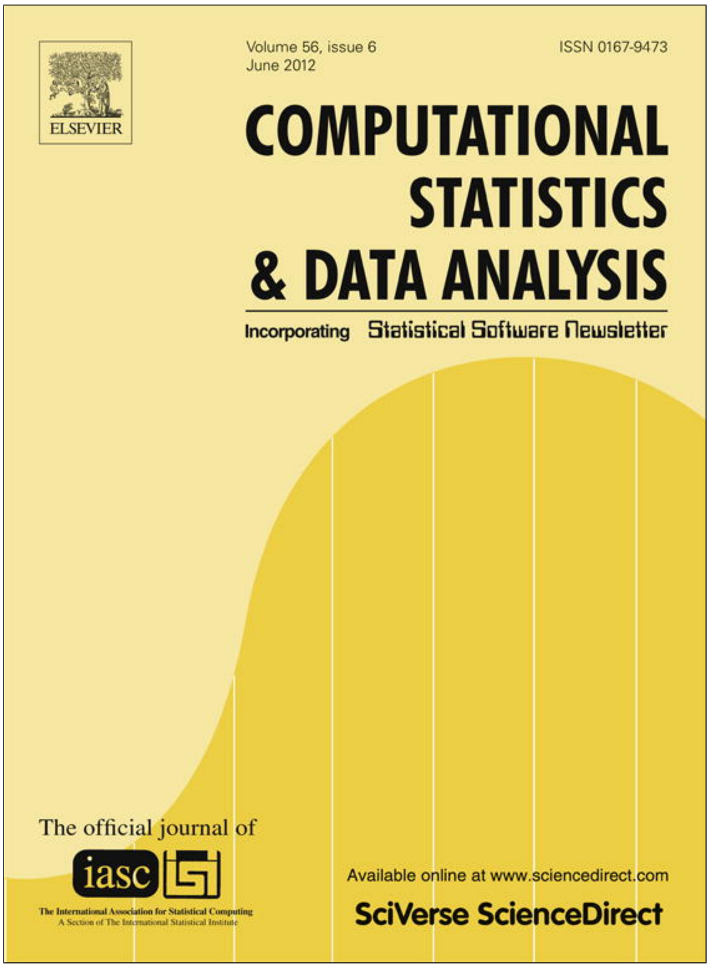
Volume 56, issue 6
June 2012
ISSN 0167-9473
ELSEVIER
COMPUTATIONAL
STATISTICS
& DATA ANALYSIS
Incorporating Statistical Software Newsletter
The official journal of
iasc
The International Association for Statistical Computing
A Section of The International Statistical Institute
Available online at www.sciencedirect.com
SciVerse ScienceDirect

Computational Statistics and Data Analysis 56 (2012) 1594–1608

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Bandwidth choice for robust nonparametric scale function estimation

Graciela Boente [a,b,*], Marcelo Ruiz [c], Ruben H. Zamar [d]

[a] Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina
[b] CONICET, Argentina
[c] Universidad Nacional de Río Cuarto, Argentina
[d] University of British Columbia, Canada

ARTICLE INFO

Article history:
Received 20 January 2011
Received in revised form 12 September 2011
Accepted 1 October 2011
Available online 12 October 2011

Keywords:
Cross-validation
Data-driven bandwidth
Heteroscedasticity
Local *M*-estimators
Nonparametric regression
Robust estimation

ABSTRACT

We introduce and compare several robust procedures for bandwidth selection when estimating the variance function. These bandwidth selectors are to be used in combination with the robust scale estimates introduced by Boente et al. (2010a). We consider two different robust cross-validation strategies combined with two ways for measuring the cross-validation prediction error. The different proposals are compared with non robust alternatives using Monte Carlo simulation. We also derive some asymptotic results to investigate the large sample performance of the corresponding robust data-driven scale estimators.

1. Introduction

We consider the heteroscedastic regression model

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

where $0 \le x_1 \le \cdots \le x_n \le 1$ are fixed design points and the errors $\{U_i\}_{i\ge 1}$ are i.i.d. random variables with a common distribution F_0. In this non-parametric setting the functions $g(x)$ and $\sigma(x) > 0$ are assumed to be unspecified and unknown.

Most research in nonparametric regression focuses on the estimation of the regression function g and the scale function σ is often assumed to be constant (homoscedasticity) and treated as a nuisance parameter. This simplistic approach has several limitations, including that: (i) homoscedasticity may not be a realistic assumption in some applications; (ii) robust and efficient estimates of σ are needed to construct confidence and prediction intervals; (iii) the performance of the robust estimate of the regression function relies on the performance of its companion scale estimator (iv) the scale function may be the parameter of main interest for the phenomena under study. A nice discussion of applications of scale function estimates can be found in Carroll and Ruppert (1988), Hall et al. (1990), Dette et al. (1998) and Levine (2003) among others.

Hall et al. (1990) proposed preliminary estimates for the scale based on differences for homoscedatic nonparametric regression models, generalizing initial proposals by Rice (1984) and Gasser et al. (1986). Müller and Stadtmüller (1987) and Brown and Levine (2007), among others, extended this class of estimates to heteroscedastic models while Fan and Yao (1998) considered a local linear variance estimator based on squared residuals obtained from a preliminary regression estimator.

* Correspondence to: Instituto de Cálculo, FCEyN, UBA, Ciudad Universitaria, Pabellón 2, Buenos Aires, C1428EHA, Argentina. Tel.: +54 11 45763375; fax: +54 11 45763375.
E-mail addresses: gboente@dm.uba.ar, gboente@fibertel.com.ar (G. Boente), mruiz@exa.unrc.edu.ar (M. Ruiz), ruben@stat.ubc.ca (R.H. Zamar).

0167-9473/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2011.10.002

Recently, Dette and Marchlewski (2008) used scale estimates based on differences to derive tests for checking whether the variance function belongs to a parametric family in a partly linear regression model. In particular, tests for homocedasticity are considered therein.

It is well known that for both types of models – homoscedastic and heterocedastic – the classical scale estimators based on minimizing square residuals are sensitive to departures from the model error distribution F_0. The need for robust estimates of the scale function is well established in the statistical literature, (see for example Härdle and Gasser (1984), Härdle and Tsybakov (1988) and Boente and Fraiman (1989), Leung et al. (1993), Boente et al. (1997), Cantoni and Ronchetti (2001), and Leung (2005), among others). Also, Dette and Marchlewski (2008) pointed out the need for robust methods when analyzing the refinery data set studied by Daniel and Wood (1980).

When $\sigma(x)$ is constant (homoscedastic nonparametric regression model), Boente et al. (1997) proposed the median of the absolute differences $|Y_{i+1} - Y_i|$, $1 \le i \le n-1$, based on ideas in Rice (1984). Subsequently, Ghement et al. (2008) introduced the general class of global M-estimators based on differences. Finally, Boente et al. (2010a) considered the heteroscedastic model and defined local M-estimates of scale based on differences (LMD). Besides, Dette and Marchlewski (2010) considered a robust test for homoscedasticity based on the empirical process of the residuals from a robust regression fit obtained using a local M-estimator of g.

As in nonparametric regression, LMD-estimators critically hinge on a smoothing parameter that must be estimated from the available data. Large bandwidths produce smooth curves with high bias, while small bandwidths produce more wiggly curves. This well known trade-off between bias and variance led, in the case of estimation of the regression function g, to several proposals for selecting the smoothing parameter, such as cross-validation and plug-in methods. Much less development has been reported so far for the case of estimation of the scale function σ. Levine (2003) derived an expression for the optimal bandwidth that led to the plug-in approach in Levine (2006). Later on, K-fold cross-validation procedures were recommended and studied for classical (non-robust) estimates based on square differences. Not surprisingly these procedures are very sensitive to outliers.

As far as we know, there are no previous proposals for robust bandwidth estimators to be used when estimating the scale function. The aim of this paper is to introduce bandwidth selectors resistant to outliers, which, combined with LMD estimates, yield robust data-driven scale estimators based on differences.

The rest of the paper is organized as follows. In Section 2, we briefly review 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. A real data example is analyzed in Section 4. The asymptotic properties of the robust local M-estimates based on random bandwidths are investigated in Section 5. The results of some numerical experiments conducted to evaluate the performance of the different procedures are described in Section 6. Finally, Section 7 provides some concluding remarks. Proofs can be found in the Appendix.

2. The estimators

We now review the definition of *local M-estimates of scale based on differences* introduced by Boente et al. (2010a) and implemented in this paper.

