Journal of Multivariate Analysis 154 (2017) 59–84

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Robust estimators in semi-functional partial linear regression models

Graciela Boente [a,b,*], Alejandra Vahnovan [c,d]

[a] Departamento de Matemáticas, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 1, 1428, Buenos Aires, Argentina

[b] IMAS, CONICET, Ciudad Universitaria, Pabellón 1, 1428, Buenos Aires, Argentina

[c] Departamento de Matemáticas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Calle 50 y 115, La Plata, Argentina

[d] CONICET, Godoy Cruz 2290. 1425, Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 20 February 2016
Available online 31 October 2016

AMS subject classifications:
62G35
62G08
62G05

Keywords:
Functional data
Kernel smoothers
Partial linear models
Robust estimation

ABSTRACT

Partial linear models have been adapted to deal with functional covariates to capture both the advantages of a semi-linear modelling and those of nonparametric modelling for functional data. It is easy to see that the estimation procedures for these models are highly sensitive to the presence of even a small proportion of outliers in the data. To solve the problem of atypical observations when the covariates of the nonparametric component are functional, robust estimates for the regression parameter and regression operator are introduced. Consistency results of the robust estimators and the asymptotic distribution of the regression parameter estimator are studied. The reported numerical experiments show that the resulting estimators have good robustness properties. The benefits of considering robust estimators is also illustrated on a real data set where the robust fit reveals the presence of influential outliers.

1. Introduction

Nonparametric regression models make possible the definition of general regression estimates when no structural assumptions are made on the relation between covariates and a scalar response variable. More precisely, these models assume that the practitioner has independent observations (Y_i, X_i), $1 \leq i \leq n$, where $Y_i = m(X_i) + \epsilon_i$ and the errors ϵ_i are independent and independent of X_i with $\mathrm{E}(\epsilon_i) = 0$. In many applications, the covariates X_i can be seen as functions recorded over a period of time instead of finite-dimensional vectors. For that reason, these variables can be viewed as realizations of a stochastic process, often assumed to be in the $L^2(\mathcal{I})$ with $\mathcal{I}$ a bounded interval and are usually called functional variables in the literature. In this general framework, statistical models adapted to infinite-dimensional data have been studied. We refer to Ramsay and Silverman [43,44], Ferraty and Vieu [27], Ferraty and Romain [25] and Horváth and Kokoszka [36] for a description of different procedures for functional data and their properties. For a summary of recent advances in infinite dimensional statistics see Cuevas [20] and Goia and Vieu [31]. In particular, for functional covariates, kernel-type estimators of the regression operator m have been considered by Ferraty and Vieu [27] and Ferraty et al. [23,22], among others. On the other hand, in some situations the researcher has both a vector of finite dimension $\mathbf{Z}$ and a functional random element X that may have effect on the responses. These data, usually called mixed or hybrid data, were described in Ramsay and

* Correspondence to: Departamento de Matemáticas, FCEyN, UBA, Ciudad Universitaria, Pabellón 1, Buenos Aires, C1428EHA, Argentina.
E-mail addresses: gboente@dm.uba.ar (G. Boente), avahnovan@gmail.com (A. Vahnovan).

http://dx.doi.org/10.1016/j.jmva.2016.10.005

Silverman [44]. Linear models for mixed data were introduced by Zhang et al. [55], while kernel methods in nonparametric regression models in this setting are given in Shang [48].

Semiparametric models have been introduced to provide an intermediate approach between a fully nonparametric model and a less flexible linear one. When the covariates are functional, a possible model is the functional single-index model, introduced by Ferraty et al. [24] to avoid the curse of dimensionality, in which the regression operator is written as an unknown function η of an unknown projection of the covariates, i.e., $m(X) = \eta(\langle\alpha, X\rangle)$ where both η and α must be estimated. Estimation methods combining spline basis or a least square approach and smoothing methods have been proposed, among others, by Ait-Saïdi et al. [1] and Chen et al. [19]. Some extensions of this model have been considered by Ferraty et al. [21] and Goia and Vieu [30]. Up to our knowledge, single-index models for hybrid covariates have not been studied yet.

A different approach was studied in Aneiros-Pérez and Vieu [3] who proposed a semi-functional partial linear regression model that relates the functional predictor with the response nonparametrically, while the finite-dimensional covariate is included in the model through a linear regression. More precisely, under a semi-functional partial linear regression model, the independent observations $(Y_i, \mathbf{Z}_i^\top, X_i)^\top$, $1 \le i \le n$, are such that $Y_i \in \mathbb{R}$, $\mathbf{Z}_i \in \mathbb{R}^p$ and $X_i \in \mathcal{H}$ are functional variables satisfying the semi-linear model

$$Y_i = \boldsymbol{\beta}^\top \mathbf{Z}_i + g(X_i) + \epsilon_i, \tag{1}$$

where the errors ϵ_i are independent and independent of $(\mathbf{Z}_i^\top, X_i)^\top$, $\boldsymbol{\beta} \in \mathbb{R}^p$, $g : \mathcal{H} \to \mathbb{R}$ is a smooth operator and $(\mathcal{H}, d)$ is a semi-metric space. Note that the semi-functional partial linear model (1) combines the advantages of a functional nonparametric component with the linear effect of some additional real-valued explanatory variables that may be available to the practitioner. One of the differences between model (1) and the single-index model mentioned above is that the latter assumes that the response dependence on the covariates is given, up to the link function η, through a linear projection, while the nonparametric structure of the semi-functional partial linear model allows for more general structures. Furthermore, under a single-index model the covariates lie in a Hilbert space, while model (1) is more flexible since in this setting $\mathcal{H}$ is a semi-metric space and it also includes the effect of an additional finite-dimensional covariate. On the other hand, under model (1) the rates of convergence for the estimators of the regression operator g will be slow as in nonparametric regression models.

Several authors have studied proposals for estimating the unknown quantities under model (1) in the finite dimensional setting, i.e., when $X_i \in \mathbb{R}$. An extensive description of different results obtained in partial linear regression models for scalar covariates can be found, for instance, in Härdle et al. [33]. For functional covariates X_i, Aneiros-Pérez and Vieu [3] combine the ideas beyond functional nonparametric regression and those used in the finite-dimensional setting to provide estimators of the unknown quantities $\boldsymbol{\beta}$ and g, when $\mathrm{E}(\epsilon) = 0$ and $\mathrm{E}\epsilon^2 < \infty$. Besides, through the analysis of the spectrometric data set described in Ferraty and Vieu [27], Aneiros-Pérez and Vieu [3] showed the advantage of the semi-functional partial linear model (1) over the functional linear regression, functional nonparametric regression and additive semi-functional models. Since then, the semi-functional partial linear model has been widely studied and we can mention among others Aneiros-Pérez and Vieu [4–6], Shin [50], Shang [49] and Aneiros-Pérez et al. [2].

This work aims to contribute to the semi-functional partial linear regression model by introducing methods that will not be distorted by the presence of some atypical observations. For that reason, throughout this paper we will not assume any moment condition on the errors, instead we will assume that the errors have a symmetric distribution $G(\cdot/\sigma_0)$, i.e., the distribution G has scale 1 to identify the scale parameter σ_0. In general, it is crucial for the practitioner to detect outliers in the data to avoid misleading inferences. Furthermore, the detected atypical data may contain important information that would redound in a deeper comprehension of the whole phenomenon under study and for that reason, it is important to identify them. In the finite-dimensional setting, the literature on robust procedures or outlier detection methods is wide. In particular, it is well known that, both in linear regression as in non-parametric regression, least squares estimators can be seriously affected by anomalous data. The same conclusion is valid for partially linear models. In particular, large values of the response variable Y_i can cause a peak on the estimates of the smooth function g in the neighbourhood of X_i and large values of the response variable Y_i combined with high leverage points $\mathbf{Z}_i$ produce, as in linear regression, breakdown of the classical estimates of the regression parameter $\boldsymbol{\beta}$. To overcome the lack of robustness of the classical least squares approach, Bianco and Boente [9] considered a robust kernel-based three-step procedure under a finite-dimensional partially linear model, i.e., when $X_i \in \mathbb{R}$, while Henry and Rodriguez [35] extended these estimators to deal with the situation in which $X_i \in \mathcal{M}$, with $\mathcal{M}$ a Riemannian manifold. However, the study of robust estimators for functional data is rather scarce. Robust proposals for nonparametric functional regression estimation have been considered by Azzedine et al. [7] who studied nonparametric robust estimation methods based on the local M-estimators assuming that the scale is known. Besides, Boente and Vahnovan [16] proposed robust equivariant M-regression estimators.

The benefits of the semi-functional partial linear model (1) for situations in which the response dependence can be explained by a nonparametric structure and the linear effect of exogenous variables, together with the possible presence of atypical data in the sample explain our interest in providing reliable estimation procedures in this particular framework. Up to our knowledge, resistant procedures for the semi-functional partial linear model (1) have not been considered yet. Hence, in this paper we aim to introduce a class of robust procedures under the semi-functional partial linear model (1) combining robust regression estimators with the robust conditional location functional defined in Boente and Fraiman [12]. In this sense, our proposal extends to the functional setting the robust proposal in Bianco and Boente [9] using the robust nonparametric estimators introduced in Boente and Vahnovan [16]. As in Bianco and Boente [9], we will not require moments of the errors and we will replace the assumption that $\mathrm{E}(\epsilon) = 0$ by the requirement that the errors have a symmetric distribution. It is worth mentioning that the procedures introduced for functional single-index model and mentioned above are not competitors for the proposals given here, since they correspond to a different model setting and do not include mixed covariates. Besides, these methods may also be sensitive to outliers since they are based on linear smoothers and least squares methods. Further research on resistant procedures for single-index models with mixed covariates is needed, but is beyond the scope of this paper. Instead, a primary focus of this paper is to provide a rigorous theoretical foundation for the approach of robust estimation under semi-functional partial linear regression models. In particular, we establish the strong consistency of the proposed estimators and the asymptotic distribution of the estimators of the regression parameter $\boldsymbol{\beta}$, under general conditions.

