the location or the regression parameters. However, as it is well known, robust location–free scale estimators are also available, see, for instance, Rousseeuw and Croux (1993). Rousseeuw and Hubert (1986) considered robust regression–free estimators of scale by considering triplets of data points. Our purpose is to construct robust estimators of the variance function under the heteroscedastic regression model (1) which do not depend on the choice of the regression estimators $\widehat{g}$. In some sense, our estimators are related to those considered by Rousseeuw and Croux (1993) for the location–scale model, but our estimates are based on $M-$functionals in a nonparametric setting.

Preliminary estimation of the scale function is motivated, basically, by two reasons. Simultaneous estimation of the regression and scale function substantially increases the algorithmic complexity and, in consequence, the computational time. Another reason, particularly important in the heteroscedastic context, is the possible lack of robustness of the regression function when considering simultaneous estimation. This conjecture is based on the fact that, in the location–scale model $Y=\mu+\sigma U$, when estimating simultaneously location and scale the location estimator $\widehat{\mu}$ does not attain a $1/2$ breakdown point (see Maronna *et al.*, 2006).

It should be noted that the asymptotic properties of the robust proposals are derived under mild conditions on the errors distribution, in particular, without imposing moments conditions. It is also worth noticing that our results are based on the asymptotic behavior of weighted sums of $r-$dependent random variables, and so, our proposal can easily be extended to robust estimators based on any difference orders. However, as mentioned by Dette (2002), "for moderate sample sizes the Rice (1984) and Gasser *et al.* (1986) estimates will be sufficient in most cases". Moreover, as it may be expected, the resistance of the estimators to contamination will decrease as the difference order increases, since contaminations propagate over the considered differences. This fact is analogous to the behavior observed in time series by Caliskan *et al.* (2009) who proposed estimators based on three consecutive observations attaining at most a breakdown point of 0.25, see also Gelper *et al.* (2009). Note also that the breakdown point of the estimators considered in Rousseeuw and Hubert (1986) is at most 20%. Hence, we shall develop the theory only for robust estimators based on differences of order 1.

The rest of the paper is organized as follows. Section 2, describes the robust local $M-$estimates of the scale function. In section 3, we discuss finite sample properties of the estimators, while in Section 4, the consistency and asymptotic distribution of our estimates are derived. Finally, Section 5 provides some concluding remarks. All the proofs are delayed to the Appendix.

2 The estimators and Robust Proposals

In this section, we introduce a family of robust estimators of the scale function $\sigma(x)$ which we call *local $M-$estimates of scale based on differences.* Throughout this paper, we consider observations satisfying model (1) with errors $\{U_i\}_{i\geq 1}$ having common distribution G from the gross error neighborhood $\mathcal{P}_\epsilon(F_0)$ defined as

$$\mathcal{P}_\epsilon(F_0) = \{G:\ G(y) = (1-\epsilon)F_0(y) + \epsilon H(y);\ H \in \mathcal{D}, y \in \mathbb{R}\}\ ,$$

where $\mathcal{D}$ denotes the set of all distribution functions, F_0 is the central model, generally the normal distribution, and H is any arbitrary distribution function modeling the contamination. The amount of contamination $\epsilon \in [0, 1/2)$ represents the fraction of outliers that we expect be present in the sample. Finally, G_x will denote the distribution of $\sigma(x)(U_2 - U_1)$ where U_1 and U_2 are independent random variables with common distribution G. Notice that, as mentioned in the Introduction, we do not assume the existence of moments for the errors distribution G nor the symmetry of the central model distribution F_0.

For $x \in (0, 1)$, we define the *local $M-$estimator of the scale function $\sigma(x)$* based on successive differences of the responses variables as

$$\widehat{\sigma}_{\mathrm{M},n}(x) = \inf\left\{s > 0:\ \sum_{i=1}^{n-1} w_{n,i}(x)\chi\left(\frac{Y_{i+1}-Y_i}{as}\right) \leq b\right\}, \tag{3}$$

where $\{w_{n,i}(x)\}_{i=1}^{n-1}$ is a sequence of weight functions (such as kernel or nearest neighbor weights), χ is a score function, the constants $a \in (0,\infty)$ and $b \in (0, 1)$ satisfy

$$E[\chi(Z_1)] = b \ \text{ and }\ E\left[\chi\left(\frac{Z_2 - Z_1}{a}\right)\right] = b, \tag{4}$$

with $\{Z_i\}_{i=1,2}$ independent random variables with common distribution $Z_1 \sim F_0$. Typically, the score function $\chi : \mathbb{R} \to \mathbb{R}$ is even with $\chi(0) = 0$, non-decreasing on $\mathbb{R}_+$ and $0 < \|\chi\|_\infty$ where $\|f\|_\infty = \sup_{x\in\mathbb{R}} |f(x)|$. It is worth noticing that the infimum in (3) is needed to define the estimates when the score function is discontinuous. When χ is continuous, it is easy to see that $\widehat{\sigma}_{\mathrm{M},n}(x)$ satisfies $\sum_{i=1}^{n-1} w_{n,i}(x)\chi\left((Y_{i+1} - Y_i)/(a\, \widehat{\sigma}_{\mathrm{M},n}(x))\right) = b$. Besides, the constant b is related to the robustness properties of the estimator while the constant a ensures the Fisher–consistency under the central model, as dicussed below.

Some Examples. Based on (3), in the sequel, we give some examples of local scale $M-$estimators.

i) When $\chi(x) = x^2$, $a = \sqrt{2}$ and $b = 1$, we obtain the classical *local Rice estimator*

$$\widehat{\sigma}_{\mathrm{RICE},n}(x) = \left[\sum_{i=1}^{n-1} w_{n,i}(x)\left(\frac{Y_{i+1}-Y_i}{\sqrt{2}}\right)^2\right]^{1/2}$$

ii) The proposal considered by Boente *et al.* (1997) can be extended to deal with heteroscedastic nonparametric regression models by choosing $\chi(y) = I_{\{u:\ |u|>\Phi^{-1}(3/4)\}}(y)$, $a = \sqrt{2}$ and $b = 1/2$ in (3). This estimator will be denoted by $\widehat{\sigma}_{\text{MSD},n}(x)$, and called from now on the *local median of the squared differences.*

iii) For $c > 0$ fixed, let

$$\chi_c(y) = \begin{cases} 3\,(y/c)^2 - 3\,(y/c)^4 + (y/c)^6 & \text{if } \ |y| \le c \\ 1 & \text{if } \ |y| > c \end{cases}$$

be the score function introduced by Beaton and Tukey (1974). Let $\widehat{\sigma}_{\text{BT},n}(x)$ stand for the *local $M-$estimador with BT function* that is, the solution of (3) with score function χ_c with $c = 0.70417$, $a = \sqrt{2}$ and $b = 3/4$.

Some Robustness Considerations. In Section 4, we show that, under regularity conditions, for all G in the contamination neighborhood, the sequence $\{\widehat{\sigma}_{\text{M},n}(x)\}_{n\ge1}$ converge almost surely to

$$S(G_x) = \inf\left\{\sigma > 0 :\ E\left[\chi\left(\frac{\sigma(x)\,(U_2 - U_1)}{a\sigma}\right)\right] \le b\right\}\,.$$

As mentioned above, if χ is a continuous function, $S(G_x)$ is the unique solution of

$$E\left[\chi\left(\frac{\sigma(x)\,(U_2 - U_1)}{aS(G_x)}\right)\right] = b\,. \tag{5}$$

For any fixed x denote $S(G) = S(G_x)$ with G the errors distribution and by $F_n(y|x)$ the empirical conditional distribution, $F_n(y|x) = \sum_{i=1}^{n-1} w_{n,i}(x) I_{(-\infty,y]}(Y_{i+1} - Y_i)$. Then, we have that $S(F_n(\cdot|x)) = \widehat{\sigma}_{\text{M},n}(x)$ and so, our estimator is related to a robust functional (defined on a wide class of distribution functions) in the sense that this functional is weakly continuous and such that at the central model F_0, S is Fisher–consistent, i.e., $S(F_0) = \sigma(x)$ which means that $\widehat{\sigma}_{\text{M},n}(x)$ estimates the true value $\sigma(x)$ at the central model. For a discussion regarding robust weakly continuous functionals in the nonparametric context see Boente *et al.* (1991).

When the scale function is constant, Ghement *et al.* (2008) showed that under certain regularity conditions and design restrictions, $M-$estimators of scale attain their maximum breakdown point of $1/2$ when $b = 3/4$. In heteroscedastic models, it might occur that the local breakdown point is lower, similar to local $M-$estimators of the regression function in nonparametric regression (see Boente and Rodriguez, 2008, and Maronna et al., 2006, Chapter 4). The empirical breakdown point and influence function of local $M-$estimates of scale are discussed in Sections 3 and 5.

