Bandwidth choice for robust nonparametric scale function estimation*

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

Some key words: Cross–validation; Data–driven bandwidth; Heteroscedasticity; Local $M-$estimators; Nonparametric regression; Robust estimation.

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

Running Head: Robust 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 heteroscedastic nonparametric regression model

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

where $0 \le x_1 \le \ldots \le x_n \le 1$ are fixed design points in $[0,1]$ and the errors $\{U_i\}_{i\ge 1}$ are i.i.d. random variables with common distribution F_0. Both functions g and σ are continuous functions defined on $[0,1]$ and they are assumed to be unknown.

In most situations, nonparametric regression research has focused on the estimation of the function g and the variance function σ is treated as a nuisance parameter. There are reasons why this approach is often not completely satisfactory. To begin with, the homoscedasticity assumption usually assumed by many authors may not be a viable option. Besides, this approach fails to take into account that confidence intervals or prediction intervals require to have precise enough local variance estimators. Mainly, two general reasons explain why the estimation of scale function has become an important problem. One reason is that the performance of the estimation procedures of the regression function relies on the behavior of the scale estimators. Another reason appears when, in direct applications, the scale is the main parameter of the phenomena under study. Discussions and applications of scale estimations can be found, for example, in Carroll and Ruppert (1988), Hall and Marron (1990), Dette *et al.* (1998) and Levine (2003).

Hall *et al.* (1990), for homoscedatic nonparametric regression models, proposed preliminary scale estimators based on differences, generalizing the initial proposals of Rice (1984) and Gasser *et al.* (1986). Considering local estimators based on kernel weights Müller and Stadtmüller (1987), Brown and Levine (2007), among others, extend this class to heteroscedastic nonparametric models.

In both types of models, homoscedastic and heterocedastic, these scale estimators are not robust under departures of the central distribution F_0 of the errors $\{U_i\}_{i\ge 1}$. There exists a wide literature discussing the necessity of the introduction of robust procedures of the scale function. For instance, Boente *et al.* (1997), Cantoni and Ronchetti (2001), Leung *et al.* (1993) and Leung (2005) show how, in the estimation of the regression function, robust scale estimators improve the accuracy of bandwidth selectors. Also, Härdle and Gasser (1984), Härdle and Tsybakov (1988) and Boente and Fraiman (1989) provide robust procedures for g using robust estimators of the scale σ that involve the estimation of the regression function.

For homoscedastic nonparametric regression models, i.e., when $\sigma(x) \equiv \sigma_0$ and based on the ideas relying behind the scale estimators considered by Rice (1984), Boente *et al.* (1997) proposed, as a robust estimator, the median of the absolute differences $|Y_{i+1} - Y_i|$, $1 \le i \le n-1$. Subsequently, Ghement *et al.* (2008) introduced a more general robust class, known as the $M-$estimator based on differences. Finally, Boente *et al.* (2010) consider the situation in which the scale function does not need to be constant and define local $M-$estimates of scale based on differences.

As in nonparametric regression estimation, all these estimators depend on a smoothing parameter that needs to be choosen by the practitioner. As it is well known, large bandwidths produce estimators with small variance but high bias, while small values produce more wiggly curves. This trade–off between bias and variance lead to several proposals to select the smoothing parameter, such as cross-validation procedures and plug–in methods when estimating the regression function g. However, when estimating the variance function, less development has been obtained until now.

Levine (2003) derived an expression for the optimal bandwidth that leads to the plug–in approach discussed in Levine (2006). Therein, also a $K-$fold cross–validation procedure was recommended. All these procedures are based on square differences and thus, are very sensitive to outliers, as it is the case when estimating the regression function. The aim of this paper is to propose bandwidth selectors resistant to atypical observations that combined with the estimators defined in Boente *et al.* (2010) lead to robust data–driven scale estimators.

The rest of the paper is organized as follows. In Section 2, we briefly remind the definition of the robust local $M-$estimates of the scale function used in subsequent sections. In Section 3, we discuss several robust procedures in order to select the smoothing parameter when using kernel weights. The asymptotic properties of the robust local $M-$estimates based on random bandwidths are investigated in Section 4. The results of a Monte Carlo study designed to compare the behavior of the different bandwidth selectors are described in Section 5. Finally, Section 6 provides some concluding remarks. Proofs can be found in the Appendix.

2 The estimators

In this section, we remind the robust estimators of the scale function $\sigma(x)$, i.e., the *local $M-$estimates of scale based on differences*, defined by Boente *et al.* (2010). Throughout this paper, we consider observations satisfying model (1) with errors $\{U_i\}_{i\geq 1}$ having common distribution G from the gross error neighbourhood $\mathcal{P}_\epsilon(F_0)$ 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 modelling the contamination. The amount of contamination $\epsilon \in [0, 1/2)$ represents the fraction of outliers that we expect be present in the sample.

For $x \in (0,1)$, the *local $M-$estimator of the scale function* $\sigma(x)$ based on successive differences of the responses variables is defined as $\widehat{\sigma}_{\mathrm{M},n}(x) = \inf \left\{ s > 0 : \ \sum_{i=1}^{n-1} w_{n,i}(x)\chi\left(\frac{Y_{i+1}-Y_i}{as}\right) \leq b \right\}$, where $w_{n,i}(x) = w_{n,i}(x, h_n)$, $i = 1, \ldots, n-1$, are kernel weights, such as the Nadaraya–Watson or the Rosenblatt weights, defined respectively as $w_{n,i}(x,h) = K\left((x-x_i)/h\right)\left\{\sum_{j=1}^n K\left((x-x_j)/h\right)\right\}^{-1}$ and $w_{n,i}(x,h) = (nh)^{-1}K\left((x-x_i)/h\right)$. The non–negative parameter h_n is the bandwidth parameter that regulates the trade–off between bias and variance of the estimator. Moreover, χ is a score function, $a \in (0,\infty)$ is chosen to attain Fisher–consistency at the central model and $b \in (0,1)$ gives the robustness level of the estimator, and both tunning constants satisfy $\mathbb{E}[\chi(Z_1)] = b$ and $\mathbb{E}\left[\chi\left((Z_2 - Z_1)/a\right)\right] = b$, with $\{Z_i\}_{i=1,2}$ i.i.d. random variables with $Z_1 \sim F_0$. Note that, when χ is a smooth function, $\widehat{\sigma}_{\mathrm{M},n}(x)$ satisfies

$$\sum_{i=1}^{n-1} w_{n,i}(x)\chi\left(\frac{Y_{i+1}-Y_i}{a\widehat{\sigma}_{\mathrm{M},n}(x)}\right) = b. \tag{2}$$

Remark 2.1. The family of estimators defined through (2), include among others the classical *local Rice estimator* taking $\chi(x) = x^2$, $a = \sqrt{2}$ and $b = 1$. Two other estimators will also be considered in the simulation study, the *local* MAD *estimator* and the *local $M-$estimador with BT function*. The *local* MAD *estimator*, denoted $\widehat{\sigma}_{\mathrm{MAD},n}(x)$, corresponds to $\chi(y) = \mathbb{I}_{\{u:\ |u|>\Phi^{-1}(3/4)\}}(y)$, $a = \sqrt{2}$ and $b = 1/2$. On the other hand, the *local $M-$estimador with BT function*, denote by $\widehat{\sigma}_{\mathrm{BT},n}(x)$, is related to the score function χ_c introduced by Beaton and Tukey (1974) with tuning constant $c = 0.70417$, $a = \sqrt{2}$ and $b = 3/4$ where the Beaton–Tukey function is defined as

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

3 Robust bandwidth selectors.

As mentioned in the introduction, an important issue regarding kernel weights is the selection of the smoothing parameter h_n, that should be chosen by the practitioner. As it is well known, when the accuracy of the estimators of the regression function is measured, a trade–off between bias and variance occurs, leading to different alternatives to select the smoothing parameter, such as cross-validation procedures and plug–in methods. A detailed exposition on these alternatives can be found in Härdle (1990) and Härdle *et al.* (2004).

However, these procedures may not be robust and their sensitivity to anomalous data was discussed by several authors, including Leung *et al.* (1993), Wang and Scott (1994), Boente *et al.* (1997), Cantoni and Ronchetti (2001) and Leung (2005). Wang and Scott (1994) note that, when estimating the regression function, in the presence of outliers, the least squares cross–validation function is nearly constant on its whole domain and thus, essentially worthless for the purpose of choosing a bandwidth.

The study of data–driven bandwidth selectors for the scale function is less developed. When considering scale estimators based on squared differences, Levine (2003) obtained an expression for the optimal bandwidth. This expression leads to the plug–in approach discussed in Levine (2006), who also mentioned its disadvantages and recommended a version of $K-$fold cross–validation for selecting the smoothing parameter. This method produces a variance estimator that, in typical cases, is not very sensitive to the choice of the mean function.

For the sake of completeness, we remind the $K-$fold cross–validation method considered in Levine (2006). Partition the data set $\{(x_i, y_i)\}$ at random into K approximately equal and disjoint subsets, the $j-$th subset having size $n_j \ge 2$, $\sum_{j=1}^{K} n_j = n$. Let $\{(\widetilde{x}_i^{(j)}, \widetilde{y}_i^{(j)})\}_{1\le i\le n_j}$ be the pairs of the $j-$th subset with the values of $\widetilde{x}_i^{(j)}$ arranged in ascending order. Similarly, let $\{(x_i^{(j)}, y_i^{(j)})\}_{1\le i\le n-n_j}$ denote the pairs in the complement of the $j-$th subset, again with the $x_i^{(j)}$ arranged in ascending order. The set $\{(x_i^{(j)}, y_i^{(j)})\}_{1\le i\le n-n_j}$ will be the training set and $\{(\widetilde{x}_i^{(j)}, \widetilde{y}_i^{(j)})\}_{1\le i\le n_j}$ the validation set. Moreover, denote $\Delta_i^{(j)} = (y_{i+1}^{(j)} - y_i^{(j)})/\sqrt{2}$ and $\widetilde{\Delta}_i^{(j)} = (\widetilde{y}_{i+1}^{(j)} - \widetilde{y}_i^{(j)})/\sqrt{2}$ the successive differences within each subset. Let $\widehat{\sigma}_{\mathrm{RICE},n}^{(j)}(x,h)$ and $\widehat{\sigma}_{\mathrm{M},n}^{(j)}(x,h)$ be the classical and robust scale estimators computed using a bandwidth h and the $j-$th training subset $\{(x_i^{(j)}, y_i^{(j)})\}_{1\le i\le n-n_j}$, i.e., using the successive differences $\Delta_i^{(j)}$, respectively. The classical $K-$fold cross–validation criterion as described in Levine (2006) is defined as

$$CV_{\mathrm{LS,KCV}}(h) = \frac{1}{n}\sum_{j=1}^{K}\sum_{i=1}^{n_j-1}\left[\left(\widetilde{\Delta}_i^{(j)}\right)^2 - \left(\widehat{\sigma}_{\mathrm{RICE},n}^{(j)}(\widetilde{x}_i, h)\right)^2\right]^2 .$$

The $K-$fold cross–validation bandwidth is defined as $\widehat{h}_{\text{LS,KCV}} = \text{argmin}_{h\in\mathcal{H}} CV_{\text{LS,KCV}}(h)$, where $\mathcal{H}$ is the grid of possible values in $[0,1]$ over which we perform the search.

An alternative $K-$fold cross–validation procedure can be considered by measuring the deviances in log scale, i.e., we consider as measure

$$CV^{\log}_{\text{LS,KCV}}(h) = \frac{1}{n}\sum_{j=1}^{K}\sum_{i=1}^{n_j-1}\left[\log\left(\widetilde{\Delta}_i^{(j)}\right) - \log\left(\widehat{\sigma}^{(j)}_{\text{RICE},n}(\widetilde{x}_i, h)\right)\right]^2$$

and defining $\widehat{h}^{\log}_{L,\text{LS,KCV}} = \text{argmin}_{h\in\mathcal{H}} LCV_{\text{LS,KCV}}(h)$.

Even if robust scale estimators are considered, i.e., even if we perform the cross–validation with $\widehat{\sigma}^{(j)}_{\text{M},n}(x,h)$ instead of $\widehat{\sigma}^{(j)}_{\text{RICE},n}(x,h)$, the $K-$fold cross–validation bandwidth will be sensitive to outliers, since high residuals corresponding to an atypical observation are not downweighted. For that reason, when using a robust estimator, one needs to define robust scale–based procedures as a robust alternative to the classical $K-$th fold cross–validation criterion. This can be done by defining

$$CV_{\text{ROB,KCV}}(h) = \frac{1}{n}\sum_{j=1}^{K} s_j^2(h)\sum_{i=1}^{n_j-1}\psi^2\left(\frac{e_i^{(j)}}{s_j(h)}\right) ,$$

where $e_i^{(j)} = \left(\widetilde{\Delta}_i^{(j)}\right)^2 - \left(\widehat{\sigma}^{(j)}_{\text{M},n}(\widetilde{x}_i,h)\right)^2$, $s_j(h) = \text{median}\,|e_i^{(j)}|$ and ψ is a bounded score function such as the Huber function. The robust cross–validation bandwidth is then defined as $\widehat{h}_{\text{ROB,KCV}} = \text{argmin}_{h\in\mathcal{H}} CV_{\text{ROB,KCV}}(h)$.