The paper is organized as follows. In Section 2 we review the classical approach and we introduce the robust estimators. Consistency results are given in Section 3.1, while the asymptotic normality of the regression parameter estimators is studied in Section 3.2. The results of a numerical experiments conducted to evaluate the performance of the classical and robust procedures are described in Section 4, while the analysis of a real data set, which shows the advantage of using a robust procedure, is presented in Section 5. All proofs are relegated to the Appendix.

2. The robust estimators

Let $(\mathcal{H}, d)$ be a semi-metric functional space, that is, d satisfies the metric properties but $d(x, y) = 0$ does not imply $x = y$. Denote $(Y, \mathbf{Z}^\top, X)^\top \in \mathbb{R}^{p+1} \times \mathcal{H}$ a random element with the same distribution as $(Y_i, \mathbf{Z}_i^\top, X_i)^\top$, i.e., $Y = \boldsymbol{\beta}^\top \mathbf{Z} + g(X) + \epsilon$, with ϵ independent of $(\mathbf{Z}^\top, X)^\top$. When $\mathrm{E}(\epsilon) = 0$ and $\mathrm{E}(\epsilon^2) < \infty$, Aneiros-Pérez and Vieu [3] proposed a least squares approach to estimate $\boldsymbol{\beta}$, noting that $Y - \rho_0(X) = \boldsymbol{\beta}^\top(\mathbf{Z} - \boldsymbol{\rho}(X)) + \epsilon$, with $\rho_0(x) = \mathrm{E}(Y|X = x)$, $\rho_j(x) = \mathrm{E}(Z_j|X = x)$ and $\boldsymbol{\rho}(x) = (\rho_1(x), \dots, \rho_p(x))^\top$. To be more precise, Aneiros-Pérez and Vieu [3] inserted kernel estimators of $\rho_0(x)$ and $\boldsymbol{\rho}(x)$, $\widehat{\rho}_0(x)$ and $\widehat{\boldsymbol{\rho}}(x)$, prior to the estimation of the regression parameter. The functional Nadaraya–Watson kernel estimates of $\rho_0(x)$ and $\rho_j(x)$ are defined, respectively, as $\widehat{\rho}_0(x) = \sum_{i=1}^n w_i(x, h) Y_i$ and $\widehat{\rho}_j(x) = \sum_{i=1}^n w_i(x, h) Z_{ij}$, where the weights $w_i(x, h)$ are given by

$$w_i(x, h) = \frac{K\left(\frac{d(x,X_i)}{h}\right)}{\sum_{i=1}^n K\left(\frac{d(x,X_i)}{h}\right)}, \tag{2}$$

with K a kernel function, i.e., a nonnegative integrable function on $\mathbb{R}$ and h the bandwidth parameter. The least squares estimator of $\boldsymbol{\beta}$ introduced in Aneiros-Pérez and Vieu [3] is defined as the minimizer $\widehat{\boldsymbol{\beta}}$ of $\sum_{i=1}^n \left\{Y_i - \widehat{\rho}_0(X_i) - \mathbf{b}^\top(\mathbf{Z}_i - \widehat{\boldsymbol{\rho}}(X_i))\right\}^2$, while the estimator of g equals $\widehat{g}(x) = \widehat{\rho}_0(x) - \widehat{\boldsymbol{\beta}}^\top \widehat{\boldsymbol{\rho}}(x)$.

As in the purely parametric setting, the regression estimators of $\boldsymbol{\beta}$ are sensitive to the presence of outliers both in the responses and in the covariates $\mathbf{Z}$. This sensitivity is inherited by the nonparametric component estimator which is based on a local average of the response variables. In particular, as mentioned in the Introduction, atypical responses located in the neighbourhood of the point x will have a large influence. For that reason, resistant procedures are needed to provide reliable inferences.

Let $\psi : \mathbb{R} \to \mathbb{R}$ be a continuous, odd and bounded score function. Furthermore, denote as $F_0(y|X = x)$ and $F_j(z|X = x)$ the conditional distributions of $Y|X = x$ and $Z_j|X = x$, respectively and as $s_j(x)$ the MAD of $F_j(z|X = x)$, $0 \le j \le p$. Let $\phi_j(x)$, $0 \le j \le p$ be the robust conditional location functional introduced in Boente and Fraiman [12]. More precisely, for each $x \in \mathcal{H}$ and $0 \le j \le p$, let $\phi_j(x)$ be the solution of $\lambda_j(x, \phi_j(x), s_j(x)) = 0$ where

$$\lambda_0(x, a, \sigma) = \mathrm{E}\left\{\psi\left(\frac{Y - a}{\sigma}\right)\Big|X = x\right\} \qquad \lambda_j(x, a, \sigma) = \mathrm{E}\left\{\psi\left(\frac{Z_j - a}{\sigma}\right)\Big|X = x\right\}, \quad 1 \le j \le p. \tag{3}$$

Robust estimators of $\phi_j(x)$, $0 \le j \le p$, can be obtained by plugging into (3) estimators of the conditional distribution of $Y|X = x$ and $Z_j|X = x$. In particular, the empirical conditional distribution functions kernel estimators $\widehat{F}_0(y|X = x)$ and $\widehat{F}_j(z|X = x)$ are defined as

$$\widehat{F}_0(y|X = x) = \sum_{i=1}^n w_i(x, h) \mathbf{1}_{(-\infty, y]}(Y_i), \qquad \widehat{F}_j(z|X = x) = \sum_{i=1}^n w_i(x, h) \mathbf{1}_{(-\infty, z]}(Z_{ij}) \quad 1 \le j \le p,$$

with $w_i(x, h)$ the kernel weights given in (2). Hence, the local M-estimators $\widehat{\phi}_{j,\mathrm{M}}(x)$, $0 \le j \le p$, which provide robust estimators of $\phi_j(x)$, $0 \le j \le p$, are given as the solution of $\widehat{\lambda}_j(x, a, \widehat{s}_j(x)) = 0$, where $\widehat{s}_j(x)$ stands for a robust estimator of the conditional scale such as the MAD of $\widehat{F}_j(\cdot|X = x)$ and

$$\widehat{\lambda}_0(x, a, \sigma) = \sum_{i=1}^n w_i(x, h) \psi\left(\frac{Y_i - a}{\sigma}\right) \qquad \widehat{\lambda}_j(x, a, \sigma) = \sum_{i=1}^n w_i(x, h) \psi\left(\frac{Z_{ij} - a}{\sigma}\right), \quad 1 \le j \le p. \tag{4}$$

In particular, when $\psi(t) = \text{sgn}(t)$ the target is the conditional median which can be estimated as the median of $\widehat{F}_j(\cdot|X = x)$, in which case no scale estimator is needed. Other possible choices for the score function ψ are the Huber or the bisquare ψ-function. Consistency results and strong convergence rates for the local median and the local M-estimators were obtained in Boente and Vahnovan [16]. It is worth noting that, when $\psi(t) = t$ and the expectation involved exists, $\phi_0 = \rho_0$ and $\boldsymbol{\phi} = \boldsymbol{\rho}$.

To define robust estimators of the regression parameter $\boldsymbol{\beta}$ under the semi-functional partial linear regression model (1) we combine robust smoothers and robust regression estimators as follows:

Step 1: Estimate $\phi_0(x)$ and $\phi_j(x)$ through a robust functional smoothing, as the local medians or local M-type estimates. Let $\widehat{\phi}_0(x)$ and $\widehat{\phi}_j(x)$ stand for the obtained estimates and $\widehat{\boldsymbol{\phi}}(x) = (\widehat{\phi}_1(x), \dots, \widehat{\phi}_p(x))^\top$.

Step 2: Estimate the regression parameter by applying any robust regression estimate to the residuals $\widehat{r}_i = Y_i - \widehat{\phi}_0(X_i)$ and $\widehat{\mathbf{u}}_i = \mathbf{Z}_i - \widehat{\boldsymbol{\phi}}(X_i)$ (see [41] for a description on robust regression estimators). Denote $\widehat{\boldsymbol{\beta}}$ the obtained estimator.

To estimate the regression operator g, we may proceed as in Bianco and Boente [9] and define the regression operator estimator as $\widehat{g}(x) = \widehat{\phi}_0(x) - \widehat{\boldsymbol{\beta}}^\top \widehat{\boldsymbol{\phi}}(x)$. Alternatively, as in Bianco et al. [11], one may robustly smooth the residuals $Y_i - \widehat{\boldsymbol{\beta}}^\top \mathbf{Z}_i$. More precisely, we can define an estimator of $g(x)$, $\widehat{\widehat{g}}(x)$, as the solution of $\widehat{\widehat{\lambda}}(x, a, \widehat{\widehat{s}}(x)) = 0$, where

$$\widehat{\widehat{\lambda}}(x, a, \sigma) = \sum_{i=1}^{n} K\left(\frac{d(x, X_i)}{\widetilde{h}}\right) \psi\left(\frac{Y_i - \widehat{\boldsymbol{\beta}}^\top \mathbf{Z}_i - a}{\sigma}\right), \tag{5}$$

and $\widehat{\widehat{s}}(x)$ is a robust scale estimate. A possible choice for the scale $\widehat{\widehat{s}}(x)$ is the MAD of the conditional residual empirical distribution $\widehat{\widehat{F}}(u|X = x) = \sum_{i=1}^n w_i(x, \widetilde{h}) \mathbf{1}_{(-\infty,u]}\left(Y_i - \widehat{\boldsymbol{\beta}}^\top \mathbf{Z}_i\right)$, where $\widetilde{h}$ stands for the smoothing parameter. Indeed, as the residuals $Y_i - \widehat{\boldsymbol{\beta}}^\top \mathbf{Z}_i$ have less variability than the original variables Y_i, it may be preferred to use a different smoothing parameter $\widetilde{h}$ than the one used in **Step 1**. The estimators $\widehat{\widehat{g}}$ are those considered in the simulation study reported in Section 4.