3 Finite sample properties

Robust procedures are expected to be less sensitive to outliers than their classical counterparts. A popular measure of robustness is the finite sample breakdown point (BP). To investigate the resistance of our proposals to different amounts/sizes of contamination (and to get some insight about their finite sample BP) we conduct a simulation study comparing the performance of the classical estimator, $\widehat{\sigma}_{\text{RICE},n}(x)$, and two robust local $M-$estimators of the scale function, $\widehat{\sigma}_{\text{MSD},n}(x)$ and $\widehat{\sigma}_{\text{BT},n}(x)$, introduced in Section 2. We consider the regression model (1) with $g(x) = 2\text{sen}(4\pi x)$ and $\sigma(x) = \exp(x)$. This model has been considered for homoscedastic testing in Dette and Hetzler (2008). Similar results were obtained for others models (see Ruiz, 2008, for further details).

The design points are chosen as $x_i = i/(n+1)$, $1 \leq i \leq n$ while the error's distribution is $G(y) = (1-\epsilon)\Phi(y) + \epsilon\ H$, with Φ the standard normal distribution and H modeling two types of contamination,

a) a symmetric outlier contamination where $H(y) = \mathcal{C}(0, \tau^2)$ is the Cauchy distribution centered at 0 with scale $\tau = 4$ and

b) asymmetric contaminations where $H = \text{N}(\mu, \tau^2)$ is the normal distribution with means $\mu = 10, 100$ and $\mu = 1000$ and common variance $\tau = 0.1$.

In the first contamination scenario, we have a heavy–tailed distribution while, in the second one, there is a sub–population in data (see Maronna *et al.*, 2006). The amounts of contamination were $\epsilon = 0, 0.1, 0.2, 0.30, 0.35$ and 0.40. The main reason to incorporate high contaminations proportions and extremely asymmetric contaminations is to give some insight on the breakdown point of the estimators. The sample size considered is $n = 100$ and, the number of replications, $N = 10000$.

For both, the classical and robust estimators, we have used the Nadaraya–Watson weights, $w_{n,i}(x) = K\left((x - x_i)/h_n\right)\left[\sum_{j=1}^{n-1} K\left((x-x_j)/h_n\right)\right]^{-1}$, with a standard gaussian kernel. As in any smoothing procedure a value for the smoothing parameter must be selected. However, the study of data–driven bandwidth selectors for the scale function is less developed. When considering scale estimators based on squared differences, Levine (2006) recommended a version of $K-$fold cross–validation for selecting the smoothing parameter. As in nonparametric regression, this approach can be sensitive to outliers even when it is combined with robust scale estimators. The ideas of robust cross–validation can be adapted to the present situation, however, the study of robust selectors is beyond the scope of the paper. Based on extensive preliminary comparisons we selected a smoothing parameter $h_n = 0.20$ for our simulations.

To asses the behavior of each estimator Tables 1 and 2 report, as summary measures, the mean and the standard deviation of the integrated square error in logarithmic scale of the estimators, $\widehat{\text{ISEL}}$, defined as

$$\widehat{\text{ISEL}}_j(\widehat{\sigma}_n) = \frac{1}{n}\sum_{i=1}^{n}\left[\log\left(\frac{\widehat{\sigma}_n^{(j)}(x_i)}{\sigma(x_i)}\right)\right]^2$$

where $\widehat{\sigma}_n^{(j)}(x_i)$ denotes the scale estimator, classical or robust, obtained at the $j-$th replication.

As expected, under the central model, $\epsilon = 0$, the classical local Rice scale estimator performs better than the robust ones that show a loss of efficiency measured through the $\widehat{\text{ISEL}}$. On the other hand, the performance of the classical local Rice estimator is highly sensitive to the presence of outliers in the sample. When anomalous observations are present, regardless of the amount of contamination and the sample size $\widehat{\sigma}_{\text{RICE},n}$ has a very poor integrated square error, in both contamination scenarios. In particular, note that with only 10% of contamination the mean of the $\widehat{\text{ISEL}}(\widehat{\sigma}_{\text{RICE},n})$ suffers a considerable increase confirming the expected non–robustness of this estimator.

Estimator	$\epsilon = 0$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$	$\epsilon = 0.35$	$\epsilon = 0.40$
$\widehat{\sigma}_{\text{RICE},n}$	0.021 (0.017)	3.893 (5.467)	6.631 (6.894)	8.701 (7.623)	9.613 (7.999)	10.429 (8.231)
$\widehat{\sigma}_{\text{MSD},n}$	0.036 (0.030)	0.074 (0.060)	0.204 (0.138)	0.470 (0.271)	0.670 (0.356)	0.918 (0.442)
$\widehat{\sigma}_{\text{BT},n}$	0.052 (0.047)	0.082 (0.068)	0.177 (0.128)	0.357 (0.210)	0.487 (0.260)	0.647 (0.314)

Table 1: Mean and standard deviation (between brackets) of the $\widehat{\text{ISEL}}$ for the local scale–estimates under different amounts of symmetric contamination, i.e., when $G(y) = (1-\epsilon)\Phi(y) + \epsilon\ H$ and $H(y) = \mathcal{C}(0, \sigma^2)$ with $\sigma = 4$.

μ	Estimator	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$	$\epsilon = 0.35$	$\epsilon = 0.40$
10	$\widehat{\sigma}_{\text{RICE},n}$	1.353 (0.321)	2.060 (0.261)	2.453 (0.218)	2.574 (0.202)	2.656 (0.186)
	$\widehat{\sigma}_{\text{MSD},n}$	0.108 (0.088)	0.386 (0.352)	1.141 (0.862)	1.541 (1.015)	1.790 (1.097)
	$\widehat{\sigma}_{\text{BT},n}$	0.099 (0.081)	0.201 (0.147)	0.323 (0.236)	0.372 (0.283)	0.390 (0.307)
100	$\widehat{\sigma}_{\text{RICE},n}$	11.530 (1.152)	13.744 (0.722)	14.899 (0.675)	15.132 (0.492)	15.344 (0.443)
	$\widehat{\sigma}_{\text{MSD},n}$	0.118 (0.147)	0.829 (1.411)	5.235 (4.786)	6.380 (4.854)	8.333 (5.190)
	$\widehat{\sigma}_{\text{BT},n}$	0.107 (0.087)	0.291 (0.237)	1.002 (0.899)	1.229 (1.119)	1.636 (1.379)
1000	$\widehat{\sigma}_{\text{RICE},n}$	32.413 (2.001)	36.101 (1.180)	37.832 (0.878)	38.339(0.787)	38.678 (0.704)
	$\widehat{\sigma}_{\text{MSD},n}$	0.149 (0.329)	1.999 (3.480)	10.917 (9.110)	16.817 (10.638)	21.564 (10.964)
	$\widehat{\sigma}_{\text{BT},n}$	0.117 (0.108)	0.600 (0.891)	3.092 (3.166)	5.246 (4.286)	7.438 (4.888)

Table 2: Mean and standard deviation (between brackets) of the $\widehat{\text{ISEL}}$ for the local scale–estimates under different amounts of asymmetric contamination, i.e., when $G(y) = (1-\epsilon)\Phi(y) + \epsilon\ H$ and $H = \text{N}(\mu, \sigma^2)$, with $\mu = 10, 100, 1000$ and $\sigma^2 = 0.1$.

Under none or small (10%) symmetric contamination the behavior of $\widehat{\sigma}_{\text{MSD},n}$ and $\widehat{\sigma}_{\text{BT},n}$ are similar. On the other hand, under both contamination schemes, if the amount of the contamination is large, the local $M-$estimate $\widehat{\sigma}_{\text{BT},n}$ performs better than $\widehat{\sigma}_{\text{MSD},n}$, specially under asymmetric contamination. These results suggest that the breakdown point of $\widehat{\sigma}_{\text{MSD},n}$ is lower than that of $\widehat{\sigma}_{\text{BT},n}$.

Another useful robustness measure is the empirical influence function (EIF) introduced by Tukey (1977). EIF reflects the behavior of the estimator when a single sample point is replaced by a new observation that does not follow the original model.