As above, a robust log–scale cross–validation procedure can be considered as

$$CV^{\log}_{\text{ROB,KCV}}(h) = \frac{1}{n}\sum_{j=1}^{K} s_j^2(h)\sum_{i=1}^{n_j-1}\psi^2\left(\frac{e_i^{(j)}}{s_j(h)}\right) ;$$

where now $e_i^{(j)} = \log\left(\widetilde{\Delta}_i^{(j)}\right) - \log\left(\widehat{\sigma}^{(j)}_{\text{M},n}(\widetilde{x}_i,h)\right)$, $s_j(h) = \text{median}\,|e_i^{(j)}|$. Related to this criterion, the robust selector is defined as $\widehat{h}^{\log}_{\text{ROB,KCV}} = \text{argmin}_{h\in\mathcal{H}} CV_{\text{ROB,KCV}}(h)$.

As a particular case of the $K-$fold method, the classical leave–one–out cross–validation produces an asymptotically optimal data–driven bandwidth, $\widehat{h}_{\text{LS,CV}}$, by minimizing

$$CV_{\text{LS,CV}}(h) = \frac{1}{n-1}\sum_{i=1}^{n-1}\left[D_i^2 - (\widehat{\sigma}^{(-i)}_{\text{RICE},n}(x_i,h))^2\right]^2 \tag{3}$$

where $D_i = |Y_{i+1} - Y_i|/\sqrt{2}$ and $\sigma^{(-i)}_{\text{RICE},n}(x_i,h)$ is the classical estimator of $\sigma(x_i)$ computed with a bandwidth h and based on all the observations except (x_i, Y_i) and (x_{i+1}, Y_{i+1}), i.e., on the sample $(x_1,Y_1),\ldots,(x_{i-1},Y_{i-1})$, $(x_{i+2},Y_{i+2}),\ldots,(x_n,Y_n)$. The cross–validation method in log–scale is obtained by using

$$CV^{\log}_{\text{LS,CV}}(h) = \frac{1}{n-1}\sum_{i=1}^{n-1}\left[\log(D_i) - \log(\widehat{\sigma}^{(-i)}_{\text{RICE},n}(x_i,h))\right]^2 \tag{4}$$

with $\widehat{h}^{\log}_{\text{LS,CV}}$ denoting the optimal bandwidth.

On the other hand, the robust alternatives to the classical cross-validation procedures can be defined as follows. The robust bandwidth selector, $\widehat{h}_{\text{ROB,CV}}$, minimizes

$$CV_{\text{ROB,CV}}(h) = \frac{s_n^2(h)}{n-1}\sum_{i=1}^{n-1}\psi^2\left(\frac{u_i}{s_n(h)}\right) \tag{5}$$

where $u_i = (D_i)^2 - (\widehat{\sigma}^{(-i)}_{\text{M},n}(x_i,h))^2$ with $\sigma^{(-i)}_{\text{M},n}(x_i,h)$ a robust scale estimator computed with a bandwidth h and based on all the observations except (x_i,Y_i) and (x_{i+1},Y_{i+1}) and $s_n(h) = \text{median}\,|u_i|$.

Besides, $\widehat{h}^{\log}_{\text{ROB,CV}}$ minimizes the related robustified criterion on log–scale, i.e., $\widehat{h}^{\log}_{\text{ROB,CV}} = \text{argmin}_{h\in\mathcal{H}}\, CV^{\log}_{\text{ROB,CV}}(h)$ where

$$CV^{\log}_{\text{ROB,CV}}(h) = \frac{s_n^2(h)}{n-1}\sum_{i=1}^{n-1}\psi^2\left(\frac{u_i}{s_n(h)}\right)$$

where now $u_i = \log(D_i) - \log(\widehat{\sigma}^{(-i)}_{\text{M},n}(x_i,h))$.

A robust cross-validation criterion similar to that considered by Bianco and Boente (2007) for partly linear autoregression models and by Boente and Rodriguez (2008) in partly linear regression models can be defined. This approach splits the cross-validation error into two components, one related to the bias and the other one to the variance. Taking into account this fact, a robust alternative to the classical $K-$th fold cross–validation criterion, the split $K-$th fold, can be defined as

$$CV_{\text{ROB,SKCV}}(h) = \sum_{j=1}^{K}\mu^2_{n_j}(e_1^{(j)},\ldots,e_{n_j}^{(j)}) + \tau^2_{n_j}(e_1^{(j)},\ldots,e_{n_j}^{(j)})\ ,$$

where $e_i^{(j)} = \left(\widetilde{\Delta}_i^{(j)}\right)^2 - \left(\widehat{\sigma}^{(j)}_{\text{M},n}(\widetilde{x}_i,h)\right)^2$, $\tau_n(z_1,\ldots,z_n)$ and $\mu_n(z_1,\ldots,z_n)$ are robust scale and location estimators of the sample $z_1,\ldots,z_n$, such as a tau-scale and the median. The robust split $K-$fold cross–validation bandwidth is then defined as $\widehat{h}_{\text{ROB,SKCV}} = \text{argmin}_{h\in\mathcal{H}}\, CV_{\text{ROB,SKCV}}(h)$.

In an analogous way, the log-version split $K-$fold cross–validation criterion can be defined as

$$CV^{\log}_{\text{ROB,SKCV}}(h) = \sum_{j=1}^{K}\mu^2_{n_j}(e_1^{(j)},\ldots,e_{n_j}^{(j)}) + \tau^2_{n_j}(e_1^{(j)},\ldots,e_{n_j}^{(j)})\ ,$$

where now $e_i^{(j)} = \log\left(\widetilde{\Delta}_i^{(j)}\right) - \log\left(\widehat{\sigma}^{(j)}_{\text{M},n}(\widetilde{x}_i,h)\right)$; so, $\widehat{h}^{\log}_{\text{ROB,SKCV}} = \text{argmin}_{h\in\mathcal{H}}\, CV^{\log}_{\text{ROB,SKCV}}(h)$.

Also, one may define the leave–one–out split versions as a particular case of the two last criteria by considering $K = n$. The cross–validation errors will be denoted as $CV_{\text{ROB,SCV}}$ and $CV^{\log}_{\text{ROB,SCV}}$, leading to bandwidths $\widehat{h}_{\text{ROB,SCV}}$ and $\widehat{h}^{\log}_{\text{ROB,SCV}}$, respectively.

4 Asymptotic behaviour of data–driven local scale $M-$estimates

Based on deterministic bandwidths sequences, Boente *et al.* (2010) derived the asymptotic behavior of the local scale $M-$estimators based on differences. Under mild conditions, the estimators turn out to be strongly consistent and asymptotically normal distributed. However, in many situations as those described above, the bandwidth parameter is not fixed but random. The aim of this section is to extend the results regarding the asymptotic behavior of the local scale $M-$estimators when the smoothing parameter is random. To be more precise, if we denote by $\widehat{\sigma}_{\mathrm{M},n}(x,h_n)$ the robust local $M-$estimates computed using a sequence of bandwidths h_n, we are interested on deriving the asymptotic properties of the robust local $M-$estimates $\widehat{\sigma}_{\mathrm{M},n}(x,\widehat{h}_n)$ where $\widehat{h}_n=\widehat{h}_n(Y_1,...,Y_n)$ stands for a random bandwidth. Our results are related to the results obtained for the regression function by Boente and Fraiman (1995).

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

For simplicity, we will only consider the Rosenblatt' s weights function defined as $w_{n,i}(x,h)=(nh)^{-1}K\left((x-x_i)/h\right)$.

In order to derive the consistency of the estimators, we will need the following conditions.

C1. (i) $K:\mathbb{R}\longrightarrow\mathbb{R}$ is even, bounded and such that $\int|K(u)|du<\infty$, $\int K^2(u)du<\infty$ and $\lim_{u\to\infty}u^2K(u)=0$.
(ii) $\int K(u)du=1$.
(iii) K is continuously differentiable with derivative K' and the function $K_1(u)=uK'(u)$ is such that K_1 and K_1^2 satisfy (i).

C2. χ is Lipschitz continuous.

C3. The design points satisfy $M_n=\max_{1\le i\le n-1}|x_{i+1}-x_i|=O(n^{-1})$.

C4. There exists a sequence $\{h_n\}_{n\ge1}$ of real numbers such that

(i) $\widehat{h}_n/h_n\xrightarrow{p}1$
(ii) $\lim_{n\to\infty}nh_n=+\infty$ and $\lim_{n\to\infty}h_n=0$.

It is worth noticing that **C1** are standard conditions when dealing with kernel weights. Assumptions **C3** and **C4** were also considered in Boente and Fraiman (1995). The following result establishes the consistency of the data–driven estimators.

Theorem 4.1. *Let U_1 and U_2 be i.i.d. random variables with distribution G and G_x the distribution of $\sigma(x)(U_2-U_1)$. Assume that **C1** to **C4** hold, and, moreover, that $\lim_{n\to\infty}nh_n/\log(n)=\infty$ and $\widehat{h}_n/h_n\xrightarrow{a.s.}1$. Then, for every $x\in(0,1)$,*

$$\widehat{\sigma}_{\mathrm{M},n}(x,\widehat{h}_n)\xrightarrow{a.s.}S(G_x)$$

where $S(G_x)$, the solution of $\mathbb{E}\left[\chi\left(\sigma(x)\left(U_2-U_1\right)/(aS(G_x))\right)\right]=b$, is the Huber scale functional.

In order to obtain the asymptotic distribution of the data–driven estimators we will need some additional assumptions.

N1. g and σ are Lipschitz continuous functions.

N2. The score function χ is twice continuously differentiable with first and second derivatives χ' and χ'' such that

i) $\chi_1(u) = u\chi'(u)$ and $\chi_2(u) = u^2\chi''(u)$ are bounded.

ii) for any $u \neq 0$, $v \neq 0$, $\nu(u,v) = \mathbb{E}\,|\chi'(uU_2 + vU_1)U_2| < \infty$, where $\{U_i\}_{i=1,2}$ are i.i.d, $U_1 \sim G$.

N3. For any $x \in (0,1)$, the following limits exist

(i) $\lim_{n\to\infty} (nh_n)^{-1/2} \sum_{i=1}^{n-1} K\left(\frac{x-x_i}{h_n}\right)(\sigma(x) - \sigma(x_i)) = \beta_1$

(ii) $\lim_{n\to\infty} (nh_n)^{-1/2} \sum_{i=1}^{n-1} K\left(\frac{x-x_i}{h_n}\right)(\sigma(x_i) - \sigma(x))^2 = 0.$

Remark 4.1. The hypothesis **N1** is usual in non-parametric setting. Note also that **N2** does not necessarily impose the existence of moments of the distribution of the errors; for instance this hypothesis is fulfilled if the errors $\{U_i\}_{i\geq 1}$ have Cauchy distribution and χ belongs to the Beaton–Tukey family. **N3** is related to the asymptotic bias. Assume that $\int u^2 K(u)du < \infty$. If $nh_n^3 \to \gamma^2$, where γ is some finite constant, and the scale function is continuously differentiable then $\beta_1 = 0$ (since the kernel is an even function). Therefore, there is no asymptotic bias when the order of the bandwidth is $n^{-1/3}$. On the other hand, if $nh_n^5 \to \gamma^2$ and $\sigma(x)$ is twice continuously differentiable, then $\beta_1 = \gamma\sigma''(x)\int u^2K(u)du\left(\int K^2(u)du\right)^{1/2}$.

Theorem 4.2. *Assume **C3**, **C4** and **N1** to **N3** hold. Then*

$$(nh_n)^{1/2}\left[\widehat{\sigma}_{\mathrm{M},n}(x,\widehat{h}_n) - S(G_x)\right] \xrightarrow{\mathcal{D}} N\left(\frac{S(G_x)}{\sigma(x)}\beta_1, v\int K^2(u)du\right)$$

where $v = v(G_x) = v_1/v_2^2$, with

$$\begin{aligned} v_1 = v_1(G_x) &= \mathrm{VAR}\left[\chi\left(\frac{\sigma(x)(U_2-U_1)}{aS(G_x)}\right)\right] + 2\beta\,\mathrm{COV}\left[\chi\left(\frac{\sigma(x)(U_2-U_1)}{aS(G_x)}\right), \chi\left(\frac{\sigma(x)(U_4-U_3)}{aS(G_x)}\right)\right] \\ v_2 = v_2(G_x) &= \mathbb{E}\left[\chi'\left(\frac{\sigma(x)(U_2-U_1)}{aS(G_x)}\right)\left(\frac{\sigma(x)(U_2-U_1)}{a(S(G_x))^2}\right)\right], \end{aligned}$$

$\beta = \int K^2(u)du$ and $\{U_i\}_{i\geq 1}$ are i.i.d. random variables with distribution G.

5 Monte Carlo Study

In this section, a simulation study is carried out to compare, for moderate sample sizes, the performance of the classical estimator, $\widehat{\sigma}_{\text{RICE},n}(x,\widehat{h}_n)$, with that of two robust local $M-$estimators of the scale function, $\widehat{\sigma}_{\text{MAD},n}(x\widehat{h}_n)$ and $\widehat{\sigma}_{\text{BT},n}(x\widehat{h}_n)$, introduced in Section 2 when the bandwidth $\widehat{h}_n$ is selected using the procedures described in Section 3.