As in the finite-dimensional setting, to ensure that the model is identifiable, we will assume that the vector $\mathbf{1}_n$ is not in the space spanned by the column vectors of $\mathbf{Z}$. Identifiability means that if $\boldsymbol{\beta}_1^\top \mathbf{Z}_i + g_1(X_i) = \boldsymbol{\beta}_2^\top \mathbf{Z}_i + g_2(X_i)$ for $1 \le i \le n$ then, $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ and $g_1 = g_2$. As mentioned by several authors, identifiability implies that only "slope" coefficients may be estimated. Moreover, the assumptions to be stated below avoid any linear combination of the components of $\mathbf{Z}$ from being a function of X (see Robinson [45]), since these models are more nonparametric than semiparametric.

3. Asymptotic results

From now on, $\xrightarrow{a.co.}$, $\xrightarrow{a.s.}$ and $\xrightarrow{p}$ stand for almost complete convergence, almost sure convergence and convergence in probability, respectively, while $\xrightarrow{D}$ denotes convergence in distribution. On the other hand, for a given distribution function H or probability measure P over $\mathbb{R}^q$, $\mathbf{z} \sim H$ and $\mathbf{z} \sim P$ indicate that the vector $\mathbf{z} \in \mathbb{R}^q$ has distribution H or that the related probability measure is P, respectively.

As mentioned above, the observations to be considered are such that the covariates X belong to a semi-metric functional space $(\mathcal{H}, d)$. In this space, the open and closed balls will be indicated as $\mathcal{V}(x, \delta) = \{y \in \mathcal{H} : d(x, y) < \delta\}$ and $B(x, \delta) = \{y \in \mathcal{H} : d(x, y) \le \delta\}$, respectively. For the sake of completeness, we recall the definition of the Kolmogorov's entropy which is an important tool to obtain uniform convergence results. Given a subset $\mathcal{S}_{\mathcal{H}} \subset \mathcal{H}$ and $\epsilon > 0$, denote $N_\epsilon(\mathcal{S}_{\mathcal{H}})$ the minimal number of open balls of radius ϵ needed to cover $\mathcal{S}_{\mathcal{H}}$. Then, the quantity $\Psi_{\mathcal{S}_{\mathcal{H}}}(\epsilon) = \ln(N_\epsilon(\mathcal{S}_{\mathcal{H}}))$ is called the Kolmogorov's ϵ-entropy of the set $\mathcal{S}_{\mathcal{H}}$. As is well known, if $\mathcal{H}$ is complete, the set $\mathcal{S}_{\mathcal{H}}$ has finite entropy if and only if its closure is compact.

Throughout this paper, when no confusion will be possible, we will denote by C and C' some strictly positive generic constants.

3.1. Consistency

In this section, consistency results for the estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{g}$ defined in Section 2 are derived, under general conditions.

As mentioned in Ferraty and Vieu [27], convergence results in nonparametric statistics for functional variables are closely related to the concentration properties of the probability measure of the functional variable X given by the function ϕ defined in **H1**, while to derive uniform consistency results, a uniform upper and lower bound is needed. Sufficient conditions ensuring that the concentration property for the probability measure holds uniformly in $\mathcal{S}_{\mathcal{H}}$, are given in Section 7.2 of Ferraty et al. [22].

The following set of assumptions are needed to derive consistency results.

H1 There exists a function $\phi : \mathbb{R} \to \mathbb{R}_{>0}$ such that $\lim_{h\to 0} \phi(h) = 0$ and $0 < C\phi(h) \le \Pr\{X \in B(x, h)\} \le C'\phi(h)$, for all $x \in \mathcal{S}_{\mathcal{H}}$.

H2 The kernel K is a bounded nonnegative function with support $[0, 1]$ such that $\int K(u)du = 1$ and satisfies a Lipschitz condition of order one. Also,

(a) If $K(1) = 0$, K is differentiable with derivative K' and $-\infty < \inf_{u\in\mathbb{R}} K'(u) \le \sup_{u\in\mathbb{R}} K'(u) = \|K'\|_\infty < 0$.
(b) If $K(1) > 0$, there exist $C, C' > 0$ such that $C\, \mathbf{1}_{[0,1]}(u) < K(u) < C' \mathbf{1}_{[0,1]}(u)$.

H3 The functions ϕ and $\Psi_{\mathcal{S}_{\mathcal{H}}}$ are such that:

(a) $\phi : \mathbb{R} \to \mathbb{R}_{>0}$ is differentiable with derivative ϕ'. Moreover, there exist $C_\phi > 0$ and $\eta_0 > 0$, such that for all $\eta < \eta_0$, $\phi'(\eta) < C_\phi$.
If $K(1) = 0$, the function ϕ has to fulfil the additional condition:

$$\exists C > 0,\ \exists \eta_0 > 0 \text{ tales que } \quad \forall 0 < \eta < \eta_0 \int_0^\eta \phi(u)du > C\eta\phi(\eta)$$

(b) for n large enough,

$$\frac{\{\ln(n)\}^2}{n\phi(h)} < \Psi_{\mathcal{S}_{\mathcal{H}}}\left(\frac{\ln(n)}{n}\right) < \frac{n\phi(h)}{\ln(n)}.$$

H4 $\psi : \mathbb{R} \to \mathbb{R}$ is an odd function, strictly increasing, bounded and continuous differentiable function.
H5 The sequence $h = h_n$ is such that $h_n \to 0$, $n\phi(h_n) \to \infty$ and $n\phi(h_n)/\ln(n) \to \infty$ when $n \to \infty$.
H6 $F_j(\cdot|X = x)$ are symmetric around $\phi_j(x)$, respectively, for $0 \le j \le p$.
H7 Let $\mathcal{S}_{\mathcal{H}}$ be a compact set of $\mathcal{H}$ such as

(i) for each fixed $y \in \mathbb{R}$ and $z \in \mathbb{R}$, $F_0(y|X = x)$ and $F_j(z|X = x)$ are continuous functions of x on $\mathcal{S}_{\mathcal{H}}$, for $1 \le j \le p$.
(ii) the following equicontinuity condition holds:
$\forall \epsilon > 0\ \exists \delta > 0 : \ |u - v| < \delta \ \Rightarrow \ \sup_{x\in\mathcal{S}_{\mathcal{H}}} \max_{0\le j\le p} |F_j(u|X = x) - F_j(v|X = x)| < \epsilon$.

A general discussion on the above assumptions can be found in Boente and Vahnovan [16]. In particular, in Section 4.3 of that paper, examples of processes and compact sets satisfying **H1**, **H3** and **H5** are given.

Proposition 3.1 is a direct consequence of Theorems 4.1 and 4.3 in Boente and Vahnovan [16]. Together with Theorem 3.1, they allows us to obtain consistency results for the estimators of the regression parameter $\boldsymbol{\beta}$ and the operator g, when the kernel functional smoothing is based on local medians or M-local estimators.

Proposition 3.1. *Let $S_{\mathcal{H}} \subset \mathcal{H}$ be a compact set. Assume that **H1** to **H3** and **H5** hold for $\mathcal{S}_{\mathcal{H}}$.*

(a) *Denote as $\phi_j(x)$ the unique solution of $\lambda_j(x, a, \sigma) = 0$ for any $\sigma > 0$. Moreover, let $\widehat{s}_j(x)$ be robust scale estimators such that, with probability 1, there exist real constants $0 < A < B$ and $A < \widehat{s}_j(x) < B$ for all $x \in S_{\mathcal{H}}$ and $n \ge n_0$. Assume that, in addition, **H4** holds and either $\lambda_j(\cdot, a, \sigma) : \mathcal{H} \to \mathbb{R}$ are continuous functions on $\mathcal{S}_{\mathcal{H}}$ or **H6** and **H7** hold. Then, $\sup_{x\in\mathcal{S}_{\mathcal{H}}} |\widehat{\phi}_{j,\mathrm{M}}(x) - \phi_j(x)| \xrightarrow{a.s.} 0$ for $0 \le j \le p$.*
(b) *If, in addition, $F_j(z|X = x)$ has a unique median at $\phi_j(x)$, for $0 \le j \le p$, and **H6** and **H7** hold, we have that $\sup_{x\in\mathcal{S}_{\mathcal{H}}} |\widehat{\phi}_{j,\mathrm{MED}}(x) - \phi_j(x)| \xrightarrow{a.s.} 0$, $0 \le j \le p$.*

Lemma 3.1 and Theorem 3.1 are analogous to Lemma 1 and Theorem 1 of Bianco and Boente [9], respectively. Their proof can be found in the Appendix. The separability and completeness assumption on the functional space $\mathcal{H}$ is needed to guarantee that any probability measure on $\mathcal{H}$ will be tight.