We will follow an approach similar to that of Manchester (1996) who introduced a graphical method to display sensitivity of a kernel estimator in nonparametric regression. Given a data set $\{(x_i, y_i)\}_{1\le i\le n}$, let $\widehat{\sigma}(x)$ be the scale estimator computed at x with the Nadaraya–Watson weights. Thus, for a smooth $\chi-$function, the estimator $\widehat{\sigma}(x)$ is the solution of

$$\sum_{i=1}^{n-1} K\left(\frac{x - x_i}{h_n}\right)\left[\chi\left(\frac{Y_{i+1} - Y_i}{a\ \widehat{\sigma}(x)}\right) - b\right] = 0\ .$$

Assume that $\mathbf{z} = (x_0, y_0)$ represents a contaminating point with $x_0 \in [0, 1]$ and denote $\widehat{\sigma}_{\mathbf{z}}$ the scale estimator based on the augmented data set $\{(x_1, Y_1), \ldots (x_n, Y_n), \mathbf{z}\}$. Thus, if $x_{j_0} \le x_0 \le x_{j_0+1}$, we have that $\widehat{\sigma}_{\mathbf{z}}(x)$ is the solution of

$$\begin{aligned} 0 \;=\; & \sum_{1\le i\le n-1, i\ne j_0} K\left(\frac{x - x_i}{h_n}\right)\left[\chi\left(\frac{Y_{i+1} - Y_i}{a\ \widehat{\sigma}_{\mathbf{z}}(x)}\right) - b\right] \\ & + K\left(\frac{x - x_{j_0}}{h_n}\right)\left[\chi\left(\frac{y_0 - Y_i}{a\ \widehat{\sigma}_{\mathbf{z}}(x)}\right) - b\right] + K\left(\frac{x - x_0}{h_n}\right)\left[\chi\left(\frac{Y_{j_0+1} - y_0}{a\ \widehat{\sigma}_{\mathbf{z}}(x)}\right) - b\right]\ . \end{aligned}$$

In order to detect if a contaminating point influences the scale estimator, we can define the EIF of $\widehat{\sigma}(x)$ at (x_0, y_0) as

$$\text{EIF}(\widehat{\sigma}(x); (x_0, y_0)) = (n+1) \left|\log\left(\widehat{\sigma}_{\mathbf{z}}(x)\right) - \log\left(\widehat{\sigma}(x)\right)\right| \,.$$

The log function is introduced in order to study the influence to inliers. Figure 1 gives the surface plots for one of the samples generated under the central model described above, i.e., with $g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$, $x_i = i/(n+1)$, $n = 100$ and $\epsilon = 0$ when $x = 0.5$ to illustrate the performance at a central point. To build each surface plot, we consider a grid of values (x_0, y_0) taking values on a equidistant grid on each axis of size 40×200 on $[0.25, 0.75] \times [-100, 100]$. Thus, we have a grid of 800 points (x_0, y_0) and for each of them we have computed the empirical influence function, $\text{EIF}(\widehat{\sigma}(x); (x_0, y_0)))$ for each estimator.

As expected, the classical estimator based on square differences has an unbounded EIF, while the EIF of the robust alternatives related to bounded χ functions remain bounded. It is worth noticing that the irregularity showed by $\text{EIF}(\widehat{\sigma}_{\text{MSD},n}(x); (x_0, y_0)))$ may be related to the non–differentiability of the score function. Note that $\text{EIF}(\widehat{\sigma}_{\text{BT},n}(x); (x_0, y_0)))$ show larger values than $\text{EIF}(\widehat{\sigma}_{\text{MSD},n}(x); (x_0, y_0)))$, this fact may be related to the local–global robustness trade-off of $\widehat{\sigma}_{\text{BT},n}$. Besides, as it is well–known, the robust scale estimators may be sensitive to inliers, this feature corresponds to the behavior near $y_0 = 0$ of the EIF of both robust procedures. To give more insight on the behavior with respect to inliers, Figure 2 gives the surface plots constructed when considering a grid of values (x_0, y_0) taking values on a equidistant grid on each axis of size 40×200 on $[0.25, 0.75] \times [-5, 5]$. These plots confirm that the robust estimators may be sensitive to inliers even if their effect remains bounded. Besides the wiggly surface obtained for the $\widehat{\sigma}_{\text{MSD},n}$ near $y_0 = 0$ suggests that abrupt changes may arise when using this estimator.

4 Asymptotic behavior of the local $M-$estimates of the scale function

In this section, we derive consistency and asymptotic normality of the estimators defined in Section 2 at any distribution G from the gross–error neighborhood $\mathcal{P}_\epsilon(F_0)$, under mild conditions.

If I is an interval of $\mathbb{R}$, let $\mathcal{C}_L(I)$ be the set of bounded and Lipschitz continuous functions $f : I \to \mathbb{R}$ and denote by $\|f\|_L = \min\{k : |f(x) - f(y)| \le k|x - y|, \forall x, y \in I\}$. In order to establish the strong consistency of $\{\widehat{\sigma}_{\text{M},n}(x)\}_{n\ge 1}$, we will need the following assumptions

H1. The score function χ is continuous, even, bounded, strictly increasing on the set $C_\chi = \{x : \chi(x) < \|\chi\|_\infty\}$ with $\chi(0) = 0$. Without loss of generality, we assume that $\|\chi\|_\infty = 1$.

H2. The design points $\{x_i\}_{i=1}^n$ satisfy $\lim_{n\to\infty} M_n = 0$, where $M_n = \max_{1\le i\le n-1}(x_{i+1} - x_i)$.

H3. The regresion function $g : [0, 1] \longrightarrow \mathbb{R}$ is continuous.

H4. The scale function $\sigma : [0, 1] \to \mathbb{R}^+$ is continuous.

H5. The weights $\{w_{n,i}(x)\}_{i=1}^{n-1}$ are such that

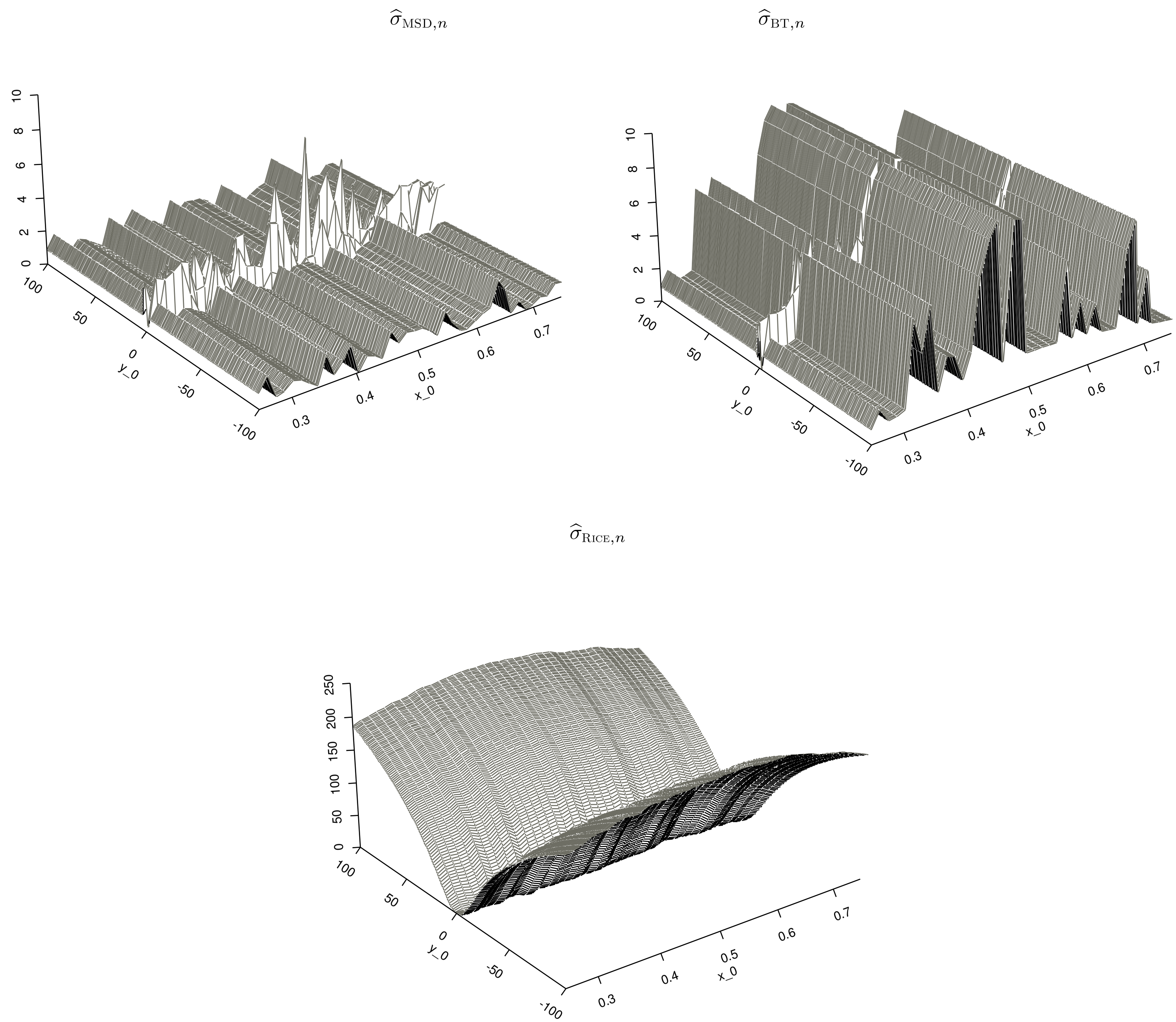

Figure 1: Empirical influence function of $\widehat{\sigma}(x)$ when $x = 0.5$.