For any $x \in (0, 1)$, let $\widehat{\sigma}_{\mathrm{M},n}(x) = \inf\left\{s > 0 : \sum_{i=1}^{n-1} w_{n,i}(x)\chi((Y_{i+1} - Y_i)/(as)) \le b\right\}$, where χ is a score function and $w_{n,i}(x) = w_{n,i}(x, h_n)$, $i = 1, \ldots, n-1$, are kernel weights, such as the Nadaraya–Watson or the Rosenblatt weights, respectively defined as $w_{n,i}(x, h) = L((x - x_i)/h)\left\{\sum_{j=1}^{n} L\left((x - x_j)/h\right)\right\}^{-1}$ and $w_{n,i}(x, h) = (nh)^{-1}L((x - x_i)/h)$. The bandwidth parameter $h_n > 0$ regulates the trade-off between bias and variance, $a \in (0, \infty)$ is chosen to attain Fisher-consistency at the central model and $b \in (0, 1)$ regulates the degree of robustness for the estimator. More precisely, the tuning constants a and b satisfy the equations $\mathbb{E}[\chi((Z_2 - Z_1)/a)] = b$ and $\mathbb{E}[\chi(Z_1)] = b$, respectively, where $\{Z_i\}_{i=1,2}$ are 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) includes, among others, the classical *local Rice estimator* by taking $\chi(x) = x^2$, $a = \sqrt{2}$ and $b = 1$. Two other particular cases considered in our simulations are the *local* MAD *estimator* and the *local M-estimator with Beaton–Tukey (BT) score 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 $b = 1/2$. The *local M-estimator with BT function*, denoted $\widehat{\sigma}_{\mathrm{BT},n}(x)$, uses the score function

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

introduced by Beaton and Tukey (1974), with tuning constant $c = 0.70417$, $a = \sqrt{2}$ and $b = 3/4$. Ghement et al. (2008) showed that for homoscedastic models, under some regularity and design conditions, 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 (as in the case of local M-estimators of the regression function, see Maronna et al., 2006, Chapter 4). The empirical breakdown point of the *local M-estimator with Beaton–Tukey (BT) score function* is discussed in Boente et al. (2010a). The value $c = 0.70417$ is numerically determined there to guarantee Fisher-consistency of the scale functional at the central Gaussian model when $b = 3/4$. More precisely, c is the solution of $\mathbb{E}[\chi_c(Z_1)] = 3/4$ where $Z_1 \sim N(0, 1)$. On the other hand, the value $a = \sqrt{2}$ provides immediately the solution to $\mathbb{E}[\chi_c((Z_2 - Z_1)/a)] = 3/4$ for Z_1, Z_2 i.i.d. such that $Z_1 \sim N(0, 1)$.

3. Robust bandwidth selectors

An important issue regarding kernel weights is the selection of the smoothing parameter h_n which regulates the trade-off between bias and variance. This led to the development of different automatic (data-driven) methods for selecting h_n, such as cross-validation and plug-in procedures. A good discussion of these methods can be found in Härdle (1990) and Härdle et al. (2004). Unfortunately, these procedures are not robust and their sensitivity to outliers 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 (smoothing parameter).

The study of automatic bandwidth selectors for the scale function is much less developed. Levine (2003) derived the optimal bandwidth for scale function estimators based on squared differences. This led to the plug-in approach discussed in Levine (2006). However, Levine (2006) also points out the limitations of the plug-in method and recommended the use of K-fold cross-validation to obtain scale estimators which are not too sensitive to the actual shape of the mean function in the case of clean data.

For completeness, we review 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 $n_j \geq 2$, $\sum_{j=1}^{K} n_j = n$. Let $\{(\widetilde{x}_i^{(j)}, \widetilde{y}_i^{(j)})\}_{1\leq i\leq n_j}$ be the pairs of the j-th subset with the values of $\widetilde{x}_i^{(j)}$ arranged in ascending order. Similarly, let $\{(x_i^{(j)}, y_i^{(j)})\}_{1\leq i\leq n-n_j}$ denote the pairs in the complement of the j-th subset, again with the $x_i^{(j)}$ arranged in ascending order. The set $\{(x_i^{(j)}, y_i^{(j)})\}_{1\leq i\leq n-n_j}$ will be the training set and $\{(\widetilde{x}_i^{(j)}, \widetilde{y}_i^{(j)})\}_{1\leq i\leq n_j}$ the validation set. Moreover, denote $\Delta_i^{(j)} = (y_{i+1}^{(j)} - y_i^{(j)})/\sqrt{2}$ and $D_{i,(j)} = |\widetilde{y}_{i+1}^{(j)} - \widetilde{y}_i^{(j)}|/\sqrt{2}$ the successive differences and the absolute value of the successive differences within each subset, respectively. Let $\widehat{\sigma}_{\text{RICE},n}^{(j)}(x, h)$ and $\widehat{\sigma}_{\text{M},n}^{(j)}(x, h)$ be the classical and robust scale estimators computed using a bandwidth h and the j-th training subset $\{(x_i^{(j)}, y_i^{(j)})\}_{1\leq i\leq n-n_j}$, 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

$$CV_{\text{LS,KCV}}(h) = \frac{1}{n}\sum_{j=1}^{K}\sum_{i=1}^{n_j-1}\left[D_{i,(j)}^2 - \left(\widehat{\sigma}_{\text{RICE},n}^{(j)}(\widetilde{x}_i, h)\right)^2\right]^2. \tag{3}$$

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 the log scale,

$$CV_{\text{LS,KCV}}^{\log}(h) = \frac{1}{n}\sum_{j=1}^{K}\sum_{i=1}^{n_j-1}\left[\log\left(D_{i,(j)}\right) - \log\left(\widehat{\sigma}_{\text{RICE},n}^{(j)}(\widetilde{x}_i, h)\right)\right]^2 \tag{4}$$

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

Even if a robust scale estimator is used, i.e., $\widehat{\sigma}_{\text{M},n}^{(j)}(x, h)$ instead of $\widehat{\sigma}_{\text{RICE},n}^{(j)}(x, h)$, the K-fold cross-validation bandwidth selector remains sensitive to outliers, because large residuals are not downweighted. For that reason, to achieve robustness it is necessary to define a robust scale-based procedure such as

$$CV_{\text{ROB,KCV}}(h) = \frac{1}{n}\sum_{j=1}^{K} s_j^2(h)\sum_{i=1}^{n_j-1}\psi^2\left(\frac{e_{i,(j)}}{s_j(h)}\right), \tag{5}$$

where $e_{i,(j)} = D_{i,(j)}^2 - \left(\widehat{\sigma}_{\text{M},n}^{(j)}(\widetilde{x}_i, h)\right)^2$, $s_j(h) = \text{median}|e_{i,(j)}|$ and ψ is a bounded score function such as the Huber function $\psi_{c_1}(y) = \min\{\max\{-c_1, y\}, c_1\}$. The robust K-fold cross-validation bandwidth is then defined as $\widehat{h}_{\text{ROB,KCV}} = \text{argmin}_{h\in\mathcal{H}} CV_{\text{ROB,KCV}}(h)$.

Similarly, a robust K-fold *log-scale* cross-validation procedure can be defined minimizing the cross-validation error

$$CV_{\text{ROB,KCV}}^{\log}(h) = \frac{1}{n}\sum_{j=1}^{K} s_{j,\log}^2(h)\sum_{i=1}^{n_j-1}\psi^2\left(\frac{e_{i,(j)}^{\log}}{s_{j,\log}(h)}\right), \tag{6}$$

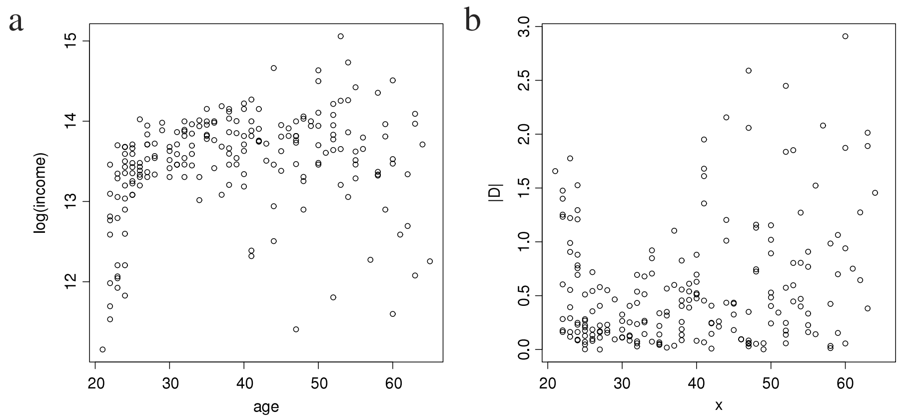

Fig. 1. (a) Plot of the Canadian log(income) data set (b) Scatter plot of the absolute differences $|D_i|$.

where now $e^{\log}_{i,(j)} = \log\left(D_{i,(j)}\right) - \log\left(\widehat{\sigma}^{(j)}_{M,n}(\widetilde{x}_i, h)\right)$, $s_{j,\log}(h) = \text{median}|e^{\log}_{i,(j)}|$. The corresponding robust selector is defined as $\widehat{h}^{\log}_{\text{ROB,KCV}} = \text{argmin}_{h\in\mathcal{H}} CV_{\text{ROB,KCV}}(h)$.