In all cases, due to the expensive computing time when using cross–validation to select the smoothing parameter, we only performed $N = 500$ replications generating independent samples of size $n = 100$. Two different models for the regression and variance components have been considered. Theses models have been considered previously in Dette and Hetzler (2008), for homoscedastic testing. They will be labeled as model $M1$ and $M2$ and they are defined as follows

- Model $M1$. This model considers as regression function $g(x) = 2\text{sen}(4\pi x)$ while the scale function is $\sigma(x) = \exp(x)$
- Model $M2$. In this model the regression function is linear $g(x) = 1 + x$ and the scale is $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$

In all cases, the design points were taken as $x_i = i/(n + 1)$, $1 \leq i \leq n$. The error's distribution was $G(y) = (1 - \epsilon)\Phi(y) + \epsilon\ H$, with Φ the standard normal distribution and H modeling two types of contamination,

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

b) an asymmetric contamination where $H = \text{N}(10, 0.1)$ is the normal distribution with mean 10 and variance 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 $\epsilon = 0, 0.1, 0.2$ and 0.3. For both, the classical and robust smoothing procedures, we have used weights with a standard gaussian kernel.

The bandwidth were selected using the cross–validation methods introduced in Section 3. More precisely, we will compare the following criteria.

C.1) The classical $K-$fold cross–validation criterion $CV_{\text{LS,KCV}}$ for Rice, and the robust $K-$fold $CV_{\text{ROB,KCV}}$ and split robust $K-$fold $CV_{\text{ROB,SKCV}}$ for the $M-$estimates. We will group this criteria under the denomination of $KCV-$criteria.

C.2) The corresponding to the previous criteria on log scale $CV^{\log}_{\text{LS,KCV}}$ (for Rice), $CV^{\log}_{\text{ROB,KCV}}$, $CV^{\log}_{\text{ROB,SKCV}}$ (for the $M-$estimates), under the denomination of $KCV^{\log}-$criteria.

C.3) The classical leave–one–out criterion $CV_{\text{LS,CV}}$ for Rice, and the robust $CV_{\text{ROB,CV}}$ and split leave–one–out $CV_{\text{ROB,SCV}}$ for the $M-$estimates. This procedures will be grouped under the name of $CV-$criteria.

C.4) The corresponding to the previous criteria on log scale $CV^{\log}_{\text{LS,CV}}$ for Rice, and the robust $CV^{\log}_{\text{ROB,CV}}$ and split leave–one–out $CV^{\log}_{\text{ROB,SCV}}$ for Rice and $M-$estimates, respectively. All of them under the denomination of $CV^{\log}-$criteria.

For the $K-$fold cross–validation methods, we choose $K = 2$ while the score function ψ used for the robust criteria was the Huber function with tuning constant 1.345. The minimization of the cross–validation functions was carried out by inspection over the grid $i/(n/2)$, $0 \leq i \leq n/2$, where n is the sample size.

To asses the behaviour of the selected bandwidth and the performance of each estimator, Tables 1 to 15 report, as summary measures, the mean and the standard deviation, between brackets, of the resulting bandwidths $\widehat{h}_n$ and the estimated integrated square error in logarithmic scale of the estimators, $\widehat{\text{ISEL}}$, defined as

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

where $\widehat{\sigma}_n^{(j)}(x_i, h)$ denotes the scale estimator, classical or robust, obtained at the $j-$th replication with the bandwidth h. Figures 1, 2, 3 and 4 show the density estimators of the ratio between the $\widehat{\text{ISEL}}$ of the robust estimators and that of the non–robust estimators, under the different criteria. The density estimates were evaluated using the normal kernel with bandwidth 0.1 in all cases.

As shown in Tables 1 to 4 and 9 to 12, when data are not contaminated (i.e., the distribution of the errors is F_0), the robust estimators exhibit a loss of efficiency compared with the classical local Rice. That is, the mean of the integrated squared errors of $\widehat{\sigma}_{\text{RICE},n}$ is smaller than the means of $\widehat{\text{ISEL}}(\widehat{\sigma}_{\text{MAD},n})$ and $\widehat{\text{ISEL}}(\widehat{\sigma}_{\text{BT},n})$.

The behavior of $\widehat{\text{ISEL}}(\widehat{\sigma}_{\text{RICE},n})$ show the poor resistance of $\widehat{\sigma}_{\text{RICE},n}$ to the presence of atypical data. As the percentage of contamination increases, Figures 1 and 4 clarify the phenomena observed in Tables 1 to 16 with respect to the better performance of the robust estimators, since the density functions move toward the left of the point 1. Notice that, in most cases, the behavior of $\widehat{\text{ISEL}}$ is much smaller than 1 for $\epsilon > 0$.

Under symmetric contamination, $\widehat{\sigma}_{\text{MAD},n}$ and $\widehat{\sigma}_{\text{BT},n}$ have a similar behavior, although for high contamination proportions (30%) and specially for the model $M1$, the local $M-$estimate $\widehat{\sigma}_{\text{BT},n}$ is slightly better than $\widehat{\sigma}_{\text{MAD},n}$. On the other hand, under asymmetric contaminations $\widehat{\sigma}_{\text{BT},n}$ is clearly more robust than $\widehat{\sigma}_{\text{MAD},n}$.

Note that, in general, the largest mean values of $\widehat{\text{ISEL}}$ appear under the model $M2$ where we also obtain the smallest mean values for the selected bandwidths. This situation indicates that the oscillating behavior of the scale function (presence of several critical points) increases the difficulties in the estimation and less smoothing is needed. But, when the scale curve is a monotone function, as it occurs under model $M1$, the mean values of the selected bandwidths tend to be larger. This implies that more smoothing is needed and the mean values of $\widehat{\text{ISEL}}$ are smaller, compared with those obtained under model $M2$.

It is also important to remark that all the bandwidth selection methods give the same conclusions regarding the performance of the estimators in both models. But we also note that, in general, the smaller values of the integrated square errors occur in SCV or in $SCV^{\log}-$ criteria.

Finally, we recommend the local $M-$estimator based on the Beaton–Tukey score function, $\widehat{\sigma}_{\mathrm{BT},n}(x)$, since it is more stable than $\widehat{\sigma}_{\mathrm{MAD},n}(x)$ even for asymmetric outliers.

kiki

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\mathrm{ISEL}}$ mean (standard deviation)	
		$\widehat{h}_{\mathrm{CV}}$	$\widehat{h}_{\mathrm{SCV}}$	$\widehat{\sigma}_n(\cdot,\widehat{h}_{\mathrm{CV}})$	$\widehat{\sigma}_n(\cdot,\widehat{h}_{\mathrm{SCV}})$
$\epsilon=0$	$\widehat{\sigma}_{\mathrm{RICE},n}$	0.190 (0.096)	0.190 (0.096)	0.032 (0.026)	0.032 (0.026)
	$\widehat{\sigma}_{\mathrm{MAD},n}$	0.181 (0.130)	0.435 (0.078)	0.066 (0.058)	0.055 (0.033)
	$\widehat{\sigma}_{\mathrm{BT},n}$	0.235 (0.131)	0.431 (0.088)	0.065 (0.061)	0.067 (0.072)
$\epsilon=0.10$	$\widehat{\sigma}_{\mathrm{RICE},n}$	0.430 (0.134)	0.430 (0.134)	4.280 (5.764)	4.280 (5.764)
	$\widehat{\sigma}_{\mathrm{MAD},n}$	0.239 (0.146)	0.442 (0.076)	0.098 (0.097)	0.085 (0.059)
	$\widehat{\sigma}_{\mathrm{BT},n}$	0.286 (0.138)	0.442 (0.082)	0.087 (0.076)	0.085 (0.064)
$\epsilon=0.20$	$\widehat{\sigma}_{\mathrm{RICE},n}$	0.435 (0.132)	0.435 (0.132)	7.121 (7.401)	7.121 (7.401)
	$\widehat{\sigma}_{\mathrm{MAD},n}$	0.301 (0.152)	0.453 (0.069)	0.220 (0.158)	0.207 (0.138)
	$\widehat{\sigma}_{\mathrm{BT},n}$	0.354 (0.130)	0.453 (0.072)	0.174 (0.122)	0.172 (0.115)
$\epsilon=0.30$	$\widehat{\sigma}_{\mathrm{RICE},n}$	0.433 (0.134)	0.433 (0.134)	9.362 (7.989)	9.362 (7.989)
	$\widehat{\sigma}_{\mathrm{MAD},n}$	0.360 (0.140)	0.459 (0.062)	0.466 (0.276)	0.458 (0.258)
	$\widehat{\sigma}_{\mathrm{BT},n}$	0.403 (0.120)	0.457 (0.069)	0.351 (0.205)	0.347 (0.204)

Table 1: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\mathrm{ISEL}}$ for the local scale–estimates. Model $M1 : g(x) = 2\mathrm{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. Errors with symmetric contamination. $CV-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\mathrm{ISEL}}$ mean (standard deviation)	
		$\widehat{h}^{\log}_{\mathrm{CV}}$	$\widehat{h}^{\log}_{\mathrm{SCV}}$	$\widehat{\sigma}_n(\cdot,\widehat{h}^{\log}_{\mathrm{CV}})$	$\widehat{\sigma}_n(\cdot,\widehat{h}^{\log}_{\mathrm{SCV}})$
$\epsilon=0$	$\widehat{\sigma}_{\mathrm{RICE},n}$	0.077 (0.071)	0.077 (0.071)	0.056 (0.040)	0.056 (0.040)
	$\widehat{\sigma}_{\mathrm{MAD},n}$	0.225 (0.137)	0.216 (0.131)	0.056 (0.043)	0.061 (0.054)
	$\widehat{\sigma}_{\mathrm{BT},n}$	0.243 (0.138)	0.256 (0.130)	0.067 (0.065)	0.062 (0.056)
$\epsilon=0.10$	$\widehat{\sigma}_{\mathrm{RICE},n}$	0.040 (0.056)	0.040 (0.056)	1.830 (1.829)	1.830 (1.829)
	$\widehat{\sigma}_{\mathrm{MAD},n}$	0.232 (0.142)	0.242 (0.134)	0.100 (0.102)	0.099 (0.086)
	$\widehat{\sigma}_{\mathrm{BT},n}$	0.260 (0.146)	0.258 (0.134)	0.097 (0.084)	0.093 (0.083)
$\epsilon=0.20$	$\widehat{\sigma}_{\mathrm{RICE},n}$	0.035 (0.039)	0.035 (0.039)	3.351 (2.373)	3.351 (2.373)
	$\widehat{\sigma}_{\mathrm{MAD},n}$	0.249 (0.155)	0.252 (0.138)	0.245 (0.180)	0.233 (0.170)
	$\widehat{\sigma}_{\mathrm{BT},n}$	0.254 (0.150)	0.268 (0.141)	0.204 (0.148)	0.199 (0.142)
$\epsilon=0.30$	$\widehat{\sigma}_{\mathrm{RICE},n}$	0.038 (0.053)	0.038 (0.053)	4.723 (2.668)	4.723 (2.668)
	$\widehat{\sigma}_{\mathrm{MAD},n}$	0.228 (0.158)	0.263 (0.150)	0.571 (0.369)	0.531 (0.340)
	$\widehat{\sigma}_{\mathrm{BT},n}$	0.251 (0.159)	0.279 (0.150)	0.414 (0.250)	0.398 (0.243)

Table 2: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\mathrm{ISEL}}$ for the local scale–estimates. Model $M1 : g(x) = 2\mathrm{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. Errors with symmetric contamination. $CV^{\log}-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\text{ISEL}}$ mean (standard deviation)	
		$\widehat{h}_{\text{KCV}}$	$\widehat{h}_{\text{SKCV}}$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{KCV}})$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{SKCV}})$
$\epsilon = 0$	$\widehat{\sigma}_{\text{RICE},n}$	0.283 (0.132)	0.283 (0.132)	0.036 (0.024)	0.036 (0.024)
	$\widehat{\sigma}_{\text{MAD},n}$	0.087 (0.124)	0.384 (0.124)	0.133 (0.078)	0.052 (0.031)
	$\widehat{\sigma}_{\text{BT},n}$	0.144 (0.169)	0.411 (0.112)	0.164 (0.122)	0.063 (0.043)
$\epsilon = 0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.422 (0.140)	0.422 (0.1405)	4.269 (5.731)	4.269 (5.731)
	$\widehat{\sigma}_{\text{MAD},n}$	0.110 (0.143)	0.397 (0.119)	0.205 (0.165)	0.085 (0.060)
	$\widehat{\sigma}_{\text{BT},n}$	0.169 (0.180)	0.432 (0.097)	0.199 (0.151)	0.085 (0.065)
$\epsilon = 0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.440 (0.190)	0.440 (0.190)	7.104 (7.326)	7.104 (7.326)
	$\widehat{\sigma}_{\text{MAD},n}$	0.133 (0.160)	0.414 (0.109)	0.408 (0.286)	0.210 (0.145)
	$\widehat{\sigma}_{\text{BT},n}$	0.182 (0.185)	0.424 (0.108)	0.317 (0.211)	0.173 (0.116)
$\epsilon = 0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.439 (0.121)	0.439 (0.121)	9.289 (7.872)	9.289 (7.872)
	$\widehat{\sigma}_{\text{MAD},n}$	0.136 (0.159)	0.410 (0.119)	0.774 (0.462)	0.463 (0.265)
	$\widehat{\sigma}_{\text{BT},n}$	0.190 (0.190)	0.428 (0.112)	0.544 (0.3201)	0.351 (0.205)