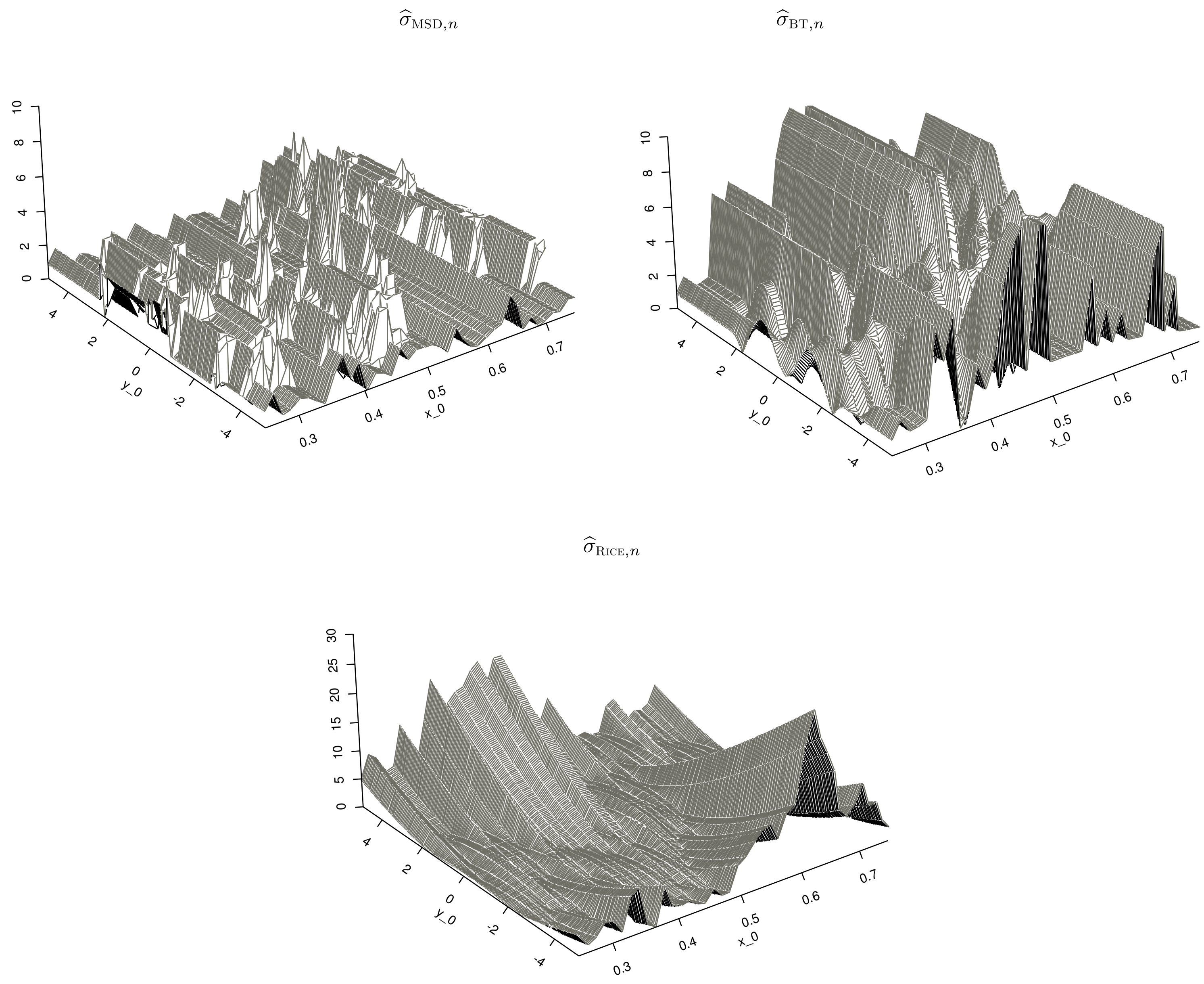

Figure 2: Empirical influence function of $\widehat{\sigma}(x)$ when $x = 0.5$.

(i) $\lim_{n\to\infty} \sum_{i=1}^{n-1} w_{n,i}(x) = 1$.

(ii) There exists $M > 0$ such that $\sum_{i=1}^{n-1} |w_{n,i}(x)| \leq M$, for all $n \geq 2$.

(iii) $\lim_{n\to\infty} \sum_{i=1}^{n-1} |w_{n,i}(x)| I_{\{|x_i - x| \geq a\}} = 0$, for any $a > 0$.

(iv) $\lim_{n\to\infty} w_n \log n = 0$, where $w_n = \max_{1\leq i\leq n-1} |w_{n,i}(x)|$.

Remark 4.1 Assumptions **H2**, **H3** and **H5** are standard conditions in nonparametric estimation. They have been considered, for instance, by Georgiev (1989) to derive the strong consistency of regression estimators. In particular, **H5** is fulfilled for the weight functions described in Section 2 if K has bounded support and the bandwidth sequence is such that $h_n \to 0$ and $nh_n/\log(n) \to \infty$ and $\max(x_{i+1} - x_i) \leq \Delta/n$. On the other hand, **H5**(ii) allows for kernels taking negative values, such as high order kernels or kernels modified to overcome boundary effects (see, for instance, Gasser and Müller, 1984). Assumption **H4** is a smoothness requirement on the scale function needed to guarantee consistency at any $x \in (0, 1)$.

Theorem 4.1 *Under **H1** to **H5**, given $x \in (0, 1)$, the local $M-$estimators are strongly consistent to $S(G_x)$ defined in* (5), *i.e.,* $\widehat{\sigma}_{\mathrm{M},n}(x) \xrightarrow{a.s.} S(G_x)$.

To derive the asymptotic distribution of the proposed local $M-$estimators, we will need some additional assumptions. From now on, we will denote by $c_n = \sum_{i=1}^{n-1} w_{n,i}^2(x)$.

H6. $g \in \mathcal{C}_L([0, 1])$.

H7. $M_n = \max_{1\leq i\leq n-1}(x_{i+1} - x_i) = O(n^{-1})$.

H8. χ is twice continuously differentiable and the functions $\chi_1(u) = u\chi'(u)$ and $\chi_2(u) = u^2\chi''(u)$ are bounded.

H9. The scale function $\sigma : [0, 1] \to \mathbb{R}^+$ satisfies (i) or (ii) where

(i) $\sigma \in \mathcal{C}_L([0, 1])$.

(ii) σ is continuous and $\lim_{n\to\infty} c_n^{-1/2} \sum_{i=1}^{n-1} |w_{n,i}(x)| \, |\sigma(x_{i+1}) - \sigma(x_i)| = 0$.

H10. Let $w_n = \max_{1\leq i\leq n-1} |w_{n,i}(x)|$.

(i) $\lim_{n\to\infty} c_n^{-1/2} w_n = 0$

(ii) $\lim_{n\to\infty} c_n^{-1/2} \left(\sum_{i=1}^{n-1} w_{n,i}(x) - 1\right) = 0$.

H11. The score function χ is such that $\nu(\alpha_1, \alpha_2) = E\,|\chi'(\alpha_1 U_1 + \alpha_2 U_2) U_2| < \infty$, for any $\alpha_1 \neq 0$, $\alpha_2 \neq 0$, where $\{U_i\}_{i=1,2}$ are i.i.d, $U_1 \sim G$.

H12. For any $x \in (0, 1)$ the following conditons hold

(i) $\lim_{n\to\infty} c_n^{-1/2} \sum_{i=1}^{n-1} w_{n,i}(x)(\sigma(x_i) - \sigma(x)) = \beta_1$

(ii) $\lim_{n\to\infty} c_n^{-1/2} \sum_{i=1}^{n-1} |w_{n,i}(x)| \ (\sigma(x_i) - \sigma(x))^2 = 0$.