Leave-one-out cross-validation is an important particular case of K-fold cross-validation obtained when $K = n$ and $n_j = 1$. We will denote by $\widehat{h}_{\text{LS,CV}}$ and $\widehat{h}^{\log}_{\text{LS,CV}}$ the classical cross-validation bandwidths obtained by minimizing the leave-one-out errors $CV_{\text{LS,CV}}$ and $CV^{\log}_{\text{LS,CV}}$ defined using (3) and (4) with $K = n$ and $n_j = 1$. Similarly, the robust cross-validation bandwidth related to (5) and (6) with $K = n$ and $n_j = 1$ will be denoted by $\widehat{h}_{\text{ROB,CV}}$ and $\widehat{h}^{\log}_{\text{ROB,CV}}$ while the corresponding cross-validation errors are $CV_{\text{ROB,CV}}$ and $CV^{\log}_{\text{ROB,CV}}$, respectively.

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 also be defined. This approach splits the cross-validation error into two components, one related to the bias and the other to the variance. Hence, the robust split K-fold cross-validation error can be defined as

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

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

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

where now $e^{\log}_{i,(j)} = \log\left(D_{i,(j)}\right) - \log\left(\widehat{\sigma}^{(j)}_{M,n}(\widetilde{x}_i, h)\right)$ and $\widehat{h}^{\log}_{\text{ROB,SKCV}} = \text{argmin}_{h\in\mathcal{H}} CV^{\log}_{\text{ROB,SKCV}}(h)$. The corresponding robust split leave-one-out errors are denoted by $CV_{\text{ROB,SCV}}$ and $CV^{\log}_{\text{ROB,SCV}}$ while $\widehat{h}_{\text{ROB,SCV}}$ and $\widehat{h}^{\log}_{\text{ROB,SCV}}$ stand for the and the related optimal bandwidths, respectively.

4. Real data example

We report the results obtained from the robust analysis of 205 observations from the 1971 Canadian Population Census. These data can be found in the library *SemiPar* in R. The response variable Y_i is the logarithm of the individual annual income while the covariate x_i is the individual age. These data were first analyzed (in Statistics) by Ullah (1985). Levine (2003) points out that the increase of income variability with age is a well known fact in Labor Economics. The scatter plot of the absolute differences $|D_i| = |Y_{i+1} - Y_i|$ versus x_i displayed on Fig. 1 shows that in fact income variability is larger for people over 45.

We estimated the scale function using Rice classical estimator and the robust procedure based on the biweight Beaton–Tukey score function with tuning constants $c = 0.70417$, $a = \sqrt{2}$ and $b = 3/4$ (see Remark 2.1).

The bandwidth $\widehat{h}_{\text{LS,CV}} = 6.6$ is obtained using leave-one-out cross-validation with Rice estimator. The smaller bandwidth $\widehat{h}^{\log}_{\text{ROB,SCV}} = 4.1$ is obtained using the robust approach described in Section 3. The numerical results in Section 6 show that this is the best performing method among the considered robust procedures. It is worth noticing that, as it will be pointed out in Section 6, $\widehat{h}^{\log}_{\text{LS,CV}}$ leads to undersmoothing (in this case $\widehat{h}^{\log}_{\text{LS,CV}} = 1$) and so, it is not a recommendable procedure.

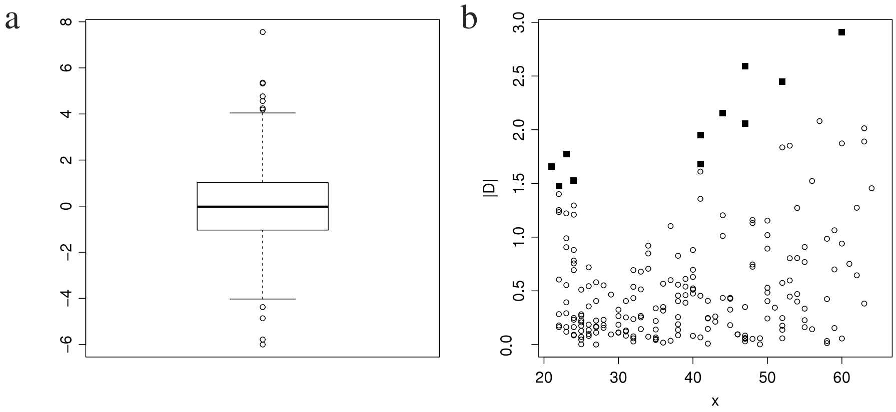

Fig. 2. (a) Boxplot of the residuals $r_i = (Y_{i+1} - Y_i)/\widehat{\sigma}_{\mathrm{BT},n}(x_i)$ (b) Scatter plot of the absolute differences $|D_i|$ with the outliers as filled squares.

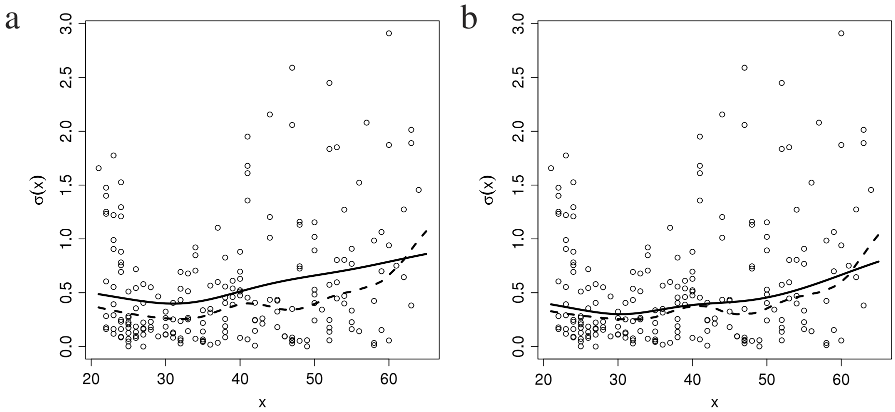

Fig. 3. Estimated variance functions, the solid and broken (- -) lines correspond to the classical and robust estimators, respectively. (a) Original data set (b) Data set without outliers.

Fig. 2 shows the boxplot of the residuals $r_i = (Y_{i+1} - Y_i)/\widehat{\sigma}_{\mathrm{BT},n}(x_i)$. There is a large outlier (about 8) and ten moderate outliers. We repeated the classical and robust analysis after removing the Y_i responsible for these outliers. The resulting bandwidths are now $\widehat{h}_{\mathrm{LS,CV}} = 5.15$ and $\widehat{h}^{\log}_{\mathrm{ROB,SCV}} = 4.2$. However, after studying the residuals we detect four additional large differences that were masked in the previous analysis. After further removing these observations, the cross-validation bandwidths are $\widehat{h}_{\mathrm{LS,CV}} = 4.85$ and $\widehat{h}^{\log}_{\mathrm{ROB,SCV}} = 4.0$. The plot of the scale function estimators with the full data and after removing the outliers are given in Fig. 3. The plot of the classical estimator without outliers is pretty similar to that of the robust scale function estimator, except for ages over 62. This difference may be due to the boundary bias effect of kernel estimators that has been widely noticed in the literature. We conclude that the outliers in the sample lead to over-smoothing, hiding some interesting features which can be uncovered using our resistant procedures.

5. Asymptotic behavior of data-driven local scale M-estimates

Boente et al. (2010a) derived the asymptotic behavior of the local scale M-estimators based on differences assuming a deterministic sequence of bandwidths (smoothing parameters). Under mild assumptions, the estimators are strongly consistent and asymptotically normal. However, in practice the bandwidth parameter is not deterministic but data-driven and, therefore, random. The aim of this section is to extend the asymptotic results of Boente et al. (2010a) to the case where the smoothing parameter depends on the data.

To be more precise, if we denote by $\widehat{\sigma}_{\mathrm{M},n}(x, h_n)$ the robust local M-estimate computed using a sequence of bandwidths h_n, we wish to derive the asymptotic properties of the robust local M-estimate $\widehat{\sigma}_{\mathrm{M},n}(x, \widehat{h}_n)$ where $\widehat{h}_n = \widehat{h}_n(Y_1, \ldots, Y_n)$ stands for a random, data-driven bandwidth.