Lemma 3.1. *Let $\mathcal{H}$ be separable and complete space and let $(r_i, \mathbf{u}_i^\top, x_i)^\top \in \mathbb{R}^{p+1} \times \mathcal{H}$, $1 \le i \le n$ be i.i.d. random vectors over $(\Omega, \mathcal{A}, \mathbb{P})$ such that $(r_i, \mathbf{u}_i^\top)^\top$ have common distribution P. Let $\widehat{\eta}_0(x)$ and $\widehat{\boldsymbol{\eta}}(x) = (\widehat{\eta}_1(x), \ldots, \widehat{\eta}_p(x))^\top$ be random functions such that, for any compact set $\mathcal{S}_{\mathcal{H}} \subset \mathcal{H}$,*

$$\sup_{x\in\mathcal{S}_{\mathcal{H}}} |\widehat{\eta}_j(x)| \xrightarrow{a.s.} 0, \quad 0 \le j \le p. \tag{6}$$

Denote as P_n and Q_n the empirical measures of $\{(r_i, \mathbf{u}_i^\top)^\top\}_{i=1}^n$ and $\{(r_i + \widehat{\eta}_0(x_i), \mathbf{u}_i^\top + \widehat{\boldsymbol{\eta}}(x_i)^\top)^\top\}_{i=1}^n$, respectively. Then,

(a) *for any bounded and continuous function $f : \mathbb{R}^{p+1} \to \mathbb{R}$, we have that $|\mathrm{E}_{Q_n}(f) - \mathrm{E}_{P_n}(f)| \xrightarrow{a.s.} 0$.*
(b) *$\pi(Q_n, P) \xrightarrow{a.s.} 0$, where π stands for the Prohorov distance.*

Theorem 3.1. *Let $\mathcal{H}$ be separable and complete space and $(Y_i, \mathbf{Z}_i^\top, X_i)^\top$, $1 \le i \le n$, independent random vectors satisfying (1). Denote P the distribution of $(r_i, \mathbf{u}_i^\top)^\top = (Y_i - \phi_0(X_i), \mathbf{Z}_i^\top - \boldsymbol{\phi}(X_i)^\top)^\top$, where $\phi_0(x) = \boldsymbol{\beta}^\top\boldsymbol{\phi}(x) + g(x)$. Let $\widehat{\phi}_j(x)$, $0 \le j \le p$ be estimators of $\phi_j(x)$ such that for any compact set $\mathcal{S}_{\mathcal{H}} \subset \mathcal{H}$*

$$\sup_{x\in\mathcal{S}_{\mathcal{H}}} |\widehat{\phi}_j(x) - \phi_j(x)| \xrightarrow{a.s.} 0, \quad 0 \le j \le p. \tag{7}$$

Furthermore, for a given probability measure Q over $\mathbb{R}^{p+1}$, denote $\boldsymbol{\beta}(Q)$ a regression functional for the model $w = \boldsymbol{\beta}^\top\mathbf{v} + \epsilon$, where $\epsilon \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^p$ are independent and $(w, \mathbf{v}^\top)^\top \sim Q$. Assume that $\boldsymbol{\beta}(Q)$ is Fisher-consistent and continuous at P. Let $\widehat{P}_n(A) = (1/n)\sum_{i=1}^n \mathbf{1}_A(\widehat{r}_i, \widehat{\mathbf{u}}_i)$ be the empirical distribution function of $\{(\widehat{r}_i, \widehat{\mathbf{u}}_i)\}_{i=1}^n$ with $\widehat{r}_i = Y_i - \widehat{\phi}_0(X_i)$ and $\widehat{\mathbf{u}}_i = \mathbf{Z}_i - \widehat{\boldsymbol{\phi}}(X_i)$, where $\widehat{\boldsymbol{\phi}}(x) = (\widehat{\phi}_1(x), \ldots, \widehat{\phi}_p(x))^\top$. Then, if $\widehat{\boldsymbol{\beta}}_{\mathrm{R}} = \boldsymbol{\beta}(\widehat{P}_n)$, we have that $\widehat{\boldsymbol{\beta}}_{\mathrm{R}} \xrightarrow{a.s.} \boldsymbol{\beta}$.

The uniform consistency of $\widehat{g}$ over compact sets follows immediately from Theorem 3.1. On the other hand, similar arguments to those considered in the proof of Theorem 4.1 of Boente and Vahnovan [16] allow to show that $\widehat{\widehat{g}}$ is also uniformly consistent. These results are stated below.

Corollary 3.1. *Let $(Y_i, \mathbf{Z}_i^\top, X_i)^\top$, $1 \le i \le n$ be independent random vectors satisfying (1).*

(a) *Assume that $\widehat{\phi}_j(x)$, $1 \le j \le p$ are estimates of $\phi_j(x)$ such that for any compact set $\mathcal{S}_\mathcal{H} \subset \mathcal{H}$ $\sup_{x \in \mathcal{S}_\mathcal{H}} |\widehat{\phi}_j(x) - \phi_j(x)| \xrightarrow{a.s.} 0$, $0 \le j \le p$. Let $\widehat{g}(x) = \widehat{\phi}_0(x) - \widehat{\boldsymbol{\beta}}_{\rm R}^\top \widehat{\boldsymbol{\phi}}(x)$. Then, under the assumptions of Theorem 3.1, we have that $\sup_{x \in \mathcal{S}_\mathcal{H}} |\widehat{g}(x) - g(x)| \xrightarrow{a.s.} 0$.*

(b) *Assume that **H1** to **H5** hold for $\mathcal{S}_\mathcal{H}$ and either $\lambda_j(\cdot, a, \sigma) : \mathcal{H} \to \mathbb{R}$ are continuous functions on $\mathcal{S}_\mathcal{H}$ or **H6** and **H7** hold. Moreover, let $\widehat{\widehat{s}}_j(x)$ be robust scale estimators such that, with probability 1, for some real constants $0 < A < B$, we have $A < \widehat{\widehat{s}}_j(x) < B$ for all $x \in S_\mathcal{H}$ and $n \ge n_0$. Let $\widehat{\widehat{g}}(x)$ be the solution of $\widehat{\widehat{\lambda}}(x, a, \widehat{\widehat{s}}(x)) = 0$, where $\widehat{\widehat{\lambda}}(x, a, \sigma)$ is given in (5), then, we have that $\sup_{x \in \mathcal{S}_\mathcal{H}} |\widehat{\widehat{g}}(x) - g(x)| \xrightarrow{a.s.} 0$.*

3.2. Asymptotic distribution

In this section, we will derive the asymptotic distribution of the regression parameter estimators, when they solve an M-estimating equation. As in Aneiros-Pérez and Vieu [3], we will require that the covariates X_i lie in a compact set $\mathcal{S}_\mathcal{H} \subset \mathcal{H}$. This is a usual condition to derive asymptotic properties of the regression parameter under a partially linear model too.

Let $\widehat{\phi}_0(x)$ and $\widehat{\boldsymbol{\phi}}(x)$ be consistent estimates of $\phi_0(x)$ and $\boldsymbol{\phi}(x)$, respectively, where $\phi_0(x)$ satisfies $\phi_0(x) = \boldsymbol{\beta}^\top \boldsymbol{\phi}(x) + g(x)$. As in Section 3.1, denote $r_i = Y_i - \phi_0(X_i)$ and $\mathbf{u}_i = \mathbf{Z}_i - \boldsymbol{\phi}(X_i)$ the nonparametric residuals and $\widehat{r}_i = Y_i - \widehat{\phi}_0(X_i)$, $\widehat{\mathbf{u}}_i = \mathbf{Z}_i - \widehat{\boldsymbol{\phi}}(X_i)$ the residual predictors. Furthermore, let ψ_1 and w_2 be a score and a weight function, respectively and denote $\widehat{\boldsymbol{\beta}}$ any consistent solution of

$$\sum_{i=1}^n \psi_1\left(\frac{\widehat{r}_i - \widehat{\boldsymbol{\beta}}^\top \widehat{\mathbf{u}}_i}{\widehat{\sigma}}\right) w_2(\|\widehat{\mathbf{u}}_i\|)\widehat{\mathbf{u}}_i = 0, \tag{8}$$

with $\widehat{\sigma}$ an estimate of the errors scale σ_0. Note that (8) corresponds to the differentiating equation that defines most robust regression estimators.

It is worth noting that the proof of the asymptotic distribution of the estimators of $\widehat{\boldsymbol{\beta}}$ given in Bianco and Boente [9] cannot be generalized to the present setting. One of the key points of their proof was to use that the covering number $N\left(\epsilon, \mathcal{F}, L^2(\mathbb{Q})\right)$ of the class of functions $\mathcal{F} = \{f \in \mathcal{C}^1[0,1] : \|f\|_\infty \le 1 \, \|f'\|_\infty \le 1\}$ is such that $\ln N\left(\nu, \mathcal{F}, L^2(\mathbb{Q})\right) \le K\nu^{-1}$. In fact, a bound for the covering number of Lipschitz functions in the functional setting can also be obtained. More precisely, let $N_\nu(\mathcal{S}_\mathcal{H})$ stand for the minimal number of balls of radius ν with respect to the metric d needed cover $\mathcal{S}_\mathcal{H}$ and denote

$$\mathcal{F} = \{g : \mathcal{S}_\mathcal{H} \subset \mathcal{H} \to \mathbb{R} \text{ continuous} : \ \|g\|_\infty \le M \text{ and } \ |g(y) - g(x)| \le M\, d(x,y) \ \forall x \in \mathcal{S}_\mathcal{H} \ \forall y \in \mathcal{S}_\mathcal{H}\}.$$

In Vahnovan [53], it is shown that

$$\ln N(\nu, \mathcal{F}, L^2(\mathbb{Q})) \le \frac{2M}{\nu} + C\, N_{(\nu/M)}(\mathcal{S}_\mathcal{H}),$$

with C a constant that does not depend on $\mathbb{Q}$ or $\mathcal{S}_\mathcal{H}$. Thus, to use the maximum inequality for covering numbers as in Bianco and Boente [9] it is needed that $\int_0^\delta \sqrt{1 + N_\nu(\mathcal{S}_\mathcal{H})}d\nu < \infty$. Note that, when $\mathcal{H} = \mathbb{R}^d$, $N_\nu(\mathcal{S}_\mathcal{H}) \simeq K\nu^{-d}$, so the required condition holds only for $d = 1$. Thus, some additional requirements will be needed. In particular, as in Speckman [51], Linton [40], He et al. [34] and González Manteiga and Aneiros Pérez [32], we will assume that the covariates $\mathbf{Z}_i$ are nonparametrically related with X_i satisfying $Z_{ij} = \phi_j(X_i) + u_{ij}$, $1 \le i \le n$, $1 \le j \le p$, where the errors u_{ij} are independent and independent of X_i.