Remark 4.2. It is worth noticing that **H7**, **H9**(i) and **H10**(i) entail **H9**(ii) which is needed when no differentiability conditions on σ are required. Moreover, **H10**(i) is needed to guarantee that the order of convergence is $c_n^{-1/2}$ while **H12**(i) deals with the asymptotic bias. Note that since

$$\nu(\alpha_1, \alpha_2) \leq \|\chi\,'\|_\infty \left[\frac{2c}{|\alpha_1|} + E\left(|U_2| \; I_{|\alpha_1 U_1 + \alpha_2 U_2| \leq c} \; I_{|\alpha_1 U_1| > 2c} \; I_{|\alpha_2 U_2| > c} \right) \right] < \infty \;, \tag{6}$$

if $\chi\,'(u) = 0$ for $|u| > c$, $\chi\,'$ is bounded and

$$E\left[|U_2| \; I_{|\alpha_1 U_1 + \alpha_2 U_2| \leq c} \; I_{|\alpha_1 U_1| > 2c} \; I_{|\alpha_2 U_2| > c} \right] < \infty \;, \tag{7}$$

is fulfilled for any α_1, α_2, then **H11** holds. Besides, the bound given in (6) and the fact that $\chi\,'$ is continuous entail that $\sup_{(\alpha_1,\alpha_2) \in \mathcal{K}_1 \times \mathcal{K}_2} \nu(\alpha_1, \alpha_2) < \infty$, for any compact set $\mathcal{K}_i \subset \mathbb{R} - \{0\}$. Note that the Beaton–Tukey family of score functions clearly satisfies the required conditions. Moreover, (7) is not as restrictive as it may seem, as it is fulfilled, for instance, when U_i has Cauchy distribution.

Theorem 4.2. *Let $x \in (0, 1)$ be fixed and let $c_n = \sum_{i=1}^{n-1} w_{n,i}^2(x)$. Assume that $\beta > 0$ and $v_i > 0$, $i = 1, 2$ where*

$$\begin{aligned}
\beta &= \lim_{n \to \infty} c_n^{-1} \sum_{i=1}^{n-2} w_{n,i+1}(x) w_{n,i}(x) \\
v_1 = v_1(G_x) &= \text{VAR}\left[\chi\left(\frac{\sigma(x) U_1^*}{a S(G_x)} \right) \right] + 2\beta \;\text{COV}\left[\chi\left(\frac{\sigma(x) U_1^*}{a S(G_x)} \right), \chi\left(\frac{\sigma(x) U_3^*}{a S(G_x)} \right) \right] \\
v_2 = v_2(G_x) &= E\left[\chi\,'\left(\frac{\sigma(x) U_1^*}{a S(G_x)} \right) \left(\frac{\sigma(x) U_1^*}{a (S(G_x))^2} \right) \right],
\end{aligned}$$

*with $U_1^* = U_2 - U_1$, $U_3^* = U_4 - U_3$ and $\{U_i\}_{i \geq 1}$ are i.i.d. random variables with distribution G. Let $v = v(G_x) = v_1 / v_2^2$. If, in addition, **H1** and **H5** to **H12** hold, we have that*

$$c_n^{-1/2} \left(\widehat{\sigma}_{\text{M},n}(x) - S(G_x) \right) \xrightarrow{\mathcal{D}} N\left(\frac{S(G_x) \beta_1}{\sigma(x)}, v \right)$$

*where β_1 is given in **H12**.*

Remark 4.3.

(a) Note that the asymptotic bias depends on χ only through the functional $S(G_x)$. Hence, at the central model, i.e., when $G = F_0$, the asymptotic bias is independent of the score function and, consequently, the asymptotic behavior of the sequence of $M-$estimates depends on χ only through its asymptotic variance.

(b) It is worth noticing that we do not obtain the usual expression for the asymptotic variance of the scale $M-$estimator based on independent observations. This fact can be explained by the intrinsic one–dependence, due to the responses differences appearing in each term of the estimator's definition, that leads to the second term in v_1.

5 Concluding Remarks

Robust estimation of the scale function, $\sigma(x)$, is an important problem in any nonparametric regression analysis. In this paper, for heteroscedastic models, we introduced a robust estimator for the scale function based on local $M-$scale estimators. These estimators are a robust version of the very well–known family of regression–free estimators based on responses differences (see, among others, Hall *et al.*, 1990 and Levine, 2003). They can also be seen as an extension to heteroscedastic models of the robust global $M-$scale estimators introduced for homoscedastic nonparametric regression models by Ghement *et al.*, 2008. Under mild regularity conditions, the local $M-$estimators turn is consistent and asymptotically normal.

As we mentioned in Section 2, robustness of the estimators can be considered in the sense of weak continuity of the scale functional. However, the determination of the breakdown point and influence function of local $M-$estimators of the scale function deserves a careful investigation as future work.

As Giloni and Simonoff (2005) indicate, when estimating the regression function, one possible approach to the breakdown point problem is to consider a conditional concept in the sense that, unlike for parametric models, the breakdown value changes depending on the evaluation point x. Although the simulation results suggest that the local $M-$estimator based on the Beaton–Tukey score function is more resistant than the local median of the squared differences, there still exists a need to define a local version of asymptotic breakdown point for scale functions, taking into account both implosion and explosion of the estimators.

Besides, when using kernel weights, the influence function of the estimator may be investigated by defining a smoothed influence function through a smoothed functional related to the kernel scale estimators as it was done for nonparametric regression by Aït Sahalia (1995) and Tamine (2002). However, unlike the notion of asymptotic breakdown point, a finite sample version of the influence function following the ideas of Tukey (1977) may be more adequate. Following the ideas of Manchester (1996) who introduced a graphical method to display sensitivity of a scatter plot smoother, we have defined an empirical influence function that takes into account the effect of both inliers and outliers on the scale estimator function.

Acknowledgment. The authors wish to thank an anonymous referee and the Associate Editor for valuable comments which led to an improved version of the original paper.

A Appendix

For the sake of simplicity, we will begin by fixing some notation. For any $i = 1, \ldots, n-1$, let

$$Y_i^* = Y_{i+1} - Y_i \ , U_i^* = U_{i+1} - U_i \text{ and } \widetilde{U}_i = \sigma(x_{i+1})U_{i+1} - \sigma(x_i)U_i. \tag{A.1}$$

Before proving Theorem 4.1, we state the following result whose proof follows easily using Hoeffding's inequality (Bosq, 1996, page 22).

Lemma A.1 *Let $Z_1, \ldots, Z_n$ be independent random variables with $E(Z_i) = 0$ and $\sup_{1 \le i \le n} |Z_i| \le c$. Let $\{w_{n,i}\}_{i=1}^{n-1}$ be a sequence of weights and $w_n = \max_{1 \le i \le n-1} |w_{n,i}(x)|$.*

*(a) If $\{w_{n,i}\}_{i=1}^{n-1}$ satisfies **H5**(ii) then, for any $\epsilon > 0$,*

$$P\left(\left|\sum_{i=1}^{n-1} w_{n,i} Z_i\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{\epsilon^2}{2c^2 M w_n}\right),$$

*(b) If, moreover, the sequence of weights verifies **H5**(iv), then $\sum_{i=1}^{n-1} w_{n,i} Z_i \to 0$, completely.*

PROOF OF THEOREM 4.1. Fix $x \in (0,1)$ and consider the (conditional) empirical distribution functions $F_n(y) = \sum_{i=1}^{n-1} w_{n,i}(x) I_{(-\infty,y]}(Y_i^*)$ and $\widetilde{F}_n(y) = \sum_{i=1}^{n-1} w_{n,i}(x) I_{(-\infty,y]}(\widetilde{U}_i)$ where we have omitted the dependence on x. Let π stand for the Prohorov distance. Note that **H1** entails that the functional S is weakly continuous and so, consistency will follow if

$$\pi(F_n, G_x) \xrightarrow{a.s.} 0 \,. \tag{A.2}$$

To derive (A.2), it is enough to show that

$$\pi(F_n, \widetilde{F}_n) \xrightarrow{a.s.} 0 \tag{A.3}$$

$$\pi(\widetilde{F}_n, G_x) \xrightarrow{a.s.} 0 \tag{A.4}$$

hold. Note that (A.3) holds since for any $f \in \mathcal{C}_L(\mathbb{R})$, $\int f(y) d(F_n - \widetilde{F}_n)(y) \xrightarrow{a.s.} 0$. Effectively, **H2**, **H3** and **H5**(ii) and the fact that

$$\left|\int f(y) d(F_n - \widetilde{F}_n)(y)\right| \leq \|f\|_L \sum_{i=1}^{n-1} |w_{n,i}(x)|\, |g(x_{i+1}) - g(x_i)| \ ,$$

entail $\int f(y) d(F_n - \widetilde{F}_n)(y) \xrightarrow{a.s.} 0$.