Table 3: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\text{ISEL}}$ for the local scale–estimates. Model $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. Errors with symmetric contamination. $KCV-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\text{ISEL}}$ mean (standard deviation)	
		$\widehat{h}^{\log}_{\text{KCV}}$	$\widehat{h}^{\log}_{\text{SKCV}}$	$\widehat{\sigma}_n(\cdot, \widehat{h}^{\log}_{\text{KCV}})$	$\widehat{\sigma}_n(\cdot, \widehat{h}^{\log}_{\text{SKCV}})$
$\epsilon = 0$	$\widehat{\sigma}_{\text{RICE},n}$	0.081 (0.083)	0.081 (0.083)	0.054 (0.038)	0.054 (0.038)
	$\widehat{\sigma}_{\text{MAD},n}$	0.140 (0.159)	0.269 (0.150)	0.118 (0.082)	0.055 (0.040)
	$\widehat{\sigma}_{\text{BT},n}$	0.160 (0.173)	0.303 (0.148)	0.158 (0.123)	0.062 (0.048)
$\epsilon = 0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.038 (0.049)	0.038 (0.049)	1.831 (1.832)	1.831 (1.832)
	$\widehat{\sigma}_{\text{MAD},n}$	0.143 (0.155)	0.273 (0.147)	0.180 (0.162)	0.092 (0.071)
	$\widehat{\sigma}_{\text{BT},n}$	0.175 (0.179)	0.311 (0.146)	0.196 (0.151)	0.091 (0.075)
$\epsilon = 0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.037 (0.053)	0.037 (0.053)	3.350 (2.370)	3.350 (2.370)
	$\widehat{\sigma}_{\text{MAD},n}$	0.143 (0.157)	0.290 (0.146)	0.388 (0.284)	0.218 (0.148)
	$\widehat{\sigma}_{\text{BT},n}$	0.174 (0.181)	0.310 (0.146)	0.320 (0.214)	0.184 (0.121)
$\epsilon = 0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.033 (0.033)	0.033 (0.033)	4.718 (2.671)	4.718 (2.671)
	$\widehat{\sigma}_{\text{MAD},n}$	0.132 (0.145)	0.293 (0.146)	0.754 (0.462)	0.489 (0.290)
	$\widehat{\sigma}_{\text{BT},n}$	0.165 (0.170)	0.308 (0.143)	0.552 (0.320)	0.367 (0.212)

Table 4: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\text{ISEL}}$ for the local scale–estimates. Model $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. Errors with symmetric contamination. $KCV^{\log}-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\text{ISEL}}$ mean (standard deviation)	
		$\widehat{h}_{\text{CV}}$	$\widehat{h}_{\text{SCV}}$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{CV}})$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{SCV}})$
$\epsilon = 0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.267 (0.169)	0.267 (0.169)	1.313 (0.352)	1.313 (0.352)
	$\widehat{\sigma}_{\text{MAD},n}$	0.292 (0.146)	0.446 (0.074)	0.121 (0.116)	0.115 (0.084)
	$\widehat{\sigma}_{\text{BT},n}$	0.338 (0.131)	0.443 (0.079)	0.099 (0.086)	0.101 (0.081)
$\epsilon = 0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.206 (0.131)	0.206 (0.131)	2.029 (0.270)	2.029 (0.270)
	$\widehat{\sigma}_{\text{MAD},n}$	0.407 (0.118)	0.459 (0.063)	0.359 (0.342)	0.354 (0.329)
	$\widehat{\sigma}_{\text{BT},n}$	0.441 (0.091)	0.458 (0.070)	0.194 (0.137)	0.196 (0.142)
$\epsilon = 0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.164 (0.099)	0.164 (0.099)	2.407 (0.227)	2.407 (0.227)
	$\widehat{\sigma}_{\text{MAD},n}$	0.401 (0.135)	0.459 (0.064)	0.996 (0.861)	0.353 (0.329)
	$\widehat{\sigma}_{\text{BT},n}$	0.472 (0.069)	0.458 (0.070)	0.310 (0.243)	0.196 (0.142)

Table 5: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\text{ISEL}}$ for the local scale–estimates. Model $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. Errors with asymmetric contamination. $CV-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\text{ISEL}}$ mean (standard deviation)	
		$\widehat{h}_{\text{CV}}^{\log}$	$\widehat{h}_{\text{SCV}}^{\log}$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{CV}}^{\log})$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{SCV}}^{\log})$
$\epsilon = 0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.058 (0.079)	0.058 (0.079)	1.152 (0.321)	1.152 (0.321)
	$\widehat{\sigma}_{\text{MAD},n}$	0.213 (0.148)	0.246 (0.138)	0.183 (0.221)	0.148 (0.156)
	$\widehat{\sigma}_{\text{BT},n}$	0.248 (0.151)	0.266 (0.141)	0.136 (0.143)	0.123 (0.124)
$\epsilon = 0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.059 (0.059)	0.059 (0.059)	1.892 (0.278)	1.892 (0.278)
	$\widehat{\sigma}_{\text{MAD},n}$	0.219 (0.184)	0.270 (0.151)	0.657 (0.523)	0.490 (0.482)
	$\widehat{\sigma}_{\text{BT},n}$	0.171 (0.179)	0.273 (0.161)	0.400 (0.259)	0.262 (0.205)
$\epsilon = 0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.058 (0.062)	0.058 (0.062)	2.309 (0.229)	2.309 (0.229)
	$\widehat{\sigma}_{\text{MAD},n}$	0.274 (0.177)	0.316 (0.162)	1.270 (0.853)	1.130 (0.884)
	$\widehat{\sigma}_{\text{BT},n}$	0.150 (0.163)	0.271 (0.161)	0.611 (0.368)	0.340 (0.298)

Table 6: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\text{ISEL}}$ for the local scale–estimates. Model $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. Errors with asymmetric contamination $CV^{\log}-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\text{ISEL}}$ mean (standard deviation)	
		$\widehat{h}_{\text{KCV}}$	$\widehat{h}_{\text{SKCV}}$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{KCV}})$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{SKCV}})$
$\epsilon = 0.10$	$\widehat{\sigma}_{\text{MAD},n}$	0.118 (0.148)	0.403 (0.113)	0.320 (0.287)	0.116 (0.096)
	$\widehat{\sigma}_{\text{BT},n}$	0.161 (0.175)	0.424 (0.108)	0.242 (0.178)	0.102 (0.085)
$\epsilon = 0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.301 (0.141)	0.301 (0.141)	2.105 (0.276)	2.105 (0.276)
	$\widehat{\sigma}_{\text{MAD},n}$	0.131 (0.158)	0.396 (0.130)	0.811 (0.597)	0.387 (0.367)
	$\widehat{\sigma}_{\text{BT},n}$	0.179 (0.184)	0.432 (0.109)	0.417 (0.300)	0.201 (0.149)
$\epsilon = 0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.279 (0.121)	0.279 (0.121)	2.494 (0.235)	2.494 (0.235)
	$\widehat{\sigma}_{\text{MAD},n}$	0.134 (0.191)	0.344 (0.172)	1.342 (0.905)	1.103 (0.879)
	$\widehat{\sigma}_{\text{BT},n}$	0.238 (0.224)	0.398 (0.150)	0.631 (0.452)	0.344 (0.281)

Table 7: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\text{ISEL}}$ for the local scale–estimates. Model $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. Errors with asymmetric contamination. $KCV-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\text{ISEL}}$ mean (standard deviation)	
		$\widehat{h}_{\text{KCV}}^{\log}$	$\widehat{h}_{\text{SKCV}}^{\log}$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{KCV}}^{\log})$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{SKCV}}^{\log})$
$\epsilon = 0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.043 (0.059)	0.043 (0.059)	1.140 (0.317)	1.140 (0.317)
	$\widehat{\sigma}_{\text{MAD},n}$	0.142 (0.162)	0.269 (0.143)	0.308 (0.286)	0.129 (0.121)
	$\widehat{\sigma}_{\text{BT},n}$	0.164 (0.179)	0.296 (0.144)	0.244 (0.179)	0.110 (0.097)
$\epsilon = 0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.051 (0.056)	0.051 (0.056)	1.887 (0.283)	1.887 (0.283)
	$\widehat{\sigma}_{\text{MAD},n}$	0.138 (0.162)	0.305 (0.149)	0.790 (0.592)	0.419 (0.381)
	$\widehat{\sigma}_{\text{BT},n}$	0.152 (0.174)	0.303 (0.151)	0.436 (0.297)	0.217 (0.154)
$\epsilon = 0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.054 (0.061)	0.293 (0.164)	2.308 (0.234)	2.308 (0.234)
	$\widehat{\sigma}_{\text{MAD},n}$	0.154 (0.176)	0.293 (0.164)	1.370 (0.891)	1.152 (0.885)
	$\widehat{\sigma}_{\text{BT},n}$	0.130 (0.158)	0.322 (0.154)	0.680 (0.447)	0.341 (0.257)

Table 8: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\text{ISEL}}$ for the local scale–estimates. Model $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$. Errors with asymmetric contamination. $KCV^{\log}-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\text{ISEL}}$ mean (standard deviation)	
		$\widehat{h}_{\text{CV}}$	$\widehat{h}_{\text{SCV}}$	$\widehat{\sigma}_n(\cdot,\widehat{h}_{\text{CV}})$	$\widehat{\sigma}_n(\cdot,\widehat{h}_{\text{SCV}})$
$\epsilon=0$	$\widehat{\sigma}_{\text{RICE},n}$	0.069 (0.044)	0.069 (0.044)	0.097 (0.056)	0.097 (0.056)
	$\widehat{\sigma}_{\text{MAD},n}$	0.174 (0.086)	0.339 (0.087)	0.223 (0.093)	0.330 (0.062)
	$\widehat{\sigma}_{\text{BT},n}$	0.195 (0.076)	0.325 (0.084)	0.267 (0.107)	0.349 (0.081)
$\epsilon=0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.421 (0.147)	0.421 (0.147)	4.877 (6.087)	4.877 (6.087)
	$\widehat{\sigma}_{\text{MAD},n}$	0.219 (0.098)	0.354 (0.082)	0.271 (0.085)	0.345 (0.064)
	$\widehat{\sigma}_{\text{BT},n}$	0.244 (0.087)	0.336 (0.091)	0.290 (0.074)	0.330 (0.057)
$\epsilon=0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.434 (0.138)	7.781 (7.43)	0.434 (0.138)	7.781 (7.43)
	$\widehat{\sigma}_{\text{MAD},n}$	0.264 (0.105)	0.384 (0.089)	0.422 (0.166)	0.476 (0.161)
	$\widehat{\sigma}_{\text{BT},n}$	0.293 (0.098)	0.349 (0.089)	0.361 (0.107)	0.380 (0.098)
$\epsilon=0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.439 (0.129)	0.439 (0.129)	10.063 (7.988)	10.063 (7.988)
	$\widehat{\sigma}_{\text{MAD},n}$	0.297 (0.109)	0.407 (0.081)	0.722 (0.305)	0.752 (0.294)
	$\widehat{\sigma}_{\text{BT},n}$	0.327 (0.097)	0.369 (0.090)	0.516 (0.186)	0.525 (0.183)

Table 9: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\text{ISEL}}$ for the local scale–estimates. Model $M2:\ g(x)=1+x$, $\sigma(x)=1+[1+\text{sen}(10x)]^2$. Errors with symmetric contamination. $CV-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\text{ISEL}}$ mean (standard deviation)	
		$\widehat{h}^{\log}_{\text{CV}}$	$\widehat{h}^{\log}_{\text{SCV}}$	$\widehat{\sigma}_n(\cdot,\widehat{h}^{\log}_{\text{CV}})$	$\widehat{\sigma}_n(\cdot,\widehat{h}^{\log}_{\text{SCV}})$
$\epsilon=0$	$\widehat{\sigma}_{\text{RICE},n}$	0.031 (0.004)	0.031 (0.004)	0.076 (0.033)	0.076 (0.033)
	$\widehat{\sigma}_{\text{MAD},n}$	0.087 (0.063)	0.080 (0.045)	0.141 (0.076)	0.140 (0.074)
	$\widehat{\sigma}_{\text{BT},n}$	0.099 (0.071)	0.098 (0.057)	0.171 (0.090)	0.168 (0.089)
$\epsilon=0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.031 (0.014)	0.031 (0.014)	1.869 (1.848)	1.869 (1.848)
	$\widehat{\sigma}_{\text{MAD},n}$	0.111 (0.087)	0.098 (0.057)	0.207 (0.108)	0.198 (0.109)
	$\widehat{\sigma}_{\text{BT},n}$	0.125 (0.094)	0.117 (0.084)	0.206 (0.094)	0.199 (0.094)
$\epsilon=0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.031 (0.021)	0.031 (0.021)	3.417 (2.397)	3.417 (2.397)
	$\widehat{\sigma}_{\text{MAD},n}$	0.155 (0.132)	0.141 (0.117)	0.410 (0.239)	0.384 (0.212)
	$\widehat{\sigma}_{\text{BT},n}$	0.168 (0.128)	0.145 (0.114)	0.322 (0.151)	0.310 (0.150)
$\epsilon=0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.033 (0.034)	0.033 (0.034)	4.811 (2.705)	4.811 (2.705)
	$\widehat{\sigma}_{\text{MAD},n}$	0.174 (0.144)	0.168 (0.134)	0.753 (0.367)	0.730 (0.359)
	$\widehat{\sigma}_{\text{BT},n}$	0.174 (0.138)	0.166 (0.132)	0.529 (0.239)	0.517 (0.232)