Our results show that to a first-order approximation the data-driven scale estimators behave as well as the related estimators obtained with the fixed bandwidth h_n provided $\widehat{h}_n/h_n \xrightarrow{p} c > 0$ and so, no knowledge of the order of convergence of $\widehat{h}_n$ to the "optimal" bandwidth h_n is needed.

This type of research results exists in the literature only for the robust regression function estimation (see, for instance Boente and Fraiman, 1995). We are not aware of similar results for local robust difference-based estimators.

Without loss of generality, throughout this section, we will assume that $c = 1$, i.e., that the deterministic sequence of bandwidths has been corrected so that $\widehat{h}_n/h_n \xrightarrow{p} 1 > 0$. Note that, in particular, if moment conditions are required instead of assumption **N2**, the asymptotic behavior of the plug-in data-driven estimators studied in Levine (2006) can be derived from our results.

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

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

We prove the consistency of our estimators under the following conditions.

C1. (i) $L : \mathbb{R} \longrightarrow \mathbb{R}$ is even, bounded and $\int |L(u)|du < \infty$, $\int L^2(u)du < \infty$ and $\lim_{u\to\infty} u^2L(u) = 0$.
(ii) $\int L(u)du = 1$.
(iii) L is continuously differentiable and $L_1(u) = uL'(u)$ is such that L_1 and L_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 5.1. *Let U_1 and U_2 be i.i.d. random variables with distribution G and let G_x be the distribution of $\sigma(x)(U_2-U_1)$. Assume that* **C1** *to* **C4** *hold and $\lim_{n\to\infty} nh_n/\log(n) = \infty$ 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.

We derive the asymptotic distribution of our estimators under the following conditions.

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 $u \ne 0$, $v \ne 0$, $\nu(u, v) = \mathbb{E}\left|\chi'(uU_2 + vU_1)U_2\right| < \infty$, 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) $\lim_{n\to\infty}(nh_n)^{-1/2}\sum_{i=1}^{n-1} L\left(\frac{x-x_i}{h_n}\right)(\sigma(x) - \sigma(x_i)) = \beta_1$
(ii) $\lim_{n\to\infty}(nh_n)^{-1/2}\sum_{i=1}^{n-1} L\left(\frac{x-x_i}{h_n}\right)(\sigma(x_i) - \sigma(x))^2 = 0$.

Remark 5.1. Assumption **N1** is usual in non-parametric settings. Note also that **N2** does not necessarily impose the existence of moments on the distribution of the errors; for instance this hypothesis is fulfilled if the errors $\{U_i\}_{i\ge1}$ have Cauchy distribution and χ belongs to the Beaton–Tukey family. **N3** is related to the asymptotic bias. Assume that $\int u^2L(u)du < \infty$. 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 $\beta_1 = \gamma\sigma''(x)\int u^2L(u)du\left(\int L^2(u)du\right)^{1/2}$.

Theorem 5.2. *Assume* **C3**, **C4** *and* **N1** *to* **N3** *hold. Then*

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

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

$$v_1 = v_1(G_x) = \mathrm{VAR}\left[\chi\left(\frac{\sigma(x)(U_2-U_1)}{aS(G_x)}\right)\right] + 2\beta\,\mathrm{cov}\left[\chi\left(\frac{\sigma(x)(U_2-U_1)}{aS(G_x)}\right), \chi\left(\frac{\sigma(x)(U_4-U_3)}{aS(G_x)}\right)\right]$$

$$v_2 = v_2(G_x) = \mathbb{E}\left[\chi'\left(\frac{\sigma(x)(U_2-U_1)}{aS(G_x)}\right)\left(\frac{\sigma(x)(U_2-U_1)}{a(S(G_x))^2}\right)\right],$$

$\beta = \int L^2(u)du$ and $\{U_i\}_{i\ge1}$ are i.i.d. random variables with distribution G.

6. Numerical experiments

In this section, we numerically explore the finite sample behavior of different data-driven scale function estimators and bandwidth selectors. In Section 6.1, we report the results of a Monte Carlo study comparing the performance of classical and robust estimators (using data-driven smoothing parameters) under different models (regression functions and scale functions), sample sizes and types of contaminations. In Section 6.2, we study the sensitivity to outliers of the optimal bandwidth selectors. In Section 6.3, we numerically show that data-driven bandwidths actually converge to 0 at a rate ensuring consistency and asymptotic normality of the estimators.

6.1. Monte Carlo study

In this subsection, we compare the finite sample performance of the classical scale function estimator $\widehat{\sigma}_{\text{RICE},n}(x, \widehat{h}_n)$ and of the two robust local M-estimators, $\widehat{\sigma}_{\text{MAD},n}(x, \widehat{h}_n)$ and $\widehat{\sigma}_{\text{BT},n}(x, \widehat{h}_n)$, which were introduced in Section 2. The bandwidth $\widehat{h}_n$ is selected using the procedures described in Section 3.

To reduce the heavy computational burden due to cross-validation we generate $N = 500$ independent samples of size $n = 100$. Two different models for the regression and variance components are considered. These models were introduced by Dette and Hetzler (2008) to test homoscedasticity. We report here the results corresponding to one of these two models, the one in which the regression function is $g(x) = 2\sin(4\pi x)$ and the scale function is $\sigma(x) = \exp(x)$. The results when the regression function is linear $g(x) = 1 + x$ and the scale is $\sigma(x) = 1 + [1 + \sin(10x)]^2$ are similar and available in the technical report Boente et al. (2010b). The design points are $x_i = i/(n+1)$, $1 \le i \le n$. The error's distribution is $G(y) = (1-\epsilon)\Phi(y) + \epsilon H(y)$, with Φ equal to the standard normal distribution and H producing two types of contamination,

(a) symmetric outlier contamination where $H = \mathcal{C}(0, \sigma^2)$ is the Cauchy distribution centered at 0 with scale $\sigma = 4$ and
(b) asymmetric contamination where $H = \text{N}(10, 0.1)$ is the normal distribution with mean 10 and variance 0.1.

In the first case, we have a heavy-tailed distribution while in the second case there is a sub-population in the data (see Maronna et al., 2006). The amounts of contamination are $\epsilon = 0, 0.1, 0.2$ and 0.3. We only report here the results for the central model and the asymmetric contamination model which corresponds to the worst scenario. Results for the symmetric contaminations are available in Boente et al. (2010b).

The bandwidths are selected using the cross-validation methods introduced in Section 3. More precisely, we compare the following procedures.

(C.1) The classical K-fold cross-validation criterion $CV_{\text{LS,KCV}}$ for the Rice estimator and the robust K-fold $CV_{\text{ROB,KCV}}$ and split robust K-fold $CV_{\text{ROB,SKCV}}$ for the M-estimates, taking $K = 2$. These procedures are globally called "KCV-procedures". The choice $K = 2$ is made to reduce the computational burden. Naturally, larger values of K would be desirable for sample sizes larger than the one considered in our study, (see Levine (2006) for a comparison of the behavior of the classical procedure for different choices of K).

(C.2) The procedures related to those considered in (C.1) but on a log scale $CV^{\log}_{\text{LS,KCV}}$ (for Rice), $CV^{\log}_{\text{ROB,KCV}}$, $CV^{\log}_{\text{ROB,SKCV}}$ (for the M-estimates) are globally called "$KCV^{\log}$-procedures".

(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. These procedures are globally called "CV-procedures".

(C.4) The leave-one-out procedures on log scale, that is, the classical $CV^{\log}_{\text{LS,CV}}$ for the local Rice estimate and the robust $CV^{\log}_{\text{ROB,CV}}$ and robust split leave-one-out $CV^{\log}_{\text{ROB,SCV}}$ for the local M-estimates. These are globally called "$CV^{\log}$-procedures".

For both the classical and robust procedures, we use Nadaraya–Watson weights with a standard Gaussian kernel. The score function ψ used for the robust procedures is the Huber function, ψ_{c_1}, with tuning constant $c_1 = 1.345$. The minimization of the cross-validation functions is carried out over the grid $i/(n/2)$, $3 \le i \le n/2$, where n is the sample size. To assess the behavior of the selected bandwidth and the performance of each estimator, Tables 1–3 report, as summary measures, the mean and the standard deviation, between brackets, of the resulting bandwidths $\widehat{h}_n$ and the mean of 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 bandwidth h. Note that for the classical Rice estimators and the quadratic loss function, the split procedures equal the cross-validation losses given in (3) and (4), so in the tables we use dot lines instead of repeating the obtained values. Fig. 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 procedures. The density estimates were evaluated using the normal kernel with bandwidth 0.1 in all cases.