To obtain (A.4), it will be enough to prove that for any $f \in \mathcal{C}_L(\mathbb{R})$

$$S_n = \sum_{i=1}^{n-1} w_{n,i}(x) f(\widetilde{U}_i) - E\left[\sum_{i=1}^{n-1} w_{n,i}(x) f\left(\widetilde{U}_i\right)\right] \xrightarrow{a.s.} 0, \tag{A.5}$$

and

$$\lim_{n\to\infty} E\left[\sum_{i=1}^{n-1} w_{n,i}(x) f\left(\widetilde{U}_i\right)\right] = \int f dG_x = E\left[f\left(\sigma(x) U_1^*\right)\right] \tag{A.6}$$

hold.

Let us begin by showing that (A.5) holds. Write $S_n = S_{1,n} + S_{2,n}$, where $S_{j,n} = \sum_{i \in I_{j,n}} w_{n,i}(x) Z_i$ with $I_{1,n} = \{1 < i \leq n-1 : i \text{ is even }\}$ and $I_{2,n} = \{1 < i \leq n-1 : i \text{ is odd }\}$. Let $w_n = \max_{1 \leq i \leq n-1} |w_{n,i}(x)|$. Applying Lemma A.1(a) to each term $S_{j,n}$ we get

$$P(|S_n| > 2\epsilon) \leq P(|S_{1,n}| > \epsilon) + P(|S_{2,n}| > \epsilon) \leq 4\exp\left(-\frac{\epsilon^2}{2\|f\|_\infty^2 M w_n}\right)$$

which together with **H5**(iv) implies (A.5).

Finally, we will show that (A.6) holds. Note that

$$E\left[\sum_{i=1}^{n-1} w_{n,i}(x) f(\widetilde{U}_i)\right] - Ef(\sigma(x)U_1^*) \;=\; \sum_{i=1}^{n-1} w_{n,i}(x) E\left[f(\widetilde{U}_i) - f(\sigma(x)U_1^*)\right] + \left(\sum_{i=1}^{n-1} w_{n,i}(x) - 1\right) E\left[f(\sigma(x)U_1^*)\right].$$

By **H5**(i), the second term on the right hand side converges to zero. To prove that the first term also converges to 0, define $h(y) = Ef\big(\sigma(y)U_1^*\big)$ and write $\sum_{i=1}^{n-1} w_{n,i}(x)E\left[f\big(\widetilde{U}_i\big) - f\big(\sigma(x)U_1^*\big)\right] = T_{1,n}+T_{2,n}$ where $T_{1,n} = \sum_{i=1}^{n-1} w_{n,i}(x)\,(h(x_i) - h(x))$ and

$$T_{2,n} \;=\; \sum_{i=1}^{n-1} w_{n,i}(x)E\left[f\big((\sigma(x_{i+1}) - \sigma(x_i))U_2 + \sigma(x_i)U_1^*\big) - f\big(\sigma(x_i)U_1^*\big)\right]\ .$$

Since h is a bounded and continuous function, we have that, given $\epsilon > 0$ there exists $\delta > 0$ such that if $|x - y| < \delta$ then $|h(x) - h(y)| < \epsilon/M$, where M is given in **H5**(ii) and so,

$$|T_{1,n}| \le 2\,\|h\|_\infty \sum_{i=1}^{n-1} |w_{n,i}(x)| I_{\{|x_i-x|\ge\delta\}} + \epsilon\ ,$$

which together with **H5**(iii), implies that $T_{1,n} \to 0$.

To conclude the proof, it remains to show that $T_{2,n} \to 0$. For all $1 \le i \le n-1$, let $u_i = (\sigma(x_{i+1}) - \sigma(x_i))U_2 + \sigma(x_i)U_1^*$ and $v_i = \sigma(x_i)U_1^*$. Given $\epsilon > 0$, let $k > 0$ be such that $P(|U_2| > k) \le \epsilon/(4\,\|f\|_\infty M)$ and $\delta > 0$ such that if

$$|s - t| < \delta \Rightarrow |\sigma(s) - \sigma(t)| < \epsilon/(2Mk\,\|f\|_L)\ . \tag{A.7}$$

Hence, using **H5**(ii) and the Lipschitz continuity of f, we get that

$$|T_{2,n}| \le \|f\|_L \sum_{i=1}^{n-1} |w_{n,i}(x)|\, E\Big[\big|\sigma(x_{i+1}) - \sigma(x_i)\big|\big|U_2\big| I_{\{|U_2|\le k\}}\Big] + \epsilon/2\ . \tag{A.8}$$

On the other hand, **H2** implies that there exists $n_0 \in \mathbb{N}$ such that $M_n = \max_{1\le i\le n-1}(x_{i+1} - x_i) < \delta$, for $n \ge n_0$. Hence, using (A.7), we obtain that the first term of the roght hand side of (6) can be majorized by $\epsilon/2$ which implies that $\lim_{n\to\infty} |T_{2,n}| = 0$ concluding the proof. □

To derive the asymptotic distribution of the local scale $M-$estimators, we will need the following

Lemma. For any $s > 0$ and $x \in (0,1)$, define

$$\begin{aligned}\lambda_{n,b}(s,x) &= \sum_{i=1}^{n-1} w_{n,i}(x)\chi\left(\frac{Y_i^*}{as}\right) - b \\ \lambda_{n,b}^*(s,x) &= \sum_{i=1}^{n-1} w_{n,i}(x)\chi\left(\frac{\sigma(x)U_i^*}{as}\right) - b \\ \lambda_{1,n}(s,x) &= \sum_{i=1}^{n-1} w_{n,i}(x)\chi'\left(\frac{\sigma(x)U_i^*}{as}\right)\left(\frac{U_i^*}{as}\right)(\sigma(x_i)-\sigma(x))\end{aligned}$$

where Y_i^* and U_i^*, $1 \le i \le n-1$ are as in (A.1).

Lemma A.2 *Under the assumptions **H1**, **H5**(ii), **H6** to **H10**(i), **H11** and **H12**, we have that*

$$c_n^{-1/2}\lambda_{n,b}(s,x) = c_n^{-1/2}\lambda_{n,b}^*(s,x) + c_n^{-1/2}\lambda_{1,n}(s,x) + o_p(1) \qquad \text{(A.9)}$$

$$c_n^{-1/2}\lambda_{1,n}(s,x) = \beta_1 E\left[\chi'\left(\frac{\sigma(x)U_1^*}{as}\right)\left(\frac{U_1^*}{as}\right)\right] + o_p(1). \qquad \text{(A.10)}$$

PROOF. To show (A.9) it is enough to prove that

$$c_n^{-1/2}\lambda_{n,b}(s,x) = c_n^{-1/2}\widetilde{\lambda}_{n,b}(s,x) + o_p(1) \qquad \text{(A.11)}$$

$$c_n^{-1/2}\widetilde{\lambda}_{n,b}(s,x) = c_n^{-1/2}\lambda_{n,b}^*(s,x) + c_n^{-1/2}\lambda_{1,n}(s,x) + o_p(1), \qquad \text{(A.12)}$$

where $\widetilde{\lambda}_{n,b}(s,x) = \sum_{i=1}^{n-1} w_{n,i}(x)\chi\left(\frac{\widetilde{U}_i}{as}\right) - b$ and $\widetilde{U}_i = \sigma(x_{i+1})U_{i+1} - \sigma(x_i)U_i$, $1 \le i \le n-1$ as in (A.1).

Using **H6**, **H8** and **H5**(ii) we conclude that

$$\left|c_n^{-1/2}\lambda_{n,b}(s,x) - c_n^{-1/2}\widetilde{\lambda}_{n,b}(s,x)\right| \le (as)^{-1}\|g\|_L \|\chi\|_L (nM_n)c_n^{-1/2}w_n$$

where $w_n = \max_{1\le i\le n-1} |w_{n,i}(x)|$ and $M_n = \max_{1\le i\le n-1}(x_{i+1}-x_i)$. Thus, (A.11) follows from **H7** and **H10**(i). Note now that (A.12) will follow if we prove that $H_n = o_p(1)$ and $T_n = c_n^{-1/2}\lambda_{1,n}(s,x) + o_p(1)$ where

$$\begin{aligned}H_n &= c_n^{-1/2}\sum_{i=1}^{n-1} w_{n,i}(x)\left[\chi\left(\frac{\sigma(x_{i+1})U_{i+1} - \sigma(x_i)U_i}{as}\right) - \chi\left(\frac{\sigma(x_i)U_i^*}{as}\right)\right] \\ T_n &= c_n^{-1/2}\sum_{i=1}^{n-1} w_{n,i}(x)\left[\chi\left(\frac{\sigma(x_i)U_i^*}{as}\right) - \chi\left(\frac{\sigma(x)U_i^*}{as}\right)\right],\end{aligned}$$