Table 10: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\text{ISEL}}$ for the local scale–estimates. Model $M2:\ g(x)=1+x$, $\sigma(x)=1+[1+\text{sen}(10x)]^2$. Errors with symmetric contamination. $CV^{\log}-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\text{ISEL}}$ mean (standard deviation)	
		$\widehat{h}_{\text{KCV}}$	$\widehat{h}_{\text{SKCV}}$	$\widehat{\sigma}_n(\cdot,\widehat{h}_{\text{KCV}})$	$\widehat{\sigma}_n(\cdot,\widehat{h}_{\text{SKCV}})$
$\epsilon=0$	$\widehat{\sigma}_{\text{RICE},n}$	0.218 (0.172)	0.218 (0.172)	0.239 (0.139)	0.239 (0.139)
	$\widehat{\sigma}_{\text{MAD},n}$	0.109 (0.126)	0.328 (0.109)	0.208 (0.095)	0.318 (0.075)
	$\widehat{\sigma}_{\text{BT},n}$	0.133 (0.136)	0.340 (0.108)	0.267 (0.105)	0.347 (0.086)
$\epsilon=0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.435 (0.128)	0.435 (0.128)	4.872 (6.018)	4.872 (6.018)
	$\widehat{\sigma}_{\text{MAD},n}$	0.127 (0.142)	0.344 (0.115)	0.285 (0.151)	0.333 (0.081)
	$\widehat{\sigma}_{\text{BT},n}$	0.158 (0.153)	0.348 (0.111)	0.301 (0.108)	0.328 (0.065)
$\epsilon=0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.442 (0.122)	0.442 (0.122)	7.779 (7.354)	7.779 (7.354)
	$\widehat{\sigma}_{\text{MAD},n}$	0.138 (0.148)	0.355 (0.120)	0.500 (0.263)	0.468 (0.169)
	$\widehat{\sigma}_{\text{BT},n}$	0.162 (0.153)	0.353 (0.119)	0.405 (0.168)	0.377 (0.108)
$\epsilon=0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.435 (0.129)	0.435 (0.129)	9.972 (7.909)	9.972 (7.909)
	$\widehat{\sigma}_{\text{MAD},n}$	0.136 (0.148)	0.372 (0.121)	0.891 (0.434)	0.749 (0.304)
	$\widehat{\sigma}_{\text{BT},n}$	0.165 (0.157)	0.372 (0.123)	0.616 (0.279)	0.528 (0.195)

Table 11: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\text{ISEL}}$ for the local scale–estimates. Model $M2:\ g(x)=1+x,\ \sigma(x)=1+[1+\text{sen}(10x)]^2$. Errors with symmetric contamination. $KCV-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\text{ISEL}}$ mean (standard deviation)	
		$\widehat{h}^{\log}_{\text{KCV}}$	$\widehat{h}^{\log}_{\text{SKCV}}$	$\widehat{\sigma}_n(\cdot,\widehat{h}^{\log}_{\text{KCV}})$	$\widehat{\sigma}_n(\cdot,\widehat{h}^{\log}_{\text{SKCV}})$
$\epsilon=0$	$\widehat{\sigma}_{\text{RICE},n}$	0.032 (0.007)	0.032 (0.007)	0.076 (0.033)	0.076 (0.033)
	$\widehat{\sigma}_{\text{MAD},n}$	0.010 (0.117)	0.080 (0.045)	0.193 (0.093)	0.139 (0.074)
	$\widehat{\sigma}_{\text{BT},n}$	0.117 (0.130)	0.098 (0.057)	0.250 (0.111)	0.168 (0.088)
$\epsilon=0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.031 (0.005)	0.031 (0.005)	1.872 (1.851)	1.872 (1.851)
	$\widehat{\sigma}_{\text{MAD},n}$	0.123 (0.140)	0.097 (0.074)	0.273 (0.155)	0.199 (0.109)
	$\widehat{\sigma}_{\text{BT},n}$	0.150 (0.152)	0.117 (0.084)	0.296 (0.112)	0.199 (0.094)
$\epsilon=0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.031 (0.005)	0.031 (0.005)	3.416 (2.396)	3.416 (2.396)
	$\widehat{\sigma}_{\text{MAD},n}$	0.133 (0.152)	0.141 (0.117)	0.490 (0.265)	0.384 (0.213)
	$\widehat{\sigma}_{\text{BT},n}$	0.158 (0.163)	0.145 (0.114)	0.404 (0.172)	0.310 (0.150)
$\epsilon=0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.030 (0.004)	0.030 (0.004)	4.803 (2.670)	4.803 (2.670)
	$\widehat{\sigma}_{\text{MAD},n}$	0.127 (0.144)	0.168 (0.134)	0.867 (0.434)	0.730 (0.359)
	$\widehat{\sigma}_{\text{BT},n}$	0.148 (0.153)	0.166 (0.132)	0.614 (0.282)	0.517 (0.232)

Table 12: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\text{ISEL}}$ for the local scale–estimates. Model $M2: g(x)=1+x,\ \sigma(x)=1+[1+\text{sen}(10x)]^2$. Errors with symmetric contamination. $KCV^{\log}-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\text{ISEL}}$ mean (standard deviation)	
		$\widehat{h}_{\text{CV}}$	$\widehat{h}_{\text{SCV}}$	$\widehat{\sigma}_n(\cdot,\widehat{h}_{\text{CV}})$	$\widehat{\sigma}_n(\cdot,\widehat{h}_{\text{SCV}})$
$\epsilon=0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.232 (0.207)	0.232 (0.207)	1.711 (0.513)	1.711 (0.513)
	$\widehat{\sigma}_{\text{MAD},n}$	0.261 (0.103)	0.376 (0.089)	0.338 (0.128)	0.390 (0.111)
	$\widehat{\sigma}_{\text{BT},n}$	0.281 (0.091)	0.344 (0.090)	0.317 (0.079)	0.341 (0.066)
$\epsilon=0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.096 (0.116)	0.096 (0.116)	2.303 (0.367)	2.303 (0.367)
	$\widehat{\sigma}_{\text{MAD},n}$	0.355 (0.113)	0.423 (0.079)	0.705 (0.373)	0.724 (0.365)
	$\widehat{\sigma}_{\text{BT},n}$	0.370 (0.093)	0.376 (0.092)	0.437 (0.142)	0.440 (0.140)
$\epsilon=0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.059 (0.042)	0.059 (0.042)	2.636 (0.290)	2.636 (0.290)
	$\widehat{\sigma}_{\text{MAD},n}$	0.364 (0.139)	0.449 (0.070)	1.331 (0.621)	1.337 (0.593)
	$\widehat{\sigma}_{\text{BT},n}$	0.439 (0.070)	0.404 (0.094)	0.617 (0.244)	0.612 (0.248)

Table 13: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\text{ISEL}}$ for the local scale–estimates. Model $M2: g(x)=1+x,\ \sigma(x)=1+[1+\text{sen}(10x)]^2$. Errors with asymmetric contamination. $CV-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\text{ISEL}}$ mean (standard deviation)	
		$\widehat{h}^{\log}_{\text{CV}}$	$\widehat{h}^{\log}_{\text{SCV}}$	$\widehat{\sigma}_n(\cdot,\widehat{h}^{\log}_{\text{CV}})$	$\widehat{\sigma}_n(\cdot,\widehat{h}^{\log}_{\text{SCV}})$
$\epsilon=0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.034 (0.031)	0.034 (0.031)	1.190 (0.325)	1.190 (0.325)
	$\widehat{\sigma}_{\text{MAD},n}$	0.178 (0.150)	0.126 (0.108)	0.320 (0.196)	0.280 (0.186)
	$\widehat{\sigma}_{\text{BT},n}$	0.181 (0.147)	0.129 (0.101)	0.264 (0.138)	0.237 (0.123)
$\epsilon=0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.032 (0.014)	0.032 (0.014)	1.955 (0.285)	1.955 (0.285)
	$\widehat{\sigma}_{\text{MAD},n}$	0.229 (0.160)	0.213 (0.163)	0.795 (0.429)	0.756 (0.429)
	$\widehat{\sigma}_{\text{BT},n}$	0.270 (0.169)	0.189 (0.152)	0.479 (0.202)	0.420 (0.204)
$\epsilon=0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.031 (0.003)	0.031 (0.003)	2.384 (0.233)	2.384 (0.233)
	$\widehat{\sigma}_{\text{MAD},n}$	0.183 (0.146)	0.222 (0.163)	1.468 (0.645)	1.413 (0.615)
	$\widehat{\sigma}_{\text{BT},n}$	0.200 (0.162)	0.178 (0.143)	0.732 (0.347)	0.625 (0.294)

Table 14: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\text{ISEL}}$ for the local scale–estimates. Model $M2: g(x)=1+x,\ \sigma(x)=1+[1+\text{sen}(10x)]^2$. Errors with asymmetric contamination. $CV^{\log}-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\text{ISEL}}$ mean (standard deviation)	
		$\widehat{h}_{\text{KCV}}$	$\widehat{h}_{\text{SKCV}}$	$\widehat{\sigma}_n(\cdot,\widehat{h}_{\text{KCV}})$	$\widehat{\sigma}_n(\cdot,\widehat{h}_{\text{SKCV}})$
$\epsilon = 0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.322 (0.181)	0.322 (0.181)	1.979 (0.543)	1.979 (0.543)
	$\widehat{\sigma}_{\text{MAD},n}$	0.116 (0.134)	0.351 (0.114)	0.397 (0.209)	0.379 (0.121)
	$\widehat{\sigma}_{\text{BT},n}$	0.152 (0.155)	0.356 (0.116)	0.337 (0.128)	0.339 (0.076)
$\epsilon = 0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.245 (0.176)	0.245 (0.176)	2.750 (0.488)	2.750 (0.488)
	$\widehat{\sigma}_{\text{MAD},n}$	0.135 (0.157)	0.373 (0.130)	0.922 (0.476)	0.718 (0.368)
	$\widehat{\sigma}_{\text{BT},n}$	0.166 (0.168)	0.372 (0.121)	0.523 (0.247)	0.436 (0.147)
$\epsilon = 0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.207 (0.170)	0.207 (0.170)	3.118 (0.444)	3.118 (0.444)
	$\widehat{\sigma}_{\text{MAD},n}$	0.127 (0.160)	0.392 (0.120)	1.584 (0.690)	1.349 (0.601)
	$\widehat{\sigma}_{\text{BT},n}$	0.185 (0.182)	0.403 (0.118)	0.787 (0.378)	0.610 (0.250)

Table 15: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\text{ISEL}}$ for the local scale–estimates. Model $M2 : g(x) = 1 + x$, $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. Errors with asymmetric contamination. $KCV-$criteria.

	Local $M-$estimates	Selected Bandwidth mean (standard deviation)		$\widehat{\text{ISEL}}$ mean (standard deviation)	
		$\widehat{h}^{\log}_{\text{KCV}}$	$\widehat{h}^{\log}_{\text{SKCV}}$	$\widehat{\sigma}_n(\cdot,\widehat{h}^{\log}_{\text{KCV}})$	$\widehat{\sigma}_n(\cdot,\widehat{h}^{\log}_{\text{SKCV}})$
$\epsilon = 0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.031 (0.005)	0.031 (0.005)	1.187 (0.322)	1.187 (0.322)
	$\widehat{\sigma}_{\text{MAD},n}$	0.122 (0.143)	0.202 (0.141)	0.390 (0.206)	0.298 (0.159)
	$\widehat{\sigma}_{\text{BT},n}$	0.149 (0.161)	0.224 (0.143)	0.336 (0.133)	0.272 (0.106)
$\epsilon = 0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.031 (0.006)	0.031 (0.006)	1.954 (0.294)	1.954 (0.294)
	$\widehat{\sigma}_{\text{MAD},n}$	0.127 (0.147)	0.241 (0.152)	0.914 (0.471)	0.710 (0.393)
	$\widehat{\sigma}_{\text{BT},n}$	0.156 (0.170)	0.253 (0.147)	0.527 (0.249)	0.403 (0.177)
$\epsilon = 0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.031 (0.004)	0.031 (0.004)	2.387 (0.237)	2.387 (0.237)
	$\widehat{\sigma}_{\text{MAD},n}$	0.124 (0.144)	0.251 (0.152)	1.557 (0.664)	1.378 (0.613)
	$\widehat{\sigma}_{\text{BT},n}$	0.148 (0.162)	0.265 (0.151)	0.791 (0.379)	0.594 (0.274)

Table 16: Mean and standard deviation (between brackets) of the selected bandwidths and of the $\widehat{\text{ISEL}}$ for the local scale–estimates. Model $M2 : g(x) = 1 + x$, $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$. Errors with asymmetric contamination. $KCV^{\log}-$criteria.