It is well known that robust procedures are nonlinear and difficult to compute. This is compounded in our case by the use of the also computationally intensive cross-validation. Our current software is fully implemented in R. For example,

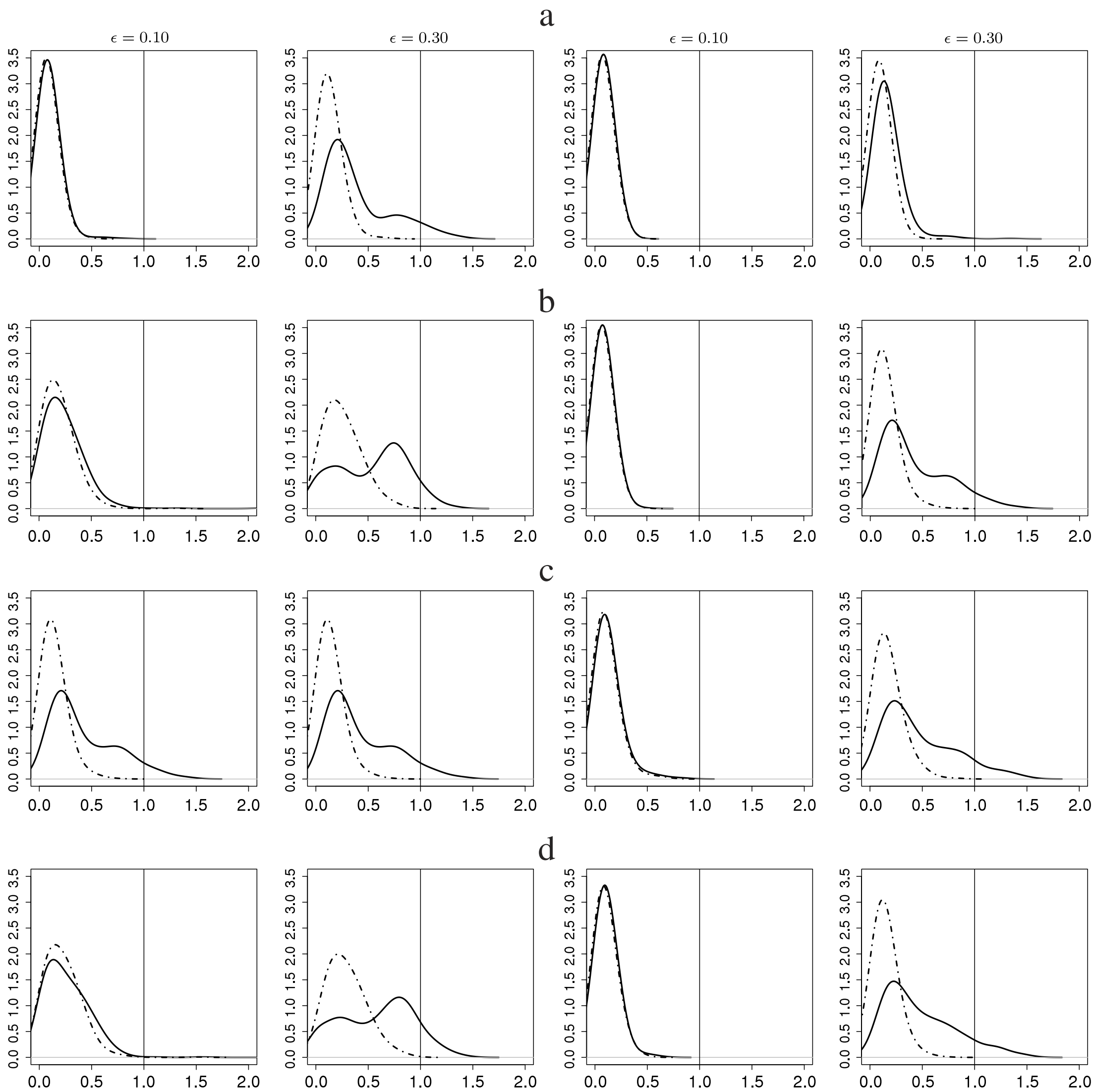

Fig. 4. Density estimator of the ratio between the $\widehat{\text{ISEL}}$ of the robust estimators and that of the non-robust estimator $\widehat{\sigma}_{\text{RICE},n}(x)$. The solid and dashed ($-\cdot-$) lines correspond to $\widehat{\sigma}_{\text{MAD},n}(x)$ and $\widehat{\sigma}_{\text{BT},n}(x)$, respectively. Model $g(x) = 2\sin(4\pi x)$, $\sigma(x) = \exp(x)$ 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).

Table 1
Mean and standard deviation (between brackets) of the selected bandwidths and mean of the $\widehat{\text{ISEL}}$ for the local scale-estimates. Model $g(x) = 2\sin(4\pi x)$, $\sigma(x) = \exp(x)$. Errors with asymmetric contamination. *CV*-procedure.

	Estimator	Selected bandwidth		$\widehat{\text{ISEL}}$	
		$\widehat{h}_{\text{CV}}$	$\widehat{h}_{\text{SCV}}$	$\widehat{\sigma}_n(\cdot,\widehat{h}_{\text{CV}})$	$\widehat{\sigma}_n(\cdot,\widehat{h}_{\text{SCV}})$
$\epsilon = 0$	$\widehat{\sigma}_{\text{RICE},n}$	0.190 (0.096)	–	0.032	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.181 (0.130)	0.435 (0.078)	0.066	0.055
	$\widehat{\sigma}_{\text{BT},n}$	0.235 (0.131)	0.431 (0.088)	0.065	0.067
$\epsilon = 0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.267 (0.169)	–	1.313	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.292 (0.146)	0.446 (0.074)	0.121	0.115
	$\widehat{\sigma}_{\text{BT},n}$	0.338 (0.131)	0.443 (0.079)	0.099	0.101
$\epsilon = 0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.206 (0.131)	–	2.029	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.407 (0.118)	0.459 (0.063)	0.359	0.354
	$\widehat{\sigma}_{\text{BT},n}$	0.441 (0.091)	0.458 (0.070)	0.194	0.196
$\epsilon = 0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.164 (0.099)	–	2.407	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.401 (0.135)	0.459 (0.064)	0.996	0.353
	$\widehat{\sigma}_{\text{BT},n}$	0.472 (0.069)	0.458 (0.070)	0.310	0.196

computing 500 replications for the leave-one-out classical $CV_{\text{LS,CV}}$ and robust $CV_{\text{ROB,CV}}$ took 102 min using an Intel 2 Quad Q9650 computer with 8 Gb of RAM.

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

	Estimator	Selected bandwidth		$\widehat{\text{ISEL}}$	
		$\widehat{h}_{\text{CV}}^{\log}$	$\widehat{h}_{\text{SCV}}^{\log}$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{CV}}^{\log})$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{SCV}}^{\log})$
$\epsilon = 0$	$\widehat{\sigma}_{\text{RICE},n}$	0.077 (0.071)	–	0.056	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.225 (0.137)	0.216 (0.131)	0.056	0.061
	$\widehat{\sigma}_{\text{BT},n}$	0.243 (0.138)	0.256 (0.130)	0.067	0.062
$\epsilon = 0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.058 (0.079)	–	1.152	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.213 (0.148)	0.246 (0.138)	0.183	0.148
	$\widehat{\sigma}_{\text{BT},n}$	0.248 (0.151)	0.266 (0.141)	0.136	0.123
$\epsilon = 0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.059 (0.059)	–	1.892	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.219 (0.184)	0.270 (0.151)	0.657	0.490
	$\widehat{\sigma}_{\text{BT},n}$	0.171 (0.179)	0.273 (0.161)	0.400	0.262
$\epsilon = 0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.058 (0.062)	–	2.309	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.274 (0.177)	0.316 (0.162)	1.270	1.130
	$\widehat{\sigma}_{\text{BT},n}$	0.150 (0.163)	0.271 (0.161)	0.611	0.340