Using **H5**(ii), **H9** and **H11** we get that

$$E|H_n| \le (as)^{-1}\sup_{1\le i\le n-1}\nu(\alpha_i,\beta_i)c_n^{-1/2}\sum_{i=1}^{n-1}|w_{n,i}(x)|\,|\sigma(x_{i+1}) - \sigma(x_i)|$$

where $\xi_i = \alpha_i U_2 + \beta_i U_1$, $\alpha_i = \widetilde{\sigma}_i/(as)$, $\beta_i = -\sigma(x_i)/(as)$ and $\widetilde{\sigma}_i$ is an intermediate point between $\sigma(x_{i+1})$ and $\sigma(x_i)$. As σ is continuous and strictly positive on the interval $[0,1]$, there exist compact sets $\mathcal{K}_j \subset \mathbb{R} - \{0\}$, $j = 1, 2$, such that $\alpha_i \in \mathcal{K}_1$ and $\beta_i \in \mathcal{K}_2$, $1 \le i \le n-1$. Thus, $\sup_{1\le i\le n-1} \nu(\alpha_i, \beta_i) < \sup_{(\alpha,\beta)\in\mathcal{K}_1\times\mathcal{K}_2} \nu(\alpha, \beta) < \infty$ which together with **H7**, **H9** and **H10**(i) implies that $E|H_n| \to 0$.

It remains to show that $T_n = c_n^{-1/2}\lambda_{1,n}(s,x) + o_p(1)$. From a second order Taylor's expansion, we have that

$$T_n = c_n^{-1/2}\lambda_{1,n}(s,x) + c_n^{-1/2}\sum_{i=1}^{n-1} w_{n,i}(x)\chi_2\left(\frac{\widetilde{\sigma}_i U_i^*}{as}\right)\frac{1}{\widetilde{\sigma}_i^2}\left(\sigma(x_i) - \sigma(x)\right)^2$$

with $\chi_2(u) = u^2\chi''(u)$ and $\widetilde{\sigma}_i$ an intermediate point between $\sigma(x_i)$ and $\sigma(x)$, $i = 1, \ldots, n-1$. Hence, noticing that $\inf_{u\in(0,1)} \sigma^2(u) > 0$ and using **H8** we obtain that

$$E\left|T_n - c_n^{-1/2}\lambda_{1,n}(s,x)\right| \le \frac{\|\chi_2\|}{\inf_{u\in(0,1)} \sigma^2(u)} c_n^{-1/2}\sum_{i=1}^{n-1} |w_{n,i}(x)| \left(\sigma(x_i) - \sigma(x)\right)^2 ,$$

which together with **H12**(ii) imply that $T_n - c_n^{-1/2}\lambda_{1,n}(s,x) = o_p(1)$.

We now prove (A.10). Let $Z_i = \chi'\left(\frac{\sigma(x)U_i^*}{aS_x}\right)\left(\frac{U_i^*}{aS_x}\right)$ and write $c_n^{-1/2}\lambda_{1,n}(s,x) = H_n + T_n$ with

$$H_n = c_n^{-1/2}\sum_{i=1}^{n-1} w_{n,i}(x)E(Z_i)\left(\sigma(x_i) - \sigma(x)\right) \text{ and } T_n = c_n^{-1/2}\sum_{i=1}^{n-1} w_{n,i}(x)(Z_i - E(Z_i))\left(\sigma(x_i) - \sigma(x)\right).$$

Since, $E(Z_i) = E(Z_1)$, from **H12**(i) we obtain easily that $H_n \longrightarrow \beta_1 E(Z_1)$. Besides, **H10**(i) and **H12** imply that $\mathrm{VAR}\left[T_n\right] \to 0$ and so, $T_n \xrightarrow{p} 0$, concluding the proof. □

Proof of Theorem 4.2. Fix $x \in (0,1)$ and let $S_x = S(G_x)$. Noting that the local $M-$estimator $\widehat{\sigma}_{\mathrm{M},n}(x)$ satisfies $\lambda_{n,b}\left(\widehat{\sigma}_{\mathrm{M},n}(x), x\right) = 0$, a Taylor's expansion of order one yields

$$0 = \lambda_{n,b}(\widehat{\sigma}_{\mathrm{M},n}(x), x) = \lambda_{n,b}(S_x, x) + \left(\widehat{\sigma}_{\mathrm{M},n}(x) - S_x\right)\lambda'_{n,b}(\widehat{\sigma}_{0,n}, x) ,$$

with $\lambda'_{n,b}(s,x) = \frac{\partial}{\partial s}\lambda_{n,b}(s,x) = -\frac{1}{s}\sum_{i=1}^{n-1} w_{n,i}(x)\chi'\left(\frac{Y_i^*}{as}\right)\left(\frac{Y_i^*}{as}\right)$ and $\widehat{\sigma}_{0,n}$ an intermediate value between S_x and $\widehat{\sigma}_{\mathrm{M},n}(x)$. Hence, $c_n^{-1/2}\left(\widehat{\sigma}_{\mathrm{M},n}(x) - S_x\right) = -c_n^{-1/2}\lambda_{n,b}(S_x, x)/\lambda'_{n,b}(\widehat{\sigma}_{0,n}, x)$ and, in consequence, it is enough to prove that

$$c_n^{-1/2}\lambda_{n,b}(S_x, x) \xrightarrow{\mathcal{D}} N\left(S_x v_2 \beta_1/\sigma(x), v_1\right) \tag{A.13}$$

$$-\lambda'_{n,b}(\widehat{\sigma}_{0,n}, x) \xrightarrow{p} v_2 . \tag{A.14}$$

Lemma A.2 implies that $c_n^{-1/2}\lambda_{n,b}(S_x, x) = c_n^{-1/2}\lambda^*_{n,b}(S_x, x) + c_n^{-1/2}\lambda_{1,n}(S_x, x) + o_p(1)$. Using that

$$c_n^{-1/2}\lambda_{1,n}(S_x, x) \xrightarrow{p} \beta_1 E\left[\chi'\left(\frac{\sigma(x)U_1^*}{aS_x}\right)\left(\frac{U_1^*}{aS_x}\right)\right] = S_x v_2\beta_1/\sigma(x)$$

and

$$\begin{aligned} c_n^{-1/2}\lambda_{n,b}^*(S_x,x) &= V_n^{1/2}\sum_{i=1}^{n-1} a_{n,i}\xi_i + bc_n^{-1/2}(\sum_{i=1}^{n-1} w_{n,i}(x)-1) \\ &= B_{1,n}+B_{2,n} \end{aligned}$$

with $a_{n,i} = V_n^{-1/2}c_n^{-1/2}w_{n,i}(x)$, $\xi_i = \chi(\sigma(x)U_i^*/(aS_x)) - b$ and $V_n = \text{VAR}\left[\sum_{i=1}^{n-1} c_n^{-1/2}w_{n,i}(x)\xi_i\right]$ to derive (A.13) it will be enough to show that

$$V_n \to v_1 \tag{A.15}$$

$$\sum_{i=1}^{n-1} a_{n,i}\xi_i \xrightarrow{\mathcal{D}} N(0,1)\ , \tag{A.16}$$

since **H10**(ii) implies that $B_{2,n}\to 0$. Considering that

$$V_n \;=\; \text{VAR}\left[\chi\left(\frac{\sigma(x)U_1^*}{aS_x}\right)\right] + 2\text{COV}\left[\chi\left(\frac{\sigma(x)U_1^*}{aS_x}\right),\chi\left(\frac{\sigma(x)U_2^*}{aS_x}\right)\right]\ c_n^{-1}\sum_{i=1}^{n-2} w_{n,i}(x)w_{n,i+1}(x).$$

and that $c_n^{-1}\sum_{i=1}^{n-2} w_{n,i}(x)w_{n,i+1}(x)\to\beta$, (A.15) follows. To obtain (A.16) we will use Theorem 2.2 in Pelligrad and Utev (1997). As $V_n\to v_1>0$, without loss of generality, we can assume that $\inf_{n>1} V_n>0$. So,

$$\sup_{n>1}\sum_{i=1}^{n} a_{ni}^2 = \sup_{n>1} V_n^{-1}c_n^{-1}\sum_{i=1}^{n} w_{n,i}^2(x) = 1/\inf_{n>1} V_n < \infty\ .$$

On the other hand (A.15) and **H10**(i) imply that $\max_{1\le i\le n}|a_{n,i}|\to 0$. It is straightforward to check that $\{\xi_i\}_{i\ge 1}$ is a uniformly square–integrable, φ–mixing (it is one–dependent) zero–mean sequence of random variables, satisfying that $\text{VAR}\left(\sum_{i=1}^{n} a_{ni}\xi_i\right)=1$. Therefore, the assumptions of Theorem 2.2 of Pelligrad and Utev (1997) are fulfilled and, in consequence, (A.16) holds, concluding the proof of (A.13).