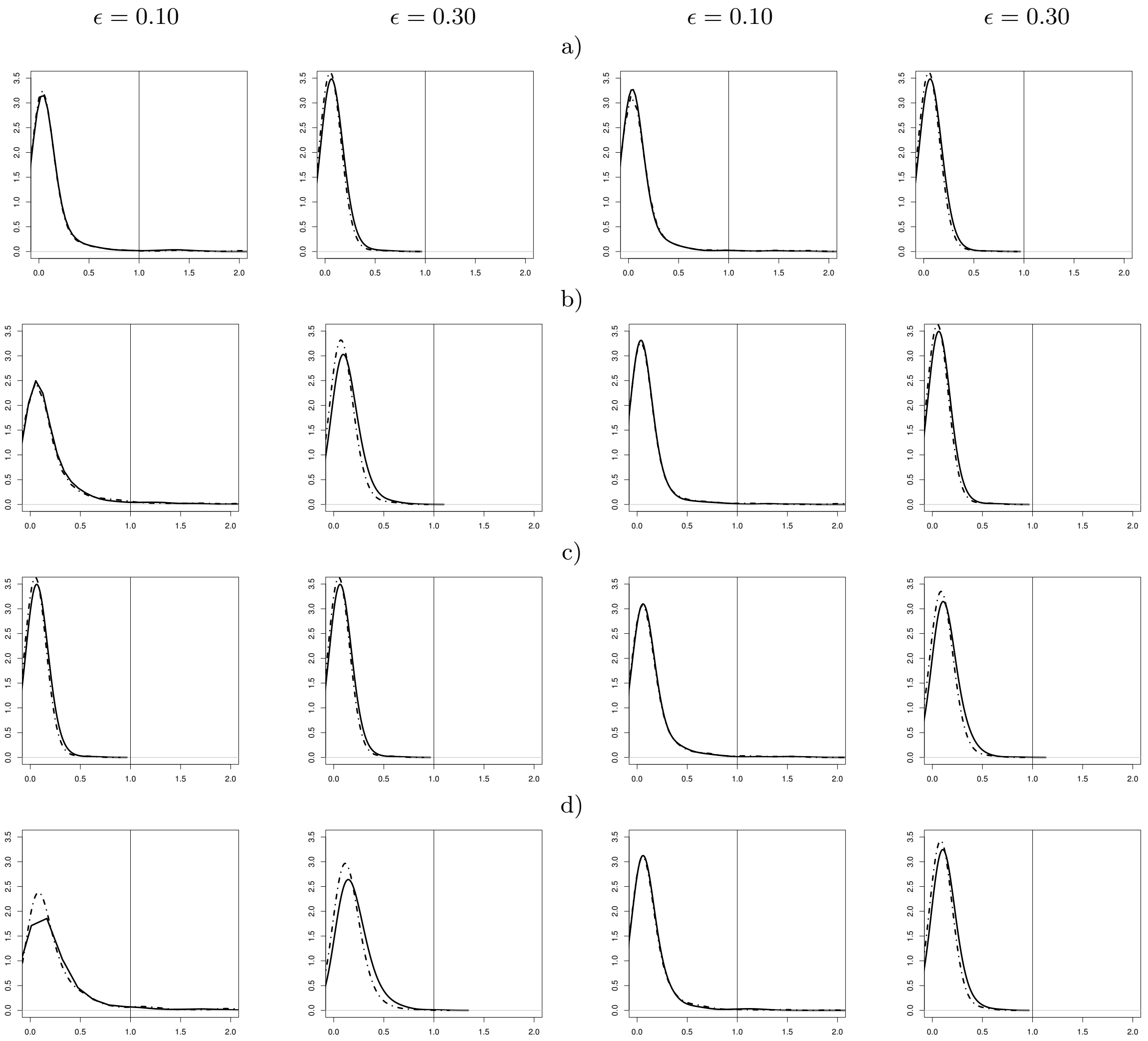

Figure 1: Density estimator of the ratio between the $\widehat{\text{ISEL}}$ of the robust estimators and that of the non–robust estimator $\widehat{\sigma}_{\text{RICE},n}(x)$. The solid line corresponds to $\widehat{\sigma}_{\text{MAD},n}(x)$ and the broken (- -) lines correspond to $\widehat{\sigma}_{\text{BT},n}(x)$. Model $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$ with symmetric contamination. a) CV(left) and SCV (right). b) KCV(left) and $SKCV$(right). c) $CV^{\log}$ (left) and $SCV^{\log}$ (right). d) $KCV^{\log}$ (left) and $SKCV^{\log}$ (right)

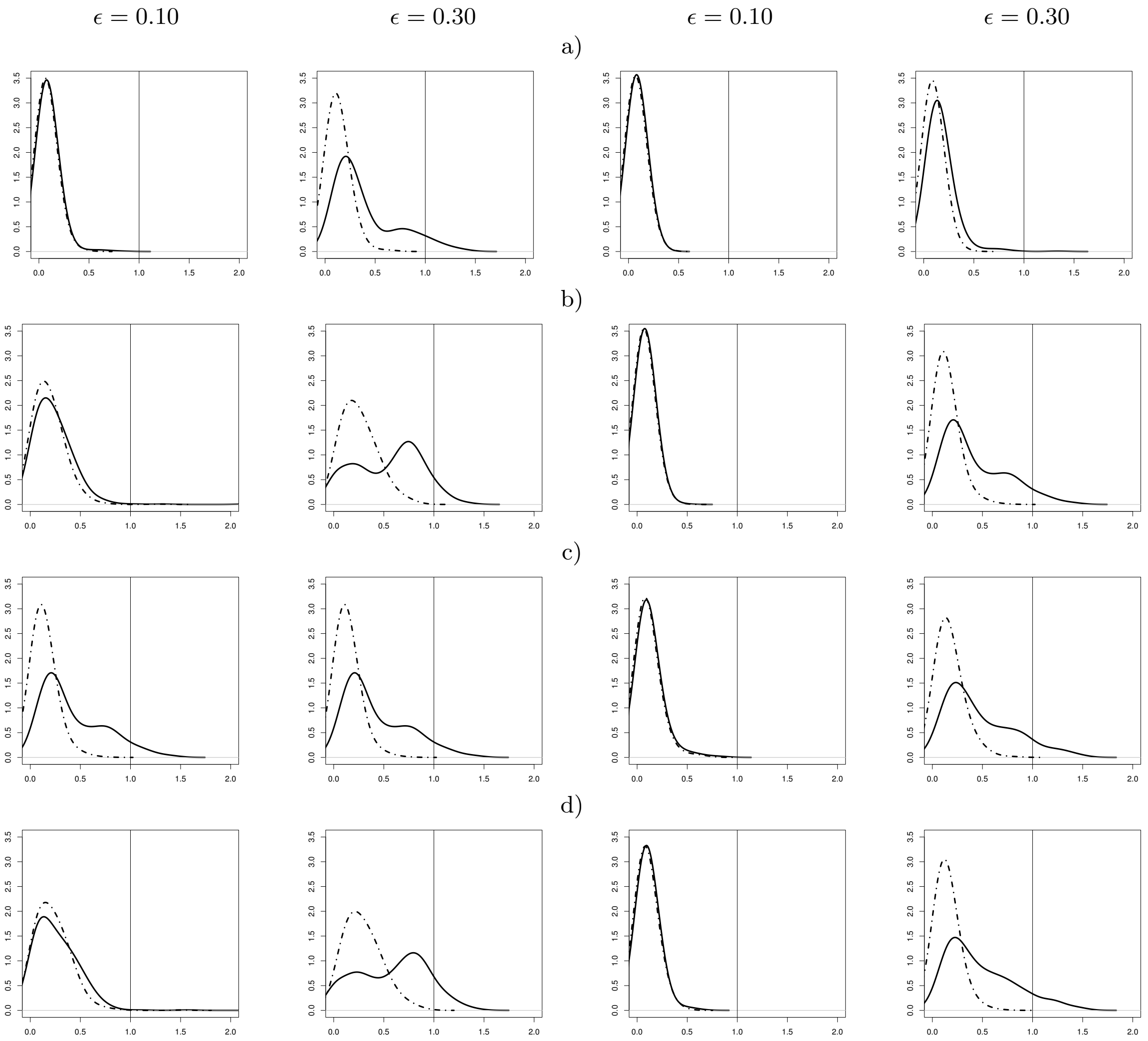

Figure 2: Density estimator of the ratio between the $\widehat{\text{ISEL}}$ of the robust estimators and that of the non–robust estimator $\widehat{\sigma}_{\text{RICE},n}(x)$. The solid line corresponds to $\widehat{\sigma}_{\text{MAD},n}(x)$ and the broken (- -) lines correspond to $\widehat{\sigma}_{\text{BT},n}(x)$. Model $M1 : g(x) = 2\text{sen}(4\pi x)$, $\sigma(x) = \exp(x)$ with asymmetric contamination. a) CV(left) and SCV (right). b) KCV(left) and $SKCV$(right). c) $CV^{\log}$ (left) and $SCV^{\log}$ (right). d) $KCV^{\log}$ (left) and $SKCV^{\log}$ (right)

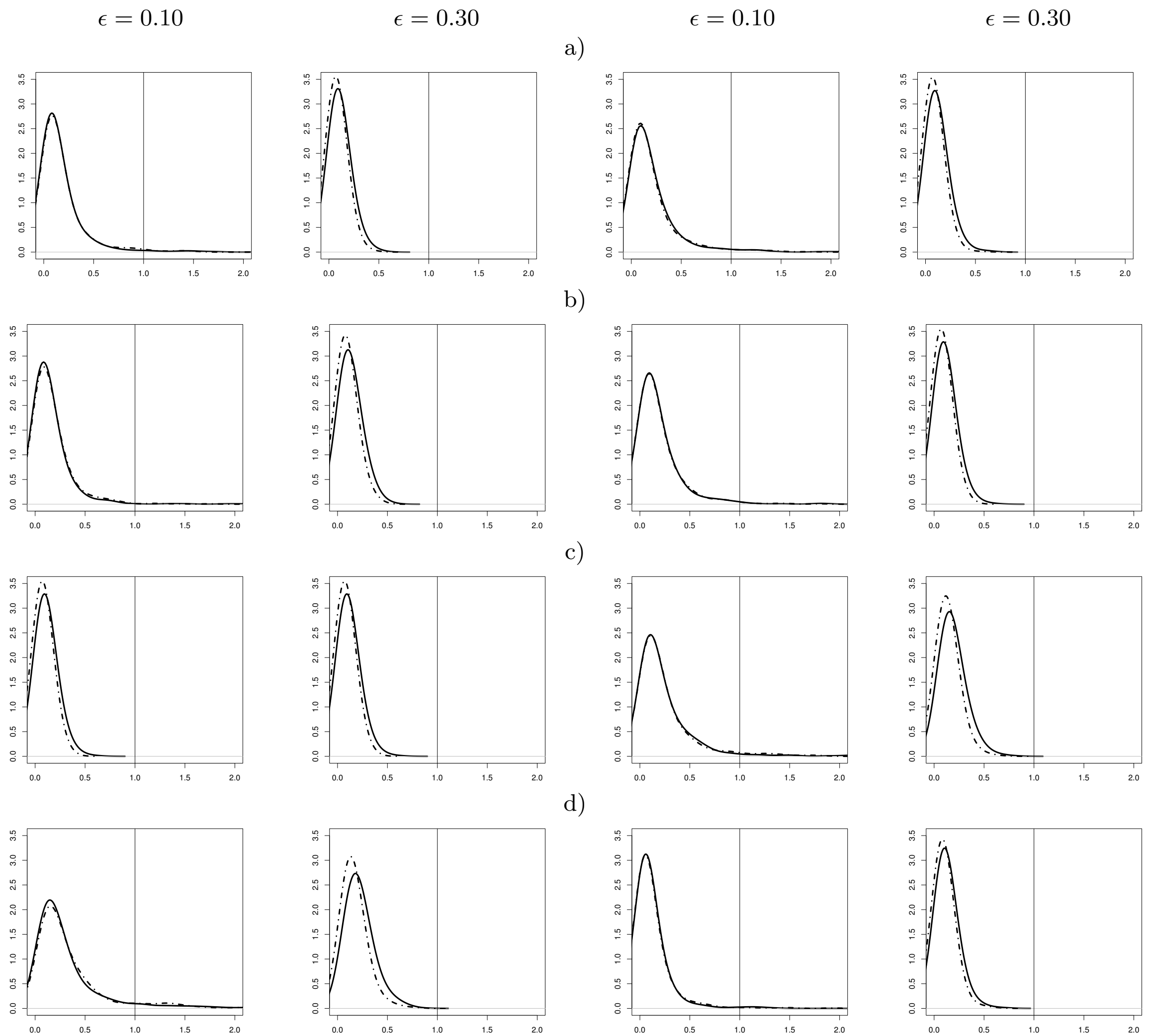

Figure 3: Density estimator of the ratio between the $\widehat{\text{ISEL}}$ of the robust estimators and that of the non–robust estimator $\widehat{\sigma}_{\text{RICE},n}(x)$. The solid line corresponds to $\widehat{\sigma}_{\text{MAD},n}(x)$ and the broken (- -) lines correspond to $\widehat{\sigma}_{\text{BT},n}(x)$. Model $M2$: $g(x) = 1 + x$, $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$ with symmetric contamination. a) CV(left) and SCV (right). b) KCV(left) and $SKCV$(right). c) $CV^{\log}$ (left) and $SCV^{\log}$ (right). d) $KCV^{\log}$ (left) and $SKCV^{\log}$ (right)