Thus, the model that will be considered in this section may be written as

$$\begin{cases} Y_i = \boldsymbol{\beta}^\top \mathbf{Z}_i + g(X_i) + \epsilon_i & 1 \le i \le n, \\ Z_{ij} = \phi_j(X_i) + u_{ij} & 1 \le i \le n, \ 1 \le j \le p, \end{cases} \tag{9}$$

where the errors ϵ_i are independent and independent of $(\mathbf{Z}_i^\top, X_i)^\top$ and the errors u_{ij} are independent and independent of X_i. Note that $r_i - \mathbf{u}_i^\top \boldsymbol{\beta} = \epsilon_i$.

From now on, we denote as $G_j(\cdot/s_j)$ the distribution of u_{1j}, for $1 \le j \le p$, and $G_0(\cdot/s_0)$ that of r_1, that is, for $0 \le j \le p$, G_j has scale 1 so as to identify the residuals scale s_j. Furthermore, let $\nu_j(\sigma) = \mathrm{E}\psi'(u_{1j}/\sigma)$, for $1 \le j \le p$, and $\nu_0(\sigma) = \mathrm{E}\psi'(r_1/\sigma)$.

It is worth noting that under model (9), if G_j is symmetric around 0, $\phi_j(x)$ is such that $\lambda_j(x, \phi_j(x), \sigma) = 0$ for any $\sigma > 0$ and for any odd score function, that is, it corresponds to the target M-conditional location functional discussed in Section 2. Furthermore, we have that $\phi_0(x) = \boldsymbol{\beta}^\top \boldsymbol{\phi}(x) + g(x)$, as required in Theorem 3.1, and $r_i = Y_i - \phi_0(X_i) = \boldsymbol{\beta}^\top \mathbf{u}_i + \epsilon_i$. Note also that, in this situation, if the errors ϵ_i and u_{ij} have finite expectation, the regression operator ϕ_j represents the conditional expectation $\rho_j(x) = \mathrm{E}(Z_j|X = x)$ as defined in Section 2, while $\phi_0(x)$ equals $\rho_0(x) = \mathrm{E}(Y|X = x)$.

Taking into account the homoscedastic structure in (9), when defining the estimators of ϕ_j, we will consider the solution $\widehat{\phi}_j(x)$ of $\widehat{\lambda}_j(x, a, \widehat{s}_j) = 0$, where $\widehat{s}_j$ stands for a robust estimator of s_j. A possible choice for $\widehat{s}_j$, when $1 \le j \le p$, is to consider the MAD of the residuals $Z_{ij}-\widehat{\phi}_{j,\text{MED}}(x)$, where $\widehat{\phi}_{j,\text{MED}}(x)$ stands for the median of $\widehat{F}_j(\cdot|X=x)$. Similarly, $\widehat{s}_0$ may be defined as the MAD of the residuals $Y_i-\widehat{\phi}_{0,\text{MED}}(x)$, with $\widehat{\phi}_{0,\text{MED}}(x)$ the median of $\widehat{F}_0(\cdot|X=x)$.

We will need the following set of assumptions:

N1 ψ_1 is an odd, bounded and twice differentiable function with bounded derivatives ψ_1' and ψ_1'', such that $\varphi_1(t)=t\psi_1'(t)$ and $\zeta_1(t)=t\psi_1''(t)$ are bounded.

N2 $w_2(t)=\psi_2(t)t^{-1}>0$ is a bounded function such that ψ_2 is an odd, bounded and continuously differentiable function with bounded derivative ψ_2'. Moreover, $w_2(t)=\varphi_2(t^2)$ where φ_2 is a twice continuously differentiable function with bounded derivative φ_2' such that $\zeta_2(t)=t\varphi_2'(t)$ is bounded.

N3 $\mathrm{E}(w_2(\|\mathbf{u}_1\|)\|\mathbf{u}_1\|^2)<\infty$ and the matrix

$$\mathbf{A}=\mathrm{E}\left\{\psi_1'\left(\frac{r_1-\mathbf{u}_1^\top\boldsymbol{\beta}}{\sigma_0}\right)w_2(\|\mathbf{u}_1\|)\mathbf{u}_1\mathbf{u}_1^\top\right\}=\mathrm{E}\left\{\psi_1'\left(\frac{\epsilon_1}{\sigma_0}\right)\right\}\mathrm{E}\{w_2(\|\mathbf{u}_1\|)\mathbf{u}_1\mathbf{u}_1^\top\}$$

is nonsingular.

N4 The functions $\phi_j(x)$, $0\le j\le p$, are continuous on $\mathcal{S}_{\mathcal{H}}$. Moreover, ϕ_j are Lipschitz of order η on $\mathcal{S}_{\mathcal{H}}$, i.e., there exists $C>0$ such that, for any $x,y\in\mathcal{S}_{\mathcal{H}}$, $|\phi_j(x)-\phi_j(y)|\le Cd(x,y)^\eta$.

N5 $\psi:\mathbb{R}\to\mathbb{R}$ is an odd function, strictly increasing, bounded and twice continuous differentiable function with bounded derivatives ψ' and ψ'' such that $\varphi(t)=t\psi'(t)$ and $\zeta(t)=t\psi''(t)$ are bounded. Furthermore, $\nu_j=\nu_j(s_j)>0$, for $0\le j\le p$.

N6 The sequence $h=h_n$ is such that $h_n\to0$, $n\phi(h_n)\to\infty$ and $n\phi(h_n)/\ln(n)\to\infty$ when $n\to\infty$. Furthermore, conditions (10) or (11) below hold

$$h\left(\frac{n}{\ln n}\right)^{1-\eta}\le C_\eta\quad\text{for all } n\ge1 \tag{10}$$

$$\phi(h)\left(\frac{n}{\ln n}\right)^{1-\eta}\le C_\eta\quad\text{for all } n\ge1, \tag{11}$$

where η is given in **N4**.

N7 The Kolmogorov ϵ-entropy of $S_{\mathcal{H}}$ satisfies $\sum_{n=1}^\infty n\exp\left\{(1-\beta)\psi_{S_{\mathcal{H}}}(\ln(n)/n)\right\}<\infty$ for some $\beta>1$.

N8 For $0\le j\le p$, G_j are symmetric around 0. Furthermore, the vector $\mathbf{u}_1$ is such that $\boldsymbol{\Lambda}\mathbf{u}_1\overset{\mathcal{D}}{\sim}\mathbf{u}_1$, for any diagonal matrix $\boldsymbol{\Lambda}$ with elements equal to 1 or -1, where $\mathbf{u}\overset{\mathcal{D}}{\sim}\mathbf{v}$ means that the vectors $\mathbf{u}$ and $\mathbf{v}$ have the same distribution.

Remark 3.1. It is worth noting that Aneiros-Pérez and Vieu [3] assumed that the compact set $\mathcal{S}_{\mathcal{H}}$ where the covariates lie is such that $N_\nu(\mathcal{S}_{\mathcal{H}})\le C\nu^{-\gamma}$. To avoid this assumption we state the assumptions in terms of the Kolmogorov entropy of $S_{\mathcal{H}}$. Examples in which assumptions **N7** and **N8** on the entropy and the measure concentration are fulfilled are discussed in Ferraty and Vieu [27], Ferraty et al. [23,22] and Boente and Vahnovan [16].

Note that **N2** and **N8** entail that $\mathrm{E}\{w_2(\|\mathbf{u}_1\|)\mathbf{u}_1\}=0$ and also $\mathrm{E}\{w_2(\|\mathbf{u}_1\|)u_{1j}\psi'(u_{1j})\}=0$. Moreover, **N2** is fulfilled when considering the bisquare weight function $w_2(t)=\left\{1-(t/c)^2\right\}^2\mathbf{1}_{[-c,c]}(t)$ since $\varphi_2(t)=\left(1-t/c^2\right)^2\mathbf{1}_{[0,c^2]}(t)$. As mentioned in Bianco and Boente [9], **N3** prevent any element of $\mathbf{Z}$ from being almost surely perfectly predictable by X. This condition was also a requirement in Aneiros-Pérez and Vieu [3] who considered as weight function $w_2\equiv1$. On the other hand, under **H1** to **H3**, **N4** to **N8**, from Boente and Vahnovan [16], we have that, for $0\le j\le p$,

$$\sup_{x\in S_{\mathcal{H}}}|\widehat{\phi}_j(x)-\phi_j(x)|=O_{a.co.}\left(h^\eta+\theta_n\right), \tag{12}$$

where $\theta_n^2=\Psi_{S_{\mathcal{H}}}(\ln(n)/n)/(n\phi(h))$.

Recall that $G_j(\cdot/s_j)$ and $G(\cdot/\sigma_0)$ stand for the distribution functions of u_{1j} and ϵ_1, respectively.