Let us show now (A.14). Denote $\eta(t)=(t/a)\chi'(t/a)$. As $\widehat{\sigma}_{0,n}\xrightarrow{p} S_x$, we have only to prove that

$$\sum_{i=1}^{n-1} w_{n,i}(x)\eta\left(\frac{Y_i^*}{\widehat{\sigma}_{0,n}}\right) \xrightarrow{p} E\Big[\eta\left(\frac{U_1^*\sigma(x)}{S_x}\right)\Big]. \tag{A.17}$$

Define, for any $s\in\mathbb{R}^+$, $f(s)=\sum_{i=1}^{n-1} w_{n,i}(x)\eta\left(Y_i^*/s\right)$. Using a Taylor's expansion of order one, we obtain

$$\begin{aligned} \sum_{i=1}^{n-1} w_{n,i}(x)\eta\left(\frac{Y_i^*}{\widehat{\sigma}_{0,n}}\right) &= \sum_{i=1}^{n-1} w_{n,i}(x)\eta\left(\frac{Y_i^*}{S_x}\right) + \frac{1}{\widehat{\sigma}_{0,n}^{(1)}}\sum_{i=1}^{n-1} w_{n,i}(x)\eta'\left(\frac{Y_i^*}{\widehat{\sigma}_{0,n}^{(1)}}\right)\left(\frac{Y_i^*}{\widehat{\sigma}_{0,n}^{(1)}}\right)(S_x-\widehat{\sigma}_{0,n}) \\ &= A_{1,n}+A_{2,n}(S_x-\widehat{\sigma}_{0,n}) \end{aligned}$$

with $\widehat{\sigma}_{0,n}^{(1)}$ an intermediate point between S_x and $\widehat{\sigma}_{0,n}$. Note that $\widehat{\sigma}_{0,n}^{(1)} \xrightarrow{p} S_x$. On the other hand, $h(t) = -a\eta'(t)t = \left[(t/a)\chi'(t/a) + (t/a)^2\chi''(t/a)\right]$ is bounded, so $A_{2,n} = O_p(1)$ which entails that, to obtain (A.17), it will be enough to show that $A_{1,n} = \sum_{i=1}^{n-1} w_{n,i}(x)\eta\left(\frac{Y_i^*}{S_x}\right) \xrightarrow{a.s.} E\left[\eta\left(\frac{U_1^*\sigma(x)}{S_x}\right)\right]$ which follows using analogous arguments to those considered to derive (A.2) in Theorem 4.1, since $\eta \in \mathcal{C}_L(\mathbb{R})$ by **H8**. $\square$

References

[1] Aït Sahalia, Y. (1995). The delta method for nonlinear kernel functionals. PhD. dissertation, University of Chicago.

[2] Beaton, A. and Tukey, J. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, **16**, 147–185.

[3] Boente, G. and Fraiman, R. (1989). Robust Nonparametric Regression Estimation for Dependent Observations. *Annals of Statistics*, **17**, 1242–1256.

[4] Boente, G. and Fraiman, R. (1991). A functional approach to robust nonparametric regression. *Directions in robust statistics and diagnostics.* Berlin/Heidelberg. Springer.

[5] Boente, G., Fraiman, R. and Meloche, J. (1997). Robust plug-in bandwidth estimators in nonparametric regression. *Journal of Statistical Planning and Inference*, **57**, 109–142.

[6] Boente, G., Rodriguez, D. (2008). Robust bandwidth selection in semiparametric partly linear regression models: Monte Carlo study and influential analysis. *Computational Statistics & Data Analysis*, **52**, 2808–2828 .

[7] Brown, L. and Levine, M. (2007). Variance Estimation in Nonparametric Regression via the Difference Sequence Method. *Annals of Statistics*, **35**, 2219–2232.

[8] Caliskan, D., Croux, C. and Gelper, S. (2009). Efficient and robust scale estimation for trended time series. *Statistics and Probability Letters*, **79**, 19001905.

[9] Cantoni, E. and Ronchetti, E. (2001). Resistant selection of the smoothing parameter for smoothing splines. *Statistics and Computing*, **11**, 141–146.

[10] Dette, H. (2002). A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. *Journal of Statistical Planning and Inference*, **103**, 311–329.

[11] Dette, H., Munk, A. and Wagner, T. (1998). Estimating the variance in nonparametric regression - What is a reasonable choice? *Journal of the Royal Statistics Society, Series B*, **60**, 751–764.

[12] Gasser, T. and Müller, H. G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, **11**, 171185.

[13] Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, **73**, 625–633.

[14] Gelper, S., Schettlinger, K., Croux, C. and Gather, U. (2009). Robust online scale estimation in time series: A regression-free approach. *Journal of Statistical Planning and Inference*, **139**, 335-339.

[15] Ghement, I., Ruiz, M. and Zamar, R. (2008). Robust Estimation of Error Scale in Nonparametric Regression Models. *Journal of Statistical Planning and Inference*, **138**, 3200–3216.

[16] Giloni, A. and Simonoff, J. (2005) The conditional breakdown properties of least absolute value local polynomial estimators. *Journal of Nonparametric Statistics*, **17**, 15–30.

[17] Hall, P., Kay, J.and Titterington, D. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, **77**, 521–528.

[18] Hannig, J. and Lee, T. (2006). Robust SiZer for Exploration of Regression Structures and Outlier Detection. *Journal of Computational and Graphical Statistics*, **15**, 101–117.

[19] Härdle, W. and Gasser, T. (1984). Robust nonparametric function fitting. *Journal of the Royal Statistical Society, Series B*, **46**, 42–51.

[20] Härdle, W. and Tsybakov, A. (1988). Robust Nonparametric Regression with Simultaneous Scale Curve Estimation. *Annals of Statistics*, **25**, 443–456.

[21] Leung, D. (2005). Cross–validation in nonparametric regression with outliers. *Annals of Statistics*, **33**, 2291–2310.

[22] Leung, D., Marriott, F. and Wu, E. (1993). Bandwidth selection in robust smoothing. *Journal of Nonparametric Statistics*, **2**, 333–339.

[23] Levine, M. (2003). *Variance estimation for nonparametric regression and its applications*. Ph.D. Dissertation, University of Pennsylvania.

[24] Levine, M. (2006). Bandwidth selection for a class of difference-based variance estimators in the nonparametric regression: A possible approach. *Computational Statistics & Data Analysis*, **50**, 3405-3431.

[25] Manchester, L. (1996). Empirical Influence for robust smoothing. *Austral. J. Statist.*, **38**, 275-296.

[26] Maronna, R., Martin, D. and Yohai, V. (2006). *Robust Statistics: Theory and Methods*, John Wiley & Sons.

[27] Martin, R. and Zamar, R. (1989). Asymptotically Min–Max Bias Robust $M-$Estimates of Scale for Positive Random Variables. *Journal of the American Statistical Association*, **84**, 494–501.

[28] Müller, H. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Annals of Statistics*, **15**, 610–625.

[29] Nadaraya, E. (1964). On Estimating Regression. *Theory of Probability and its Applications*, **9**, 141–142.

[30] Pelligrad, M. and Utev, S. (1997). Bandwidth Choice for Nonparametric Regression. *Annals of Statistics*, **25**, 443–456.

[31] Priestley, M. and Chao, M. (1972). Nonparametric function fitting. *Journal of the Royal Statistical Society, Series B*, **34**, 384–392.

[32] Rice, J. (1984). Bandwidth Choice for Nonparametric Regression. *Annals of Statistics*, **12**, 1215–1230.

[33] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, **27**, 832–837.

[34] Rousseeuw, P. and Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statististical Association*, **88**, 12731283.

[35] Rousseeuw, P. and Hubert, M. (1996). Regression-free and robust estimation of scale for bivariate data. *Computational Statistics & Data Analysis*, **21**, 6785.

[36] Ruiz, M. (2008). Contribución a la Teoría de la Estimación Robusta de Escala en Modelos de Regresión Noparamétricos. Ph.D. Dissertation.

[37] Ruppert, D., Wand, M.,Holst, U. and Hössjer. (1997). Local polynomial variance–function estimation. *Technometrics*, **39**, 262–273.

[38] Tamine, J. (2002). Smoothed influence function: another view at robust nonparametric regression. Discussion paper 62, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin.

[39] Tukey, J. (1977). *Exploratory Data Analysis*. Reading, MA: Addison–Wesley.

[40] Ullah, A. (1985) Specification analysis of econometric models. *Journal of quantitative economics*, **2**, 187–209