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

	Estimator	Selected bandwidth		$\widehat{\text{ISEL}}$	
		$\widehat{h}_{\text{CV}}^{\log}$	$\widehat{h}_{\text{SCV}}^{\log}$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{CV}}^{\log})$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{SCV}}^{\log})$
$\epsilon = 0$	$\widehat{\sigma}_{\text{RICE},n}$	0.283 (0.132)	–	0.036	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.087 (0.124)	0.384 (0.124)	0.133	0.052
	$\widehat{\sigma}_{\text{BT},n}$	0.144 (0.169)	0.411 (0.112)	0.164	0.063
$\epsilon = 0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.343 (0.157)	–	1.375	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.118 (0.148)	0.403 (0.113)	0.320	0.116
	$\widehat{\sigma}_{\text{BT},n}$	0.161 (0.175)	0.424 (0.108)	0.242	0.102
$\epsilon = 0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.301 (0.141)	–	2.105	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.131 (0.158)	0.396 (0.130)	0.811	0.387
	$\widehat{\sigma}_{\text{BT},n}$	0.179 (0.184)	0.432 (0.109)	0.417	0.201
$\epsilon = 0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.279 (0.121)	–	2.494	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.134 (0.191)	0.344 (0.172)	1.342	1.103
	$\widehat{\sigma}_{\text{BT},n}$	0.238 (0.224)	0.398 (0.150)	0.631	0.344

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

	Estimator	Selected bandwidth		$\widehat{\text{ISEL}}$	
		$\widehat{h}_{\text{CV}}^{\log}$	$\widehat{h}_{\text{SCV}}^{\log}$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{CV}}^{\log})$	$\widehat{\sigma}_n(\cdot, \widehat{h}_{\text{SCV}}^{\log})$
$\epsilon = 0$	$\widehat{\sigma}_{\text{RICE},n}$	0.081 (0.083)	–	0.054	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.140 (0.159)	0.269 (0.150)	0.118	0.055
	$\widehat{\sigma}_{\text{BT},n}$	0.160 (0.173)	0.303 (0.148)	0.158	0.062
$\epsilon = 0.10$	$\widehat{\sigma}_{\text{RICE},n}$	0.043 (0.059)	–	1.140	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.142 (0.162)	0.269 (0.143)	0.308	0.129
	$\widehat{\sigma}_{\text{BT},n}$	0.164 (0.179)	0.296 (0.144)	0.244	0.110
$\epsilon = 0.20$	$\widehat{\sigma}_{\text{RICE},n}$	0.051 (0.056)	–	1.887	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.138 (0.162)	0.305 (0.149)	0.790	0.419
	$\widehat{\sigma}_{\text{BT},n}$	0.152 (0.174)	0.303 (0.151)	0.436	0.217
$\epsilon = 0.30$	$\widehat{\sigma}_{\text{RICE},n}$	0.054 (0.061)	–	2.308	–
	$\widehat{\sigma}_{\text{MAD},n}$	0.154 (0.176)	0.293 (0.164)	1.370	1.152
	$\widehat{\sigma}_{\text{BT},n}$	0.130 (0.158)	0.322 (0.154)	0.680	0.341

Tables 1–4 show that when the data are not contaminated the robust estimators exhibit a minor loss of efficiency, compared with the classical local Rice.

The behavior of $\widehat{\text{ISEL}}\,(\widehat{\sigma}_{\text{RICE},n})$ shows the lack of robustness of $\widehat{\sigma}_{\text{RICE},n}$ in the presence of outliers. As the percentage of contamination increases, Fig. 4 confirms and explains the results observed in Tables 1–4 regarding the better performance of the robust estimators as the density functions move toward the left of 1. Notice that, in most cases, the values of $\widehat{\text{ISEL}}$ are much smaller than 1 for $\epsilon > 0$.

Table 5
Optimal bandwidths when $n = 100$ for each criteria. Model $g(x) = 2\sin(4\pi x)$, $\sigma(x) = \exp(x)$.

ϵ	M_{LS}	$M_{\mathrm{LS}}^{\log}$	M_{ROB}	$M_{\mathrm{ROB}}^{\log}$	$M_{\mathrm{ROB,S}}$	$M_{\mathrm{ROB,S}}^{\log}$
0	0.21	0.17	0.30	0.21	0.28	0.23
0.05	0.50	0.03	0.38	0.27	0.35	0.25

With respect to the performance of the two robust estimators, under asymmetric contaminations $\widehat{\sigma}_{\mathrm{BT},n}$ is clearly more robust than $\widehat{\sigma}_{\mathrm{MAD},n}$. This fact is also observed for high contamination proportions under the symmetric contamination model, as described in Boente et al. (2010b).

For the studied model, the scale curve is a monotone function and so, the mean values of the selected bandwidths tend to be larger. This implies that more smoothing is needed and the mean values of $\widehat{\mathrm{ISEL}}$ are smaller, compared with those obtained under the model in which $\sigma(x) = 1 + [1 + \sin(10x)]^2$ which are reported in Boente et al. (2010b).

It is also important to remark that all the robust bandwidth selection methods give similar conclusions regarding the performance of the estimators in both models. But in general, the smaller values of the integrated square errors correspond to SCV and $SCV^{\log}$.

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.

6.2. Performance of the asymptotic bandwidths

In this subsection, we compute the "optimal" bandwidths according to the criteria defined in Section 3 (compare with Bianco and Boente, 2007). This study shows that the classical asymptotically optimal bandwidths are very sensitive to outliers while the robust ones are much more stable. These results explain, in part, those in Section 6.1. We define the expected mean estimation errors as follows:

$$M_{\mathrm{LS}}(h) = \mathbb{E}\frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2(x_i) - \widehat{\sigma}^2_{\mathrm{LS},n}(x_i,h)\right)^2 \tag{7}$$

$$M_{\mathrm{LS}}^{\log}(h) = \mathbb{E}\frac{1}{n}\sum_{i=1}^{n}\left(\log(\sigma(x_i)) - \log\left(\widehat{\sigma}_{\mathrm{LS},n}(x_i,h)\right)\right)^2 \tag{8}$$

$$M_{\mathrm{ROB}}(h) = \mathbb{E}s^2(h)\frac{1}{n}\sum_{i=1}^{n}\psi^2\left(\frac{u_i(h)}{s(h)}\right) \tag{9}$$

$$M_{\mathrm{ROB}}^{\log}(h) = \mathbb{E}s^2_{\log}(h)\frac{1}{n}\sum_{i=1}^{n}\psi^2\left(\frac{u_i^{\log}(h)}{s_{\log}(h)}\right) \tag{10}$$

$$M_{\mathrm{ROB,S}}(h) = \mathbb{E}\left[\mu_n^2\left(\{u_i(h)\}_{1\le i\le n}\right) + \tau_n^2\left(\{u_i(h)\}_{1\le i\le n}\right)\right] \tag{11}$$

$$M_{\mathrm{ROB,S}}^{\log}(h) = \mathbb{E}\left[\mu_n^2\left(\{u_i^{\log}(h)\}_{1\le i\le n}\right) + \tau_n^2\left(\{u_i^{\log}(h)\}_{1\le i\le n}\right)\right], \tag{12}$$

where $u_i(h) = \sigma^2(x_i) - \widehat{\sigma}^2_{\mathrm{BT},n}(x_i,h)$, $u_i^{\log}(h) = \log(\sigma(x_i)) - \log\left(\widehat{\sigma}_{\mathrm{BT},n}(x_i,h)\right)$, $s(h)$ and $s_{\log}(h)$ stand for the median of the absolute values $|\sigma^2(x_i) - \widehat{\sigma}^2_{\mathrm{BT},n}(x_i,h)|$ and $\left|\log(\sigma(x_i)) - \log\left(\widehat{\sigma}_{\mathrm{BT},n}(x_i,h)\right)\right|$, respectively, and μ_n and τ_n are the median and a tau-scale. These are approximations to the corresponding "integrated mean estimation error" which in practice are estimated using cross-validation.

The expected values are approximated by simulating independent samples of size $n = 100$ (as in Section 6.1, i.e., with $g(x) = 2\,\sin(4\pi x)$, $\sigma(x) = \exp(x)$, $x_i = i/(n+1)$) and averaging over 100 replications. The "optimal" bandwidths $h_{n,\mathrm{OPT}}$ are then obtained by minimizing the above mean estimation errors over a grid of points.

We consider two types of errors: standard normal and asymmetrically contaminated standard normal ($G(y) = 0.95\,\Phi(y) + 0.05\,H(y)$ with $H = \mathrm{N}(10, 0.1)$). The optimal bandwidth is obtained by inspection over the grid $i/(n/2)$, $3 \le i \le n/2$. The obtained bandwidths are given in Table 5. Fig. 5 gives plots of the estimated errors $M_{\mathrm{LS}}(h)$, $M_{\mathrm{LS}}^{\log}(h)$, $M_{\mathrm{ROB}}(h)$, $M_{\mathrm{ROB}}^{\log}(h)$, $M_{\mathrm{ROB,S}}(h)$ and $M_{\mathrm{ROB,S}}^{\log}(h)$. The plots corresponding to contaminated samples using $M_{\mathrm{LS}}(h)$ and $M_{\mathrm{LS}}^{\log}(h)$ are given in separate plots due to the scale difference. As expected the non-robust measures $M_{\mathrm{LS}}(h)$ and $M_{\mathrm{LS}}^{\log}(h)$ are strongly influenced by contamination, while the robust measures combined with robust estimators are much more stable. The best performance is achieved when measuring deviances on the log scale and using the split bias-variance procedure. Note that these findings are consistent with the conclusions of the Monte Carlo study described in Section 6.1.