Proposition 3.2. *Let $(Y_i,\mathbf{Z}_i^\top,X_i)^\top$, $1\le i\le n$ be independent vectors satisfying (9). Assume that G and G_j, $1\le j\le p$, are symmetric around 0 and that the random elements X_i satisfy $\Pr(X_i\in\mathcal{S}_{\mathcal{H}})=1$ with $\mathcal{S}_{\mathcal{H}}$ a compact set. Let $\widehat{\phi}_j(x)$, $0\le j\le p$ be estimators of $\phi_j(x)$ defined as the solution of $\widehat{\lambda}_j(x,a,\widehat{s}_j)=0$, where $\widehat{s}_j$ stands for a robust consistent scale estimator and denote $\theta_n^2=\Psi_{S_{\mathcal{H}}}(\ln(n)/n)/(n\phi(h))$. Then, under **H1** to **H3**, **N4** to **N8**, we have that*

$$\sup_{x\in\mathcal{S}_{\mathcal{H}}}\left|\widehat{\phi}_j(x)-\phi_j(x)-\nu_j(\widehat{s}_j)^{-1}\widehat{s}_j\sum_{i=1}^n\frac{K_i(x)}{n\mathrm{E}K_1(x)}\psi\left(\frac{Z_{ij}-\phi_j(x)}{\widehat{s}_j}\right)\right|=O_{a.s.}(h^{2\eta}+\theta_n^2),\quad 1\le j\le p,$$

$$\sup_{x\in\mathcal{S}_{\mathcal{H}}}\left|\widehat{\phi}_0(x)-\phi_0(x)-\nu_0(\widehat{s}_0)^{-1}\widehat{s}_0\sum_{i=1}^n\frac{K_i(x)}{n\mathrm{E}K_1(x)}\psi\left(\frac{Y_i-\phi_0(x)}{\widehat{s}_0}\right)\right|=O_{a.s.}(h^{2\eta}+\theta_n^2)$$

where $K_i(x)=K\left(d(x,X_i)/h\right)$.

Lemma 3.2 follows using analogous arguments to those considered in Bianco and Boente [9]. Details of its proof can be found in Vahnovan [53].

Lemma 3.2. *Let $(Y_i, \mathbf{Z}_i^\top, X_i)^\top$, $1 \le i \le n$ be independent vectors satisfying (1) where the errors distribution G is symmetric around 0 and the random elements X_i are such that $\Pr(X_i \in \mathcal{S}_{\mathcal{H}}) = 1$ with $\mathcal{S}_{\mathcal{H}}$ a compact set. Let $\widehat{\phi}_j(x)$, $0 \le j \le p$, be estimators of $\phi_j(x)$ such that $\sup_{x\in\mathcal{S}_{\mathcal{H}}} |\widehat{\phi}_j(x) - \phi_j(x)| \xrightarrow{p} 0$, $1 \le j \le p$. Given $\widetilde{\boldsymbol{\beta}}$ and $\widehat{\sigma}$ weakly consistent estimators of $\boldsymbol{\beta}$ and σ_0, denote as $\mathbf{A}_n = \widehat{\mathbf{A}}(\widetilde{\boldsymbol{\beta}}, \widehat{\sigma})$ with*

$$\widehat{\mathbf{A}}(\mathbf{b}, s) = \frac{1}{n} \sum_{i=1}^{n} \psi_1' \left(\frac{\widehat{r}_i - \widehat{\mathbf{u}}_i^\top \mathbf{b}}{s} \right) w_2\left(\|\widehat{\mathbf{u}}_i\|\right) \widehat{\mathbf{u}}_i\, \widehat{\mathbf{u}}_i^\top. \tag{13}$$

*Then, if **N1** to **N3** hold, we have that $\mathbf{A}_n \xrightarrow{p} \mathbf{A}$ where $\mathbf{A}$ is given in **N3**.*

Theorem 3.2. *Let $(Y_i, \mathbf{Z}_i^\top, X_i)^\top$, $1 \le i \le n$ be independent vectors satisfying (9). Assume that G and G_j, $1 \le j \le p$, are symmetric around 0 and that the covariates X_i satisfy $\Pr(X_1 \in \mathcal{S}_{\mathcal{H}}) = 1$ with $\mathcal{S}_{\mathcal{H}}$ a compact set. Furthermore, assume that **H1** to **H3** and **N1** to **N8** hold and that $n h^{4\eta} \to 0$ and $n^{1/2}\theta_n^2 \to 0$. Let $\widehat{\phi}_j(x)$, $0 \le j \le p$, be estimators of $\phi_j(x)$ defined as the solution of $\widehat{\lambda}_j(x, a, \widehat{s}_j) = 0$, where $\widehat{s}_j$ stand for robust consistent scale estimators. Then, if $n^{1/4}(\widehat{\sigma} - \sigma_0) = O_{\mathbb{P}}(1)$ and $n^{1/4}(\widehat{s}_j - s_j) = O_{\mathbb{P}}(1)$, for $0 \le j \le p$, we have that $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} \mathcal{N}(0, \mathbf{A}^{-1}\boldsymbol{\Sigma}(\mathbf{A}^{-1})^\top)$, where $\mathbf{A}$ is defined in **N3** and*

$$\boldsymbol{\Sigma} = \sigma_0^2 \mathrm{E}\left\{\psi_1^2\left(\frac{r_1 - \mathbf{u}_1^\top\boldsymbol{\beta}}{\sigma_0}\right) w_2^2(\|\mathbf{u}_1\|)\mathbf{u}_1\mathbf{u}_1^\top\right\} = \sigma_0^2 \mathrm{E}\left\{\psi_1^2\left(\frac{\epsilon_1}{\sigma_0}\right)\right\} \mathrm{E}\{w_2^2(\|\mathbf{u}_1\|)\mathbf{u}_1\mathbf{u}_1^\top\}.$$

Remark 3.2. In order to make inferences on the regression parameter, usually the practitioner needs, besides the estimator $\widehat{\boldsymbol{\beta}}$, an estimator of its asymptotic covariance matrix $\mathbf{B} = \mathbf{A}^{-1}\boldsymbol{\Sigma}(\mathbf{A}^{-1})^\top$. It is worth noting that an estimator of $\mathbf{B}$ can be defined plugging-in the unknown quantities by their estimates and replacing the expectations by averages. More precisely, recall that $\widehat{\phi}_0(x)$ and $\widehat{\boldsymbol{\phi}}(x)$ stand for consistent estimates of $\phi_0(x)$ and $\boldsymbol{\phi}(x)$, respectively, while $\widehat{r}_i = Y_i - \widehat{\phi}_0(X_i)$, $\widehat{\mathbf{u}}_i = \mathbf{Z}_i - \widehat{\boldsymbol{\phi}}(X_i)$ are the residual predictors. Hence, an estimator of the asymptotic variance of $\widehat{\boldsymbol{\beta}}$ can be defined as $\widehat{\mathbf{B}} = \widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}}, \widehat{\sigma})^{-1}\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}, \widehat{\sigma})\widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}}, \widehat{\sigma})^{-1}$, where $\widehat{\mathbf{A}}(\mathbf{b}, s)$ is the symmetric matrix defined in (13) and

$$\widehat{\boldsymbol{\Sigma}}(\mathbf{b}, s) = \widehat{\sigma}^2\, \frac{1}{n} \sum_{i=1}^{n} \psi_1^2 \left(\frac{\widehat{r}_i - \widehat{\mathbf{u}}_i^\top \mathbf{b}}{s} \right) w_2^2\left(\|\widehat{\mathbf{u}}_i\|\right) \widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i^\top.$$

Lemma 3.2 entails that $\widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}) \xrightarrow{p} \mathbf{A}$, while similar arguments to those considered in the proof of Lemma 3.2 allow to show that $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}) \xrightarrow{p} \boldsymbol{\Sigma}$, so that $\widehat{\mathbf{B}}$ provides a weakly consistent estimator of $\mathbf{B}$.

Remark 3.3. Note that, under the assumptions of Theorem 3.2, the robust estimators $\widehat{\phi}_j(x)$ satisfy

$$n^{1/4} \sup_{x\in\mathcal{S}_{\mathcal{H}}} |\widehat{\phi}_j(x) - \phi_j(x)| \xrightarrow{p} 0, \quad 0 \le j \le p. \tag{14}$$

Effectively, Theorem 4.4 in Boente and Vahnovan [16] entails that $\sup_{x\in S_{\mathcal{H}}} |\widehat{\phi}_j(x) - \phi_j(x)| = O_{a.co.}(h^\eta + \theta_n)$ which together with the fact that $n h^{4\eta} \to 0$ and $n^{1/2}\theta_n^2 \to 0$ leads to (14). The condition $n h^{4\eta} \to 0$ was also a requirement in Aneiros-Pérez and Vieu [3] to deal with the bias term. On the other hand, when $\mathcal{H} = \mathbb{R}^d$, the requirement $n^{1/2}\theta_n^2 \to 0$ reduces to the condition $\sqrt{n h^{2d}}/\ln n \to \infty$, which corresponds to assumption **A9** in Boente and Pardo-Fernández [14] for $d = 1$. Note that when $h = n^{-\tau}$, $\sqrt{n h^{2d}}/\ln n \to \infty$ is satisfied if $\tau < 1/(2d)$, while $n h^{4\eta} \to 0$ requires $\tau > 1/(4\eta)$. Hence, more regularity is needed as the dimension increases to avoid the bias effect.

In the infinite-dimensional setting, the choice of the semi-metric plays an important role since it may increase the concentration of $\phi(h)$ around 0, to avoid that the rate of convergence of $\widehat{\phi}_j$ deteriorates with the dimension. For instance, the projection semi-metric may be a useful tool. More precisely, let $\mathcal{H}$ be a separable Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and orthonormal basis $\{e_j : j \ge 1\}$. For a fixed integer $k > 0$, define the semi-metric $d^{(k)}(x_1, x_2) = (\sum_{j=1}^{k}\langle x_1 - x_2, e_j\rangle^2)^{1/2}$. Let X be a random element in $(\mathcal{H}, d^{(k)})$ and denote as $\chi : \mathcal{H} \to \mathbb{R}^k$ the operator $\chi(x) = (\langle x, e_1\rangle, \ldots, \langle x, e_k\rangle)$. Then, as shown in Ferraty et al. [22], for any compact set $S_{\mathcal{H}}$ of $(\mathcal{H}, d^{(k)})$, we have that $\chi(S_{\mathcal{H}})$ is a compact subset of $\mathbb{R}^k$, so the ϵ-entropy of $S_{\mathcal{H}}$ has order $\ln(1/\epsilon)$. On the other hand, Lemma 13.6 in Ferraty and Vieu [27] implies that we can take $\phi(h) = h^k$ in **H1** which means that the process is fractal of order k with respect to the semi-metric d_k and the same comments as in the Euclidean setting may be given.