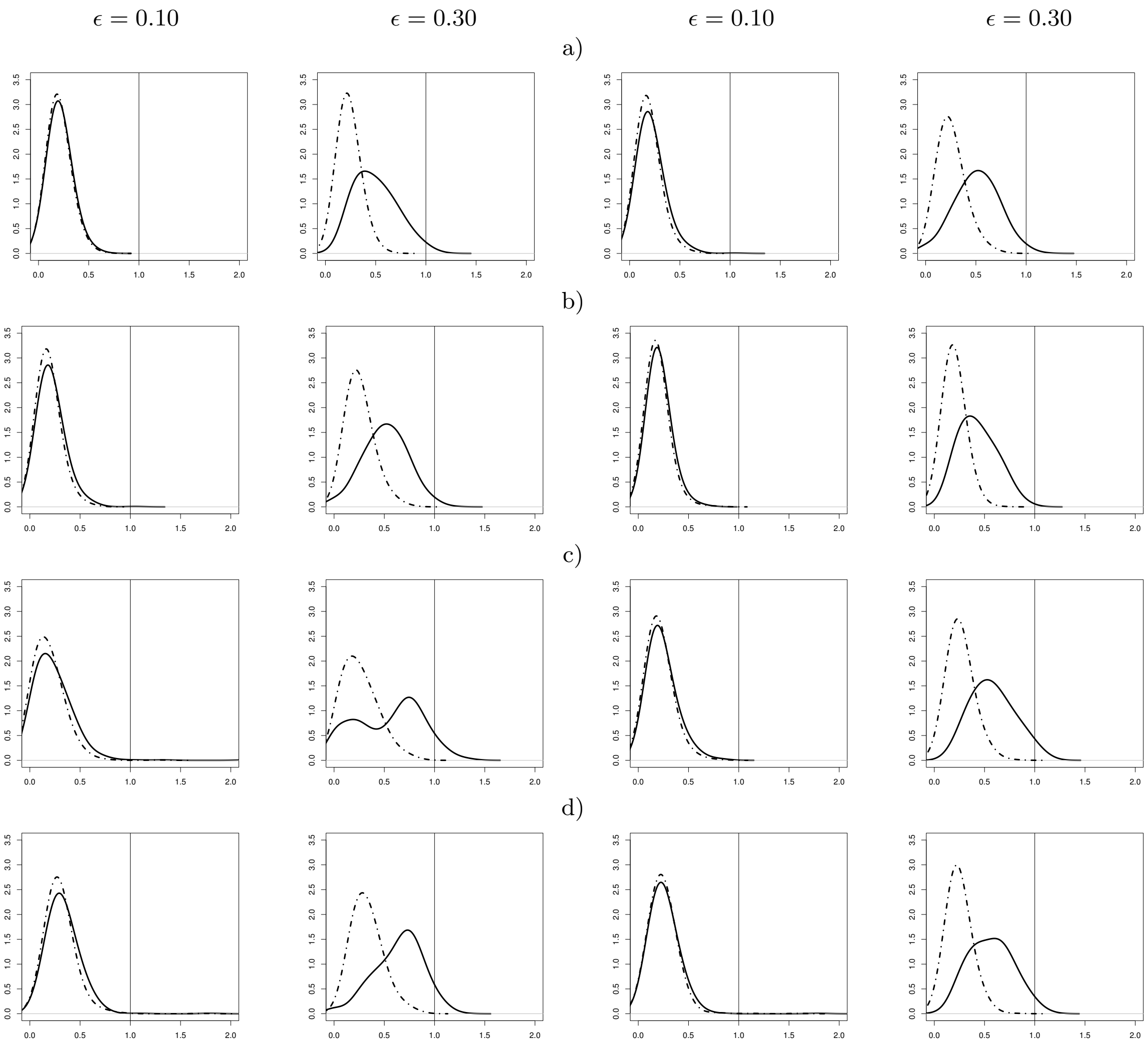

Figure 4: Density estimator of the ratio between the $\widehat{\text{ISEL}}$ of the robust estimators and that of the non–robust estimator $\widehat{\sigma}_{\text{RICE},n}(x)$. The solid line corresponds to $\widehat{\sigma}_{\text{MAD},n}(x)$ and the broken (- -) lines correspond to $\widehat{\sigma}_{\text{BT},n}(x)$. Model $M2 : g(x) = 1 + x$, $\sigma(x) = 1 + [1 + \text{sen}(10x)]^2$ with asymmetric contamination. a) CV(left) and SCV (right). b) KCV(left) and $SKCV$(right). c) $CV^{\log}$ (left) and $SCV^{\log}$ (right). d) $KCV^{\log}$ (left) and $SKCV^{\log}$ (right)

6 Concluding Remarks

Robust estimation of the scale function has become an important research problem. Although classical and robust proposals, based on differences of the responses variables, have been introduced to deal with the scale function estimation, the robust bandwidth selection problem has been less studied.

In this paper, the asymptotic behaviour of the local robust data–driven $M-$estimators for the scale function based on successive differences was studied. To be more precise, under regularity conditions, the data–driven robust kernel–based scale estimators based on differences turn out to be strong consistent and asymptotically normally distributed.

As in regression, the selection of the smoothing parameter is an important issue when considering robust estimators of the scale function. In order to select the smoothing parameter, we proposed several robust cross–validation procedures. The performance of the classical and robust bandwidth selection methods as well as the behaviour of robust and non-robust estimators based on these selected bandwidths was compared under the central model and under different contaminations.

Independently of the robust data–driven bandwidth selected, $M-$estimators based on smooth and bounded score functions have the best performance for the models and contaminations studied. Moreover, the robust approaches obtained splitting the cross-validation error into two components one related to bias and the other to the variance tends to perform better.

A Appendix

Proof of Theorem 4.1. For the sake of simplicity, we will begin by fixing some notation. For any $i = 1, \ldots, n-1$, let $Y_i^* = Y_{i+1} - Y_i$, $U_i^* = U_{i+1} - U_i$, $S_x = S(G_x)$ and

$$\begin{aligned} \lambda_{n,b}(x,s,h) &= (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) \chi\left(\frac{Y_i^*}{as}\right) - b \\ \lambda(x,s) &= \mathbb{E}\left[\chi\left(\frac{\sigma(x)U_1^*}{as}\right)\right] - b. \end{aligned}$$

Theorem 4.1 in Boente *et al.* (2010) implies that $\lambda_n(x,s,h_n) \xrightarrow{a.s.} \lambda(x,s)$. Hence, if we assume that

$$\lambda_n(x,s,\widehat{h}_n) - \lambda_n(x,s,h_n) \xrightarrow{a.s.} 0 \tag{A.1}$$

holds, we have that $\lambda_n(x,s,\widehat{h}_n) \xrightarrow{a.s.} \lambda(x,s)$. Using that $\lambda(x,S_x)=0$ and that χ is strictly increasing on $[0,\ \|\chi\|_\infty)$, we have that, for any $\epsilon > 0$, $\lambda(x,S_x-\epsilon) < 0 < \lambda(x,S_x+\epsilon)$. Therefore, for n large enough, we have that $\lambda_n(x,S_x-\epsilon,\widehat{h}_n) < 0 < \lambda_n(x,S_x+\epsilon,\widehat{h}_n)$, *a.s.*, which implies that $\widehat{\sigma}_{\text{M},n}(x,\widehat{h}_n) \xrightarrow{a.s.} S(G_x)$.

It remains to show that (A.1) holds. Define $Z_i = \chi\left(\sigma(x)Y_i^*/(as)\right)$ and write $\lambda_n(x,s,\widehat{h}_n) -$

$\lambda_n(x,s,h_n) = S_{1,n} + S_{2,n}$ with

$$\begin{aligned} S_{1,n} &= (nh_n)^{-1}\sum_{i=1}^n K\left(\frac{x-x_i}{h_n}\right) Z_i \left[h_n/\widehat{h}_n - 1\right] \\ S_{2,n} &= (n\widehat{h}_n)^{-1}\sum_{i=1}^n \left\{K\left(\frac{x-x_i}{\widehat{h}_n}\right) - K\left(\frac{x-x_i}{h_n}\right)\right\} Z_i \end{aligned}$$

In order to derive (A.1) it is enough to show that

$$S_{1,n} \xrightarrow{a.s.} 0 \tag{A.2}$$

$$S_{2,n} \xrightarrow{a.s.} 0. \tag{A.3}$$

Using that $\widehat{h}_n/h \xrightarrow{a.s.} 1$, $|Z_i| \le \|\chi\|_\infty$ and that $(nh_n)^{-1}\sum_{i=1}^n |K((x-x_i)/h_n)| \to \int |K(u)|\,du$, (A.2) follows easily. To obtain (A.3), write

$$S_{2,n} = (nh_n)^{-1}\sum_{i=1}^n K_1\left(\frac{x-x_i}{\xi_n}\right)\left[\frac{h_n}{\xi_n}\right]\left[\frac{h_n}{\widehat{h}_n} - 1\right] Z_i\ ,$$

where $K_1(u) = uK\,'(u)$ and ξ_n is an intermediate point between $\min(h_n,\widehat{h}_n)$ and $\max(h_n,\widehat{h}_n)$. Since $\widehat{h}_n/h_n \xrightarrow{a.s.} 1$, there exists a set $\mathcal{N}$ such that $\mathbb{P}(\mathcal{N}) = 0$ and for all $\omega \notin \mathcal{N}$, $(1/2) < \widehat{h}_n/h_n < 2$ holds, which implies that $\xi_n \in [h^{(m)}, h^{(M)}]$ with $h^{(m)} = h_n/2$ and $h^{(M)} = 2h_n$. From now on, we restrict our attention to those points $\omega \notin \mathcal{N}$. Noting that

$$|S_{2,n}| \le 2\left|\frac{h_n}{\widehat{h}_n} - 1\right| \|\chi\|_\infty\,(nh_n)^{-1}\sum_{i=1}^n \left|K_1\left(\frac{x-x_i}{\xi_n}\right)\right|,$$

it is enough to show that $\limsup |A_n| < \infty$ where $A_n = (\xi_n/h_n)\,C_n$ and $C_n = (n\xi_n)^{-1}\sum_{i=1}^n |K_1((x-x_i)/\xi_n)|$. Using that $\xi_n \in [h^{(m)}, h^{(M)}]$, we get $(\xi_n/h_n) \in [1/2, 2]$ and so, $\xi_n \to 0$ and $n\xi_n \to \infty$ implying that $C_n \to \displaystyle\int |K(u)|\,du$ which concludes the proof. □

Proof of Theorem 4.2. Let $\{Y_i^*\}_{i\ge1}$, $\{U_i^*\}_{i\ge1}$ and $S_x = S(G_x)$ be as in the proof of Theorem 4.1. Also, let $S_n(h) = (nh_n)^{1/2}\lambda_{n,b}(x, S_x, h)$ with

$$\lambda_{n,b}(x,s,h) = (nh)^{-1}\ \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)\chi\left(\frac{Y_i^*}{as}\right) - b$$

and

$$\widetilde{\lambda}_{1n}(x,s,h) = (nh)^{-1}S_x^{-1}\sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)\chi_1\left(\frac{Y_i^*}{as}\right).$$

Using a second order Taylor's expansion, we obtain

$$0 = (nh_n)^{1/2}\lambda_{n,b}(x, \widehat{\sigma}_{\text{M},n}(x,\widehat{h}_n), \widehat{h}_n) = S_n(\widehat{h}_n) - (nh_n)^{1/2}\left(\widehat{\sigma}_{\text{M},n}(x,\widehat{h}_n) - S_x\right) A_n(\widehat{h}_n)$$

where

$$\begin{aligned} A_n(\widehat{h}_n) &= \widetilde{\lambda}_{1n}(x, S_x, \widehat{h}_n) - \left(\widehat{\sigma}_{\mathrm{M},n}(x, \widehat{h}_n) - S_x\right) B_n(\widehat{h}_n) \\ B_n(\widehat{h}_n) &= \xi_n^{-2}(n\widehat{h}_n)^{-1}\sum_{i=1}^n K\left(\frac{x-x_i}{\widehat{h}_n}\right)\chi_3\left(\frac{Y_i^*}{a\xi_n}\right), \end{aligned}$$

with $\chi_3(u) = \chi_1(u) + \chi_2(u)$ and $\xi_n = \xi_n(x, \widehat{h}_n)$ an intermediate point between $\widehat{\sigma}_{\mathrm{M},n}(x, \widehat{h}_n)$ and S_x. Hence, we have that

$$(nh_n)^{1/2}\left(\widehat{\sigma}_{\mathrm{M},n}(x, \widehat{h}_n) - S_x\right) = S_n(\widehat{h}_n)/A_n(\widehat{h}_n).$$