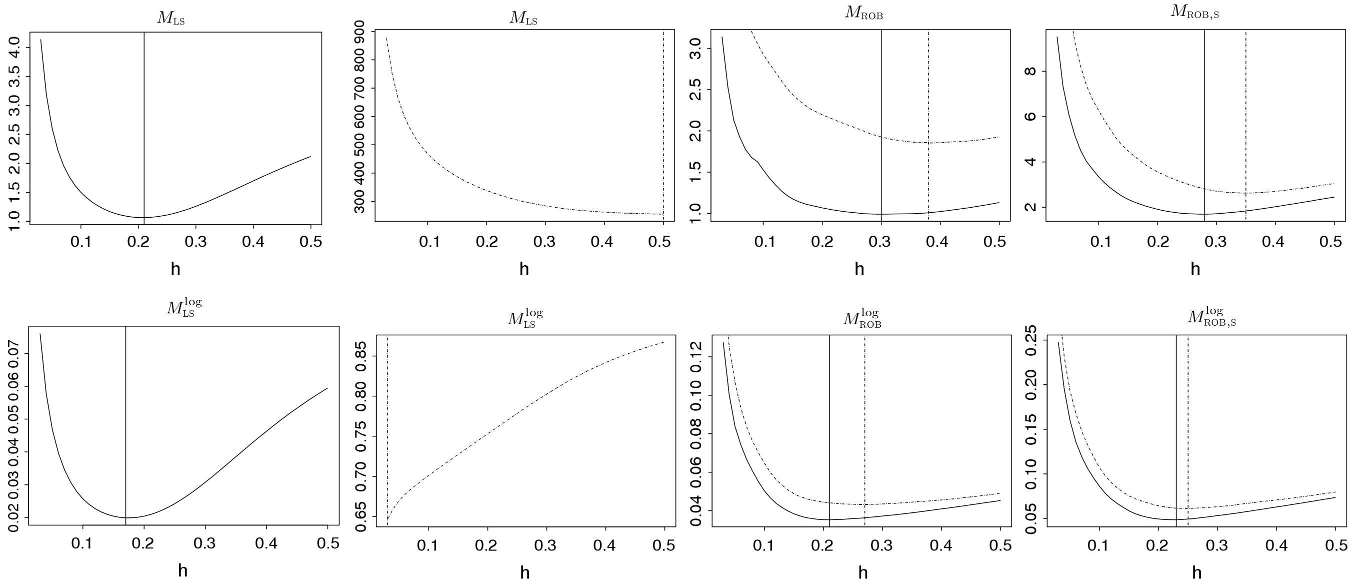

Fig. 5. In solid and dashed lines, (— · —) are plotted with the results over non-contaminated and contaminated samples, respectively. The vertical lines show the point where the minimum value is attained.

Table 6
Mean of the ratio R between the data-driven and $n^{-\alpha}$. Model $g(x) = 2\sin(4\pi x)$, $\sigma(x) = \exp(x)$.

n	$\alpha = -1/3$					
	R_{LS}	R_{ROB}	$R_{ROB,S}$	R_{LS}^{log}	R_{ROB}^{log}	$R_{ROB,S}^{log}$
100	0.882	1.091	2.001	0.357	1.128	1.188
200	0.904	1.314	2.694	0.252	1.376	1.327
500	1.028	1.530	3.865	0.180	1.641	1.639
1000	1.032	2.095	4.923	0.119	1.734	1.857
2000	1.034	2.490	6.253	0.100	1.672	2.159
n	$\alpha = -1/5$					
	R_{LS}	R_{ROB}	$R_{ROB,S}$	R_{LS}^{log}	R_{ROB}^{log}	$R_{ROB,S}^{log}$
100	0.477	0.590	1.083	0.193	0.610	0.643
200	0.446	0.648	1.329	0.124	0.679	0.655
500	0.449	0.668	1.688	0.079	0.716	0.716
1000	0.411	0.834	1.960	0.047	0.690	0.739
2000	0.375	0.904	2.270	0.036	0.607	0.784

6.3. *Asymptotic performance of the data-driven bandwidths*

In this subsection, we compute the cross-validation data-driven bandwidths for samples of sizes $n = 100, 200, 500, 1000$ and 2000 to numerically investigate their convergence rates under the central Gaussian model. That is, we consider the model described in Section 6.1 with $\epsilon = 0$.

To simplify the notations, the ratio between the cross-validation bandwidths $\widehat{h}_{LS,CV}$, $\widehat{h}^{log}_{LS,CV}$, $\widehat{h}_{BT,CV}$, $\widehat{h}^{log}_{BT,CV}$, $\widehat{h}_{BT,SCV}$ and $\widehat{h}^{log}_{BT,SCV}$ and $n^{-\alpha}$ for $\alpha = 1/3$ and $1/5$ are denoted R_{LS}, R_{ROB}, $R_{ROB,S}$, R^{log}_{LS}, R^{log}_{ROB} and $R^{log}_{ROB,S}$, respectively. Table 6 reports the averages over 100 replications.

The results in Table 6 suggest that the assumptions in Section 5 are valid for the classical bandwidth $\widehat{h}_{LS,CV}$ and for all the considered robust data-driven bandwidths. In fact, for the classical cross-validation bandwidth $\widehat{h}_{LS,CV}$, the order seems to be $n^{-1/3}$. But note the bad performance of $\widehat{h}^{log}_{LS,CV}$ which leads to very small bandwidth values and so, to undersmoothing as mentioned in Section 4. The convergence order of the robust cross-validation bandwidths appears to be $n^{-1/5}$.

In summary, the numerical results obtained in this subsection suggest that the assumptions of Theorem 5.2 are satisfied by the robust scale estimators computed with the robust data-driven bandwidths introduced in this paper. Therefore, these data driven estimators are consistent and asymptotically normally distributed.

7. Concluding remarks

Robust estimation of the scale function is an important research problem. There are classical (non-robust) and robust proposals for the estimation of the scale function based on differences of the response variable. But the critical problem of estimation of the bandwidth parameter has been less studied.

We have shown that the asymptotic behavior of the local robust data-driven M-estimators for the scale function based on successive differences is the same as that of the related estimators obtained with a fixed bandwidth h_n provided $\widehat{h}_n/h_n \xrightarrow{p} 1$. Therefore, under mild regularity conditions, the corresponding robust kernel-based scale estimators based on differences are strongly consistent and asymptotically normal.

We proposed several robust cross-validation procedures to estimate the bandwidth parameter. The performance of the classical and robust bandwidth selection methods as well as the behavior of the corresponding estimators based on these bandwidths was compared using Monte Carlo simulation under the central model and different contamination models. We also illustrated their performances in a real data example.

We have shown that the robust approaches perform better than their non-robust counterparts. We have also found that using the log scale and splitting the cross-validation error into its bias and variance components lead to the best performances in our numerical numerical studies and simulations.

Acknowledgments

We wish to thank two anonymous referees for valuable comments which led to an improved version of the original paper. This research was partially supported by Grants X-018 and 20020100100276 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.

Appendix

Proof of Theorem 5.1. 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

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

$$\lambda(x, s) = \mathbb{E}\left[\chi\left(\frac{\sigma(x)U_1^*}{as}\right)\right] - b.$$

Theorem 4.1 in Boente et al. (2010a) 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 $\widehat{\sigma}_{M,n}(x, \widehat{h}_n) \xrightarrow{a.s.} S(G_x)$.