The requirement $n^{1/4}(\widehat{\sigma} - \sigma_0) = O_{\mathbb{P}}(1)$ and $n^{1/4}(\widehat{s}_j - s_j) = O_{\mathbb{P}}(1)$ in Theorem 3.2 is a weak requirement, since M-scale estimators usually have a root $-n$ order of convergence. Effectively, in homoscedastic nonparametric regression with fixed real carriers, Ghement et al. [29] define a robust M-scale estimator based on differences and derive their asymptotic distribution under mild assumptions. In the present setting, a similar order of convergence may be obtained when using as estimators of s_j the MAD or more generally an M-scale estimator based on the residuals $Z_{ij} - \widehat{\phi}_{j,\text{MED}}(x)$ or $Y_i - \widehat{\phi}_{0,\text{MED}}(x)$. On the other hand, a possible choice for the estimate $\widehat{\sigma}$ of the errors scale σ_0 is the scale related to an S-regression estimator, that will lead to consistent estimators and will have the order of convergence required in our assumptions.

4. Monte Carlo study

In this section, we numerically explore the finite sample behaviour of different regression parameter and regression operator estimators. More precisely, we report the results of a Monte Carlo study comparing the classical estimators, i.e., those related to $\psi(t) = \psi_1(t) = t$ and $w_2(t) = 1$ with the robust estimators defined in Section 2, under different types of contaminations, when $p = 2$.

4.1. General description

We performed $NR = 1000$ replications generating independent samples of size $n = 100$. The clean data sets were generated according to the model used by Aneiros-Pérez and Vieu [5], that is, $Y_i = \mathbf{Z}_i^{\top}\boldsymbol{\beta} + g(X_i) + \epsilon_i$, for $1 \le i \le n$, where $\boldsymbol{\beta} = (-1, 3)^{\top}$, Z_{ij} and ϵ_i are independent random variables such that $Z_{ij} \sim \mathcal{N}(0, 1)$ and $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon = 0.1\{\max_{x\in\mathcal{H}} g(x) - \min_{x\in\mathcal{H}} g(x)\}$. The functional data were defined as $X_i(t) = a_i(t - 0.5)^2 + b_i$, $0 \le t \le 1$, where a_i and b_i are i.i.d., $a_i \sim \mathcal{U}(0, 1)$ and $b_i \sim \mathcal{U}(-0.5, 0.5)$.

The regression operator g equals $g(X_i) = \exp(-8f(X_i)) - \exp(-12f(X_i))$, with

$$f(X_i) = \text{sign}(X_i'(1) - X_i'(0))\left[3\int_0^1 \{X_i'(z)\}^2 dz\right]^{1/2},$$

where $X_i'(t)$ stands for the first derivative of $X_i(t)$.

Due to the smoothness of X_i, we may consider semi-metrics based on the L_2 norm of some derivative of the curves. As in Aneiros-Pérez and Vieu [5], we used the semi-metric

$$d_m(X, X^*) = \left[\int_0^1 \{X^{(m)}(t) - X^{*(m)}(t)\}^2 dt\right]^{1/2},$$

where $X^{(m)}$ stands for the m-derivative of X. We considered several orders $m = 0$, 1 and 2 for the derivative, but we report here only the results for $m = 1$. In particular, for the selected process, the order $m = 2$ is equivalent to $m = 1$. Note also that, for this process, the choice $m = 1$ leads, up to a constant, to the same semi-metric as that defined through the projection semi-metric $d^{(1)}(x_1, x_2) = \left(\langle x_1 - x_2, e_1\rangle^2\right)^{1/2}$, taking $\{e_j, j \ge 1\}$ as the Legendre polynomials basis, where we have reordered the first two basis elements as $e_1(t) = 12(t - 0.5)$ and $e_2(t) \equiv 1$.

In the smoothing procedure, we have used the kernel $K(u) = (35/16)(1 - u^2)^3 \mathbf{1}_{[0,1]}(u)$. We also performed a simulation with the Epanechnikov kernel that leads to similar conclusions.

To compute the robust estimators, we used in **Step 1** a local M-estimator with bisquare score function $\psi_{c,\text{T}}(t) = t\left(1 - (t/c)^2\right)^2 \mathbf{1}_{[-c,c]}(t)$ with tuning constant $c = 4.685$. We report the results obtained using, as robust estimate of the regression operator g, the local M-estimate $\widehat{\widehat{g}}(x)$ defined in Section 2 computed also with the bisquare score function. In **Step 2**, we used different estimators for $\boldsymbol{\beta}$. More precisely, we report the results obtained using as regression estimators

- M-estimators with bisquare score function with tuning constant $c = 4.685$,
- GM-estimators with Huber function $\psi_1(t) = \psi_{c_1,\text{H}}(t) = \min(c_1, \max(-c_1, t))$ with $c_1 = 1.6$ on the residuals and with bisquare weight function $w_2(t) = \psi_{c_2,\text{T}}(t)/t$ with constant $c_2 = \chi^2_{2,0.95}$ where $\Pr(\chi^2_2 \ge \chi^2_{2,1-\alpha}) = \alpha$,
- LMS-estimators introduced by Rousseeuw [46],
- LTS-estimators with 33% trimmed observations.

The residuals calibration constants chosen for the estimators are those considered in Bianco and Boente [9] in the finite dimensional case. With these constants, the M-estimates and the LTS-estimate have an asymptotic efficiency of 95% and 80% respectively. Moreover, the efficiency of GM-estimate for the chosen model is 60%, while the LMS-estimate has efficiency 0 since converges at a lower rate.

In all the tables LS denotes the least squares estimate of Aneiros-Pérez and Vieu [3], MT the M-estimates obtained with the Tukey function, GM the GM-estimates, while LTS and LMS denote the estimates obtained using the "least trimmed of squares" and the "least median of squares", respectively.

The results for clean data sets, i.e., for normal errors, are indicated by C_0, while C_1, C_2 and C_3 denote the following contaminations

- C_1: $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $0.9\,\mathcal{N}(0, \sigma_\epsilon^2) + 0.1\,\mathcal{N}(0, 25\,\sigma_\epsilon^2)$. This contamination corresponds to inflating the error and thus, will affect the variability of the regression estimators.

- C_2: $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $0.9\,\mathcal{N}(0, \sigma_\epsilon^2) + 0.1\,\mathcal{N}(0, 25\sigma_\epsilon^2)$ and 10 observations were artificially modified in $\mathbf{Z}$, but not in the responses Y, as $\mathbf{Z} = (-20, 20)^\top$. This case corresponds to introduce high leverage points besides inflating the errors. The aim of this contamination is to study how it affects the bias of the regression estimators and also the data-driven bandwidths.
- C_3: $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $0.9\,\mathcal{N}(0, \sigma_\epsilon^2)+0.1\,\mathcal{N}(5, 25\,\sigma_\epsilon^2)$. This contamination corresponds to an asymmetric contamination with large responses and will affect the regression operator.

We have also considered a contamination in which the errors are i.i.d. with distribution $0.9\,\mathcal{N}(0, \sigma_\epsilon^2) + 0.1\,\mathcal{N}(10, \sigma_\epsilon^2/25)$ that leads to similar conclusions to those given below under C_3. For that reason, the obtained results are not reported here.

The behaviour of an estimator of $\boldsymbol{\beta}$ was measured through $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2$, while that of an estimator $\widehat{g}$ of g was measured by using the mean squared error

$$MSE = \frac{1}{n}\sum_{i=1}^{n} \{\widehat{g}(X_i) - g(X_i)\}^2 .$$

To evaluate the performance under contamination, we considered the following two measures which describes the behaviour of the estimators both for contaminated data and for clean data

$$MSPE_{\text{CLEAN}} = \frac{1}{n - \sum_{i=1}^{n} \gamma_i} \frac{\sum_{j=1}^{n} (1-\gamma_j)(Y_j - \widehat{Y}_j)^2}{\text{MAD}^2(Y)}$$

$$MSPE_{\text{OUT}} = \frac{1}{\sum_{i=1}^{n} \gamma_i} \frac{\sum_{j=1}^{n} \gamma_j (Y_j - \widehat{Y}_j)^2}{\text{MAD}^2(Y)}$$

where $\gamma_i = 1$ if $(Y_i, \mathbf{Z}_i, X_i)$ corresponds to a contaminated datum and $\gamma_i = 0$ otherwise. A good estimator will produce good predictions for clean data so that one expects small values of $MSPE_{\text{CLEAN}}$ and large values of $MSPE_{\text{OUT}}$.

4.2. Bandwidth selection

The problem of bandwidth selection in nonparametric and partially linear models has been extensively studied, specially when the covariates of the nonparametric component are real. It is well known that the estimation of the nonparametric regression function/operator heavily depends on the choice of the smoothing parameter h that must be chosen to provide a trade-off between bias and variance. For functional nonparametric regression models, leave-one-out cross validation was considered in Ferraty and Vieu [26], while its asymptotic properties were studied in Rachdi and Vieu [42]. Some other proposals based on local leave-one-out cross validation, Bayesian strategies and a functional adaptation of the minimax techniques, were introduced in Benhenni et al. [8], Shang [47,48] and Chagny and Roche [18], respectively. Under a semi-functional partial linear model, Aneiros-Pérez and Vieu [5] discussed an approach based on the cross-validation criterion including the choice of a local adaptive bandwidth, while Shang [49] proposed a Bayesian approach related to that given in Shang [47].

Under a fully nonparametric regression model with finite-dimensional covariates, the need of a robust criterion for selecting smoothing parameters even when considering robust estimators, has been discussed among others by Leung et al. [39], Wang and Scott [54], Boente et al. [13], Cantoni and Ronchetti [17] and Leung [38]. The ideas of robust cross-validation have been adapted to partially linear models in the finite-dimensional setting by Bianco and Boente [10] and Boente and Rodriguez [15] who also considered a plug-in approach.