In the proof of Theorem 4.2 in Boente *et al.* (2010), it is shown that

$$S_n(h_n) \xrightarrow{\mathcal{D}} N\left(\frac{S(G_x)}{\sigma(x)}\beta_1\left(\int K^2(u)du\right)^{\frac{1}{2}}, v\int K^2(u)du\right),$$

hence, to conclude the proof, it will be enough to prove that

$$A_n(\widehat{h}_n) \xrightarrow{p} v_2 \tag{A.4}$$

$$S_n(\widehat{h}_n) - S_n(h_n) \xrightarrow{p} 0. \tag{A.5}$$

Since $\widehat{\sigma}_{\mathrm{M},n}(x, \widehat{h}_n) - S_x \xrightarrow{p} 0$ and considering that (A.12) in Boente *et al.* (2010) implies that $\widetilde{\lambda}_{1n}(x, S_x, h_n) \xrightarrow{p} v_2$, (A.4) follows if we show that

$$\widetilde{\lambda}_{1n}(x, S_x, h_n) - \widetilde{\lambda}_{1n}\left(x, S_x, \widehat{h}_n\right) \xrightarrow{p} 0 \tag{A.6}$$

$$B_n(\widehat{h}_n) = O_p(1). \tag{A.7}$$

The same arguments considered to derive (A.1) can be used to obtain (A.6). Using that $\xi_n = \xi_n(x\,\widehat{h}_n) \xrightarrow{p} S_x$, the bound

$$\left|(n\widehat{h}_n)^{-1}\sum_{i=1}^n K\left(\frac{x-x_i}{\widehat{h}_n}\right)\chi_3\left(\frac{Y_i^*}{a\xi_n}\right)\right| \le \|\chi_3\|_\infty (n\widehat{h}_n)^{-1}\sum_{i=1}^n \left|K\left(\frac{x-x_i}{\widehat{h}_n}\right)\right|,$$

and that analogous arguments to those considered above lead to

$$(n\widehat{h}_n)^{-1}\sum_{i=1}^n \left|K\left(\frac{x-x_i}{\widehat{h}_n}\right)\right| - (nh_n)^{-1}\sum_{i=1}^n \left|K\left(\frac{x-x_i}{h_n}\right)\right| \xrightarrow{p} 0,$$

(A.7) follows from fact that $(nh_n)^{-1}\sum_{i=1}^n \left|K\left(\frac{x-x_i}{h_n}\right)\right| \to \int |K(u)|\,du.$

We now prove (A.5). The fact that $\widehat{\tau}_n = \widehat{h}_n/h_n \xrightarrow{p} 1$, implies that $\mathbb{P}(\widehat{\tau}_n \in [r,s]) \to 1$, with r and s constants satisfying $0 < r < 1 < s$. We now define the stochastic process $V_n(\tau) = (nh_n)^{1/2}\lambda_{n,b}(x, S_x, \tau h_n)$ with $\tau \in [r,s]$, and note that $V_n(\widehat{\tau}_n) = S_n(\widehat{h}_n)$ and $V_n(1) = S_n(h_n)$.

Assume that there exists an stochastic process V belongs to $C[r,s]$, the space of continuous functions on $[r,s]$, such that

$$V_n \xrightarrow{\mathcal{D}} V \,. \tag{A.8}$$

Using that $\widehat{\tau}_n \xrightarrow{p} 1$, we have that for any $\eta > 0$, there exists $\delta > 0$ and $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}\left(|\widehat{\tau}_n - 1| > \delta\right) < \eta/2, \quad \forall n \geq n_0.$$

On the other hand, (A.8) implies that there exists $n_1 \in \mathbb{N}$ such that

$$\mathbb{P}\left(\sup_{|\tau-1|<\delta} |V_n(\tau) - V_n(1)| > \epsilon\right) \leq \eta/2, \quad \forall n \geq n_1.$$

These inequalities imply that $\forall n \geq \max(n_0, n_1)$

$$\mathbb{P}\left(|V_n(\widehat{\tau}_n) - V_n(1)| > \epsilon\right) \leq \mathbb{P}\left(|\widehat{\tau}_n - 1| > \delta\right) + \mathbb{P}\left(\sup_{|\tau-1|<\delta} |V_n(\tau) - V_n(1)| > \epsilon\right) \leq \eta \,,$$

and (A.5) follows. To prove (A.8), define $U_n(\tau) = (nh_n)^{1/2}\left\{\lambda_{n,b}(x, S_x, \tau h_n) - \mathbb{E}\left[\lambda_{n,b}(x, S_x, \tau h_n)\right]\right\}$. Then, $V_n(\tau) = U_n(\tau) + (nh_n)^{1/2}\mathbb{E}\left[\lambda_{n,b}(x, S_x, \tau h_n)\right] = U_n(\tau) + \gamma_n(\tau)$. Using analogous arguments to those considered to prove Lemma A.2 in Boente *et al.* (2010), it is easy to show that

$$\sup_{\tau\in[r,s]} \left|(nh_n)^{1/2}\mathbb{E}\left[\lambda_{n,b}(x, S_x, \tau h_n)\right] - \beta_1 \frac{S_x}{\sigma(x)} \frac{v_2}{\tau^{1/2}}\right| \to 0,$$

i.e., $\gamma_n(\tau) \to \gamma(\tau) = \beta_1 S_x v_2 (\sigma(x)\tau^{1/2})^{-1}$ uniformly on $[r,s]$. Hence (A.8) follows if we show that $U_n \xrightarrow{\mathcal{D}} U$, where U is a gaussian stochastic process on $C[r,s]$. Therefore, it remains to show that

(i) For any $\tau_1, \dots \tau_k \in [r,s]$, $(U_{\tau_1}, \dots U_{\tau_k})$ converge to a multivariate normal distribution $N(0,\Sigma)$.

(ii) The sequence $\{U_n(r)\}_{n\geq 1}$ is tight.

(iii) There exists a constant c such that $\mathbb{E}\left[U_n(\tau_2) - U_n(\tau_1)\right]^2 \leq c\,(\tau_2 - \tau_1)^2$, for $0 < r < \tau_1 < \tau_2 < s < 1$.

As it is well known, to derive (i) it is enough to show that, for any vector $\mathbf{a} = (a_1, \dots, a_k)^{\mathrm{T}} \in \mathbb{R}^k$, $W_n = \sum_{j=1}^k a_j U_n(\tau_j)$ converge to a normal distribution. Note that

$$\begin{aligned} U_n(\tau) &= (nh_n)^{1/2} \frac{1}{n\tau h_n} \sum_{i=1}^n K\left(\frac{x - x_i}{\tau h_n}\right) Z_i \\ W_n &= (nh_n)^{-1/2} \sum_{i=1}^n K^*\left(\frac{x - x_i}{h_n}\right) Z_i \end{aligned}$$

with $Z_i = \chi\left(Y_i^*/(aS_x)\right) - \mathbb{E}\chi\left(Y_i^*/(aS_x)\right)$ and $K^*(u) = \sum_{j=1}^k (a_j/\tau_j) K\left(u/\tau_j\right)$. The convergence of $\{W_n\}_{n\geq 1}$ to the normal distribution is now an immediate consequence of Theorem 4.2. in Boente *et al.* (2010).

The proof of (ii) follows immediately from the fact that $U_n(r)$ converges in distribution.

We now prove (iii). Since χ is bounded, there exists a constant k_1 such that $\text{VAR}\,(Z_i) \le k_1$ and $\text{COV}\,(Z_i, Z_{i+1}) \le k_1$ for any $i \ge 1$. Hence $\mathbb{E}\,[U_n(\tau_2) - U_n(\tau_1)]^2 \le H_{1,n} + H_{2,n}$, where

$$\begin{aligned} H_{1,n} &= k_1 \frac{1}{nh_n} \sum_{i=1}^{n} \left(\frac{1}{\tau_2} K\left(\frac{x - x_i}{\tau_2 h_n}\right) - \frac{1}{\tau_1} K\left(\frac{x - x_i}{\tau_1 h_n}\right)\right)^2 \\ H_{2,n} &= 2k_1 \frac{1}{nh_n} \sum_{i=1}^{n-1} \left(\frac{1}{\tau_2} K\left(\frac{x - x_i}{\tau_2 h_n}\right) - \frac{1}{\tau_1} K\left(\frac{x - x_i}{\tau_1 h_n}\right)\right) \times \left(\frac{1}{\tau_2} K\left(\frac{x - x_{i+1}}{\tau_2 h_n}\right) - \frac{1}{\tau_1} K\left(\frac{x - x_{i+1}}{\tau_1 h_n}\right)\right). \end{aligned}$$

The Lipschitz continuity of K implies that $H_{1,n} \le T_{1,n} + T_{2,n}$, where

$$\begin{aligned} T_{1,n} &= 2k_1 \frac{1}{\tau_2^2} \left(\frac{1}{\tau_2} - \frac{1}{\tau_1}\right)^2 \frac{1}{nh_n} \sum_{i=1}^{n} \left(K'\left(\frac{x - x_i}{\xi_i h_n}\right)\right)^2 \left(\frac{x - x_i}{h_n}\right)^2 = 2k_1 \frac{1}{\tau_2^2} \left(\frac{1}{\tau_2} - \frac{1}{\tau_1}\right)^2 R_{1,n} \\ T_{2,n} &= 2k_1 \tau_1 \left(\frac{1}{\tau_2} - \frac{1}{\tau_1}\right)^2 \frac{1}{n\tau_1 h_n} \sum_{i=1}^{n} K^2 \left(\frac{x - x_i}{\tau_1 h_n}\right) = 2k_1 \tau_1 \left(\frac{1}{\tau_2} - \frac{1}{\tau_1}\right)^2 R_{2,n} \end{aligned}$$

and $\xi_i \in (\tau_1, \tau_2)$. Using that $\tau_1, \tau_2 \in [r, s]$ and the assumptions on the design points and proceeding as in Theorem 3.1 of Boente *et al.* (1997), it is easy to show that, for all $n \ge 1$, $R_{1,n} \le B$, with B a fixed constant. If we note that for any $n \ge 1$, $R_{2,n} \le C$ with C a fixed constant, we get easily that, $T_{1,n} \le 2k_1(s^2/(r^4 B))\,(\tau_2 - \tau_1)^2$ and $T_{2,n} \le 2k_1(s/(r^2 C))\,(\tau_2 - \tau_1)^2$. So, $H_{1,n} \le c_1\,(\tau_2 - \tau_1)^2$ with $c_1 = 2k_1\left(s^2\,B/r^4 + s\,C/r^2\right)$. Using analogous arguments, it can be shown that there exists a constant c_2 such that $H_{2,n} \le c_2\,(\tau_2 - \tau_1)^2$. Hence (iii) follows with $c = c_1 + c_2$. $\square$

References

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

[2] Billingsley, P. (1999). *Convergence of Probability Measures*. John Wiley & Sons.

[3] Bianco, A. and Boente, G. (2007). Robust estimators under a semiparametric partly linear autoregression model: Asymptotic behavior and bandwidth selection. *J. Time Series Anal.*, **28**, 274-306.

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

[5] Boente, G. and Fraiman, R. (1995). Asymptotic Distribution of Data-driven smoothers in density and regression estimation under dependence. *Canadian Journal of Statistics*, **23**, 383-397.

[6] 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.

[7] Boente, G. and Rodriguez, D. (2008). Robust bandwidth selection in semiparametric partly linear regression models: Monte Carlo study and influential analysis . *Computational Statistics and Data Analysis*, **52**, 383-397.

[8] Boente, G., Ruiz, M. and Zamar, R. (2010). On a robust local estimator for the scale function in heteroscedastic nonparametric regression. *Statistics and Probability Letters*, **80**, 1185–1195.

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

[10] Buckley, M., Eagleson, G.and Silverman, B. (1988). The Estimation of Residual Variance in Nonparametric Regression. *Biometrika*, **75**, 189–199.

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

[12] Carter, C. and Eagleson, G. (1992). A comparison of variance estimators in nonparametric regression. *Journal of the Royal Statistical Society, Series B*, **54**, 773–780.

[13] Carroll, R. and Ruppert, D. (1988). *Transformation and Weighting in Regression.* Chapman & Hall.

[14] 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.

[15] Dette, H. (2002). A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. (2002) . *Journal of statistical planning and inference*, **103**, 311–329.

[16] Dette, H. and Hetzler, B. (2008). A simple test for the parametric form of the variance function in nonparametric regression. To appear in *The Annals of the Institute of Statistical Mathematics.*

[17] Gasser, T. and Müller, H. G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, **11**, 171185.

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

[19] 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.

[20] Hall, P. and Marron, J. (1990). On variance estimation in nonparametric regression. *Biometrika*, **77**, 415–419.

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

[22] 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.

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

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

[25] Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge. University Press.

[26] Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and Semiparametric Models.* Springer.

[27] Huber, P. (1981). *Robust Statistics.* John Wiley & Sons.

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

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

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

[31] Levine, M. (2006). Bandwidth selection for a class of difference-based variance estimators in the nonparametric regression: A possible approach. *Comput. Statist. Data Anal.*, **50**, 3405-3431.

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

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

[34] Müller, H. and Stadtmüller, U. (1987). Variable bandwidth kernel estimators of regression. *The Annals of Statistics*, **15**, 182–201.

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

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

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

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

[39] Wang F. and Scott D. (1994). The L1 method for robust nonparametric regression. *Journal of American Statistical Association*, **89**, 65–76.

[40] Watson, G. (1964). Smooth regression analysis. *Sankhyā, Series A*, **26**, 359–372.