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

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

$$S_{2,n} = (n\widehat{h}_n)^{-1} \sum_{i=1}^{n} \left\{L\left(\frac{x - x_i}{\widehat{h}_n}\right) - L\left(\frac{x - x_i}{h_n}\right)\right\} Z_i.$$

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 $(nh_n)^{-1}\sum_{i=1}^n |L((x - x_i)/h_n)| \to \int |L(u)|\,du$, (A.2) follows easily. To obtain (A.3), write

$$S_{2,n} = (nh_n)^{-1} \sum_{i=1}^{n} L_1\left(\frac{x - x_i}{\xi_n}\right) \left[\frac{h_n}{\xi_n}\right] \left[\frac{h_n}{\widehat{h}_n} - 1\right] Z_i,$$

where $L_1(u) = uL'(u)$ 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

$$\left|S_{2,n}\right| \le 2\left|\frac{h_n}{\widehat{h}_n} - 1\right| \|\chi\|_\infty (nh_n)^{-1}\sum_{i=1}^n \left|L_1\left(\frac{x-x_i}{\xi_n}\right)\right|,$$

it is enough to show that $\limsup |A_n| < \infty$ where $A_n = (\xi_n/h_n)\, C_n$ and $C_n = (n\xi_n)^{-1}\sum_{i=1}^n |L_1\left((x-x_i)/\xi_n\right)|$. 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 $C_n \to \int |L(u)|\, du$ which concludes the proof. □

Proof of Theorem 5.2. Let $\{Y_i^*\}_{i\ge1}$, $\{U_i^*\}_{i\ge1}$ and $S_x = S(G_x)$ be as in the proof of Theorem 5.1. Also, let $S_n(h) = (nh_n)^{1/2}\lambda_{n,b}(x, S_x, h)$ with

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

and

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

Using a second order Taylor's expansion, we obtain

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

where

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

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

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

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

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

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 $\widehat{\sigma}_{M,n}(x, \widehat{h}_n) - S_x \xrightarrow{p} 0$ and considering that (A.12) in Boente et al. (2010a) 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

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

and that analogous arguments to those considered above lead to

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

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

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 which 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

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

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 for any $n \geq \max(n_0, n_1)$

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

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. (2010a), it is easy to show that

$$\sup_{\tau\in[r,s]} \left|(nh_n)^{1/2}\mathbb{E}\left[\lambda_{n,b}(x, S_x, \tau h_n)\right] - \beta_1 \frac{S_x}{\sigma(x)} \frac{\upsilon_2}{\tau^{1/2}}\right| \to 0,$$

i.e., $\gamma_n(\tau) \to \gamma(\tau) = \beta_1 S_x \upsilon_2 (\sigma(x)\tau^{1/2})^{-1}$ 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 $\tau_1, \ldots \tau_k \in [r,s]$, $\left(U_{\tau_1}, \ldots, U_{\tau_k}\right)$ 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 $\mathbf{a} = (a_1, \ldots, a_k)^{\mathrm{T}} \in \mathbb{R}^k$, $W_n = \sum_{j=1}^k a_j U_n(\tau_j)$ converge to a normal distribution. Note that

$$U_n(\tau) = (nh_n)^{1/2} \frac{1}{n\tau h_n} \sum_{i=1}^n L\left(\frac{x - x_i}{\tau h_n}\right) Z_i$$

$$W_n = (nh_n)^{-1/2} \sum_{i=1}^n L^*\left(\frac{x - x_i}{h_n}\right) Z_i$$

with $Z_i = \chi\left(Y_i^*/(aS_x)\right) - \mathbb{E}\chi\left(Y_i^*/(aS_x)\right)$ and $L^*(u) = \sum_{j=1}^k (a_j/\tau_j) L\left(u/\tau_j\right)$. 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. (2010a).

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 $\mathrm{VAR}(Z_i) \leq k_1$ and $\mathrm{cov}(Z_i, Z_{i+1}) \leq k_1$ for any $i \geq 1$. Hence $\mathbb{E}\left[U_n(\tau_2) - U_n(\tau_1)\right]^2 \leq H_{1,n} + H_{2,n}$, where

$$H_{1,n} = k_1 \frac{1}{nh_n} \sum_{i=1}^n \left(\frac{1}{\tau_2} L\left(\frac{x - x_i}{\tau_2 h_n}\right) - \frac{1}{\tau_1} L\left(\frac{x - x_i}{\tau_1 h_n}\right)\right)^2$$

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

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

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

$$T_{2,n} = 2k_1 \tau_1 \left(\frac{1}{\tau_2} - \frac{1}{\tau_1}\right)^2 \frac{1}{n\tau_1 h_n} \sum_{i=1}^n L^2\left(\frac{x - x_i}{\tau_1 h_n}\right) = 2k_1 \tau_1 \left(\frac{1}{\tau_2} - \frac{1}{\tau_1}\right)^2 R_{2,n}$$

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 $n \geq 1$, $R_{1,n} \leq B$, with B a fixed constant. If we note that for any $n \geq 1$, $R_{2,n} \leq C$ with C a fixed constant, we get easily that, $T_{1,n} \leq 2k_1(s^2/(r^4 B))\,(\tau_2 - \tau_1)^2$ and $T_{2,n} \leq 2k_1(s/(r^2 C))\,(\tau_2 - \tau_1)^2$. So, $H_{1,n} \leq c_1\,(\tau_2 - \tau_1)^2$ 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 $H_{2,n} \leq c_2\,(\tau_2 - \tau_1)^2$. Hence (iii) follows with $c = c_1 + c_2$. □

References

Beaton, A., Tukey, J., 1974. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. Technometrics 16, 147–185.

Bianco, A., Boente, G., 2007. Robust estimators under a semiparametric partly linear autoregression model: Asymptotic behavior and bandwidth selection. Journal of Time Series Analysis 28, 274–306.

Boente, G., Fraiman, R., 1989. Robust nonparametric regression estimation for dependent observations. The Annals of Statistics 17, 1242–1256.

Boente, G., Fraiman, R., 1995. Asymptotic distribution of data-driven smoothers in density and regression estimation under dependence. Canadian Journal of Statistics 23, 383–397.

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

Boente, G., 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.

Boente, G., Ruiz, M., Zamar, R., 2010a. On a robust local estimator for the scale function in heteroscedastic nonparametric regression. Statistics & Probability Letters 80, 1185–1195.

Boente, G., Ruiz, M., Zamar, R., 2010b. Bandwidth choice for robust nonparametric scale function estimation. URL: www.ic.fcen.uba.ar/preprints/BRZ_TR.pdf.

Brown, L., Levine, M., 2007. Variance estimation in nonparametric regression via the difference sequence method. The Annals of Statistics 35, 2219–2232.

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

Carroll, R., Ruppert, D., 1988. Transformation and Weighting in Regression. Chapman & Hall.

Daniel, C., Wood, F., 1980. Fitting Equations to Data: Computer Analysis of Multifactor Data. Wiley, New York.

Dette, H., Hetzler, B., 2008. A simple test for the parametric form of the variance function in nonparametric regression. The Annals of the Institute of Statistical Mathematics 61, 861–886.

Dette, H., Marchlewski, M., 2008. A test for the parametric form of the variance function in a partial linear regression model. Journal of Statistical Planning and Inference 138, 3005–3021.

Dette, H., Marchlewski, M., 2010. A robust test for homoscedasticity in nonparametric regression. Journal of Nonparametric Statistics 22, 723–736.

Dette, H., Munk, A., 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.

Fan, J., Yao, Q., 1998. Efficient estimation of conditional variance functions in stochastic regression. Biometrika 85, 645–660.

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

Ghement, I., Ruiz, M., Zamar, R., 2008. Robust estimation of error scale in nonparametric regression models. Journal of Statistical Planning and Inference 138, 3200–3216.

Hall, P., Kay, J., Titterington, D., 1990. Asymptotically optimal difference-based estimation of variance in nonparametric regression. Biometrika 77, 521–528.

Härdle, W., 1990. Applied Nonparametric Regression. Cambridge University Press.

Härdle, W., Gasser, T., 1984. Robust nonparametric function fitting. Journal of the Royal Statistical Society, Series B 46, 42–51.

Härdle, W., Müller, M., Sperlich, S., Werwatz, A., 2004. Nonparametric and Semiparametric Models. Springer.

Härdle, W., Tsybakov, A., 1988. Robust nonparametric regression with simultaneous scale curve estimation. The Annals of Statistics 25, 443–456.

Leung, D., 2005. Cross-validation in nonparametric regression with outliers. The Annals of Statistics 33, 2291–2310.

Leung, D., Marriott, F., Wu, E., 1993. Bandwidth selection in robust smoothing. Journal of Nonparametric Statistics 2, 333–339.

Levine, M., 2003. Variance estimation for nonparametric regression and its applications. Ph.D. Dissertation, University of Pennsylvania.

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.

Maronna, R., Martin, D., Yohai, V., 2006. Robust Statistics: Theory and Methods. John Wiley & Sons.

Müller, H., Stadtmüller, U., 1987. Estimation of heteroscedasticity in regression analysis. The Annals of Statistics 15, 610–625.

Rice, J., 1984. Bandwidth choice for nonparametric regression. The Annals of Statistics 12, 1215–1230.

Ullah, A., 1985. Specification analysis of econometric models. Journal of Quantitative Economics 2, 187–209.

Wang, F., Scott, D., 1994. The L1 method for robust nonparametric regression. Journal of American Statistical Association 89, 65–76.