When considering robust estimators as those defined in Section 2, a fully automatic robust procedure is needed to select the smoothing parameters involved in the computation of the regression estimator $\widehat{\boldsymbol{\beta}}$ and the regression operator estimators $\widehat{g}$ and $\widehat{\widehat{g}}$. In this simulation study, we use a K-fold cross-validation procedure related to that described in Bianco and Boente [10] but adapted to our functional setting.

To define the criterion, we first randomly partition the data set into K disjoint subsets of approximately equal sizes, with indices $\mathcal{I}_j$, $1 \le j \le K$, so that $\bigcup_{j=1}^K \mathcal{I}_j = \{1, \ldots, n\}$. Let $\mathcal{H}_n \subset \mathbb{R}$ be the set of bandwidths to be considered and denote $\widehat{\boldsymbol{\beta}}_h^{(-j)}$ and $\widehat{g}_h^{(-j)}$ the robust estimators of the regression parameter and the nonparametric operator computed without the observations with indices in $\mathcal{I}_j$ and using as smoothing parameter $h \in \mathcal{H}_n$. For each $1 \le i \le n$, taking into account that for some $1 \le j \le K$ we have that $i \in \mathcal{I}_j$, we can define the prediction residuals $\widehat{\epsilon}_i$ as $\widehat{\epsilon}_i = Y_i - \mathbf{Z}_i^\top \widehat{\boldsymbol{\beta}}_h^{(-j)} - \widehat{g}_h^{(-j)}(X_i)$. Noting that the cross-validation criterion defined in Aneiros-Pérez and Vieu [5] can be decomposed into the sum of the squared bias and the variance, when the weights given to each observation are equal to 1, it seems sensible to use robust measures of bias and dispersion instead. Denote as $\mu_n(V_1, \ldots, V_n)$ and $\sigma_n(V_1, \ldots, V_n)$ robust estimators of location and dispersion based on the sample $\{V_1, \ldots, V_n\}$, such as the sample median and the MAD (median of the absolute deviations with respect to the median). The robust Kth fold cross-validation smoothing parameter is defined as $\widehat{h} = \text{argmin}_{h\in\mathcal{H}_n} RCV(h)$ where

$$RCV(h) = \mu_n^2\left(\widehat{\epsilon}_1, \ldots, \widehat{\epsilon}_n\right) + \sigma_n^2\left(\widehat{\epsilon}_1, \ldots, \widehat{\epsilon}_n\right). \tag{15}$$

As in Boente and Rodriguez [15] and similar to Aneiros-Pérez and Vieu [5], a local bandwidth can be defined using a local criterion that weights each residual according to a function $W_x(X_i)$ when computing μ_n and σ_n. When μ_n is the mean and σ_n is the standard deviation, the expression given in (15) reduces to the classical criterion defined as

$$CCV(h) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\epsilon}_i^2. \tag{16}$$

It is worth noticing that the leave-one-out cross-validation method is a particular case of the K-fold method and is obtained taking $K = n$ and $\mathcal{I}_j = \{j\}$, $1 \le j \le n$. Note also that when $K = n$, μ_n is the mean and σ_n is the standard deviation, $CCV(h)$ differs from the cross-validation criterion considered in Aneiros-Pérez and Vieu [5]. Effectively, in our procedure the observation j is removed to compute the regression parameter and the regression operator $\widehat{g}_h^{(j)}$, while in the criterion defined in Aneiros-Pérez and Vieu [5], the regression parameter is estimated with all the sample and after its estimation, the ith observation $(Y_i, \mathbf{Z}_i^\top, X_i)^\top$ is removed from the sample to compute the leave-one-out estimator of g.

Once the data driven-bandwidth $\widehat{h}$ is obtained, the regression estimator denoted $\widehat{\boldsymbol{\beta}}_{\widehat{h}}$ can be computed using this bandwidth. To provide an estimator of the regression operator, a second K-fold step can be used. More precisely, denote as $\widehat{V}_i = Y_i - \mathbf{Z}_i^\top \widehat{\boldsymbol{\beta}}_{\widehat{h}}$ and recall that $\widehat{\widehat{g}}_h$ stands for the robust M-smoother of $\widehat{V}_i$ defined as the solution of (5) when the bandwidth h is used. As above, to select the bandwidth of the M-smoother divide the data set into K disjoint subsets of approximately equal sizes, with indices $\mathcal{I}_j$, $1 \le j \le K$ and let $\mathcal{H}_n \subset \mathbb{R}$ be the set of possible bandwidths. Denote as $\widehat{\widehat{g}}_h^{(-j)}$ the regression estimator obtained computed with the observations $\{(\widehat{V}_i, X_i) : i \notin \mathcal{I}_j\}$ and bandwidth h. Hence, one may select $\widetilde{h}$ as the value $\widetilde{h} = \text{argmin}_{h\in\mathcal{H}_n} RCV_2(h)$, where $RCV_2(h)$ is defined as in (15) using as residuals $\widehat{\widehat{\epsilon}}_i = \widehat{V}_i - \widehat{\widehat{g}}_h^{(-j)}(X_i)$ instead of $\widehat{\epsilon}_1$.

In our simulation study, to select the bandwidths of the classical estimator we used the standard K-fold cross-validation procedure (16) while for the robust estimates, we used the robust K-fold method (15) with μ_n the median and σ_n the MAD. In all these cases, we choose $K = 5$. The set H_n of possible values of h was set as an equidistant grid of length 21 between 0.04 and 0.40. When the minimum of the objective function is attained at 0.04, a new search was performed between 0.008 and 0.04 with a step of 0.008. On the other hand, while if the minimum is attained at 0.40, the grid was enlarge to include the values $\{0.45, 0.50, 0.55, 0.60\}$. Once the data-driven bandwidth $\widehat{h}$ is obtained, two robust estimators of the regression operator g are computed. The first one solves (5) with $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_{\widehat{h}}$ and weights $K\left(d(x, X_i)/\widehat{h}\right)$, while the second one selects the bandwidth using the K-fold cross-validation criterion $RCV_2(h)$. To distinguish both estimators we denote them as $\widehat{\widehat{g}}_{\widehat{h}}$ and $\widehat{\widehat{g}}_{\widetilde{h}}$, respectively. It is worth noticing that then considering the least squares estimators $\widehat{\widehat{g}}_{\widehat{h}}(x) = \widehat{g}_{\widehat{h}}(x)$.

4.3. Results and comments

Tables 1–5 summarize the results of the simulations. The simulation confirms the expected inadequate behaviour of the least-squares estimates in the presence of outliers in the carriers. With respect to the estimation of the regression parameter $\boldsymbol{\beta}$, from the results in Table 1, we notice that the classical method presents some advantage over the robust method, under C_0. The M-estimators give the better results among the robust alternatives under C_0, since they are the most efficient. On the contrary, under C_2, the estimates based on least median and least trimmed estimates and specially those based on GM-estimators show a much better performance than the least squares or the M-estimator. Under C_0 and when considering the GM-estimator, the mean over replications of $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2$ is about 5 times larger than that of the LS-estimator, while under C_2 the situation is reversed, since the average over replications of the LS-estimator is 175 times larger than that of the GM-estimator. Moreover, the ratio between the mean of $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2$ under C_2 and under C_0 is smaller than 2 when using a GM-estimator, while for the least median and least trimmed estimators this ratio is larger than 2. Furthermore, when considering the least squares or the M-estimators the ratio equals 1277 and 574, respectively. This shows the lack of robustness of the classical estimators as well as that of the M-estimators that take extreme values under the presence of high leverage points. Finally, all methods appear to be mainly unaffected by C_1, which corresponds to the presence of errors with larger variances. As shown in Table 2, contamination C_1 only increases the variability of the least squares regression estimators but do not affect its bias, while, as is other regression settings, contaminating with high leverage points has a high impact on the bias leading to large values of $\|\widehat{\boldsymbol{\beta}}_{\widehat{h}} - \boldsymbol{\beta}\|^2$. It is worth noticing that, even if the variability of the least squares estimator under C_1 is almost twice that obtained for clean data, the MAD of the robust estimators is at most 10% larger under C_1 than under C_0. This improvement of the robust procedures over the classical one under C_1 is also reflected on the values reported in Table 1. With respect to C_3, this asymmetric contamination is more harmful to the GM-estimators than to the M-estimators, both in bias and dispersion. However, the robust estimators show a better performance than the classical ones under this contamination which is more harmful to the nonparametric regression component.

With respect to the estimation of the regression operator, Tables 3 and 4 show that, under C_2, the least squares and M-estimators estimate inadequately the regression operator and provide bad predictions. Furthermore, smoothing the

Table 1
Mean over replications of $\|\widehat{\boldsymbol{\beta}}_{\widehat{h}} - \boldsymbol{\beta}\|^2$.

	LS	MT	GM	LMS	LTS
C_0	0.0022	0.0064	0.0102	0.0249	0.0241
C_1	0.0038	0.0068	0.0105	0.0261	0.0239
C_2	2.8112	2.8149	0.0157	0.0586	0.0571
C_3	0.1923	0.0317	0.0868	0.1474	0.1458

Table 2
Median and MAD over replications of $\widehat{\beta}_{1,\widehat{h}} - \beta_1$ and $\widehat{\beta}_{2,\widehat{h}} - \beta_2$.

		$\widehat{\beta}_{1,\widehat{h}} - \beta_1$					$\widehat{\beta}_{2,\widehat{h}} - \beta_2$				
		LS	MT	GM	LMS	LTS	LS	MT	GM	LMS	LTS
MEDIAN	C_0	0.0000	0.0015	0.0011	0.0022	0.0031	0.0000	−0.0005	−0.0006	−0.0005	−0.0011
	C_1	0.0000	0.0017	0.0013	0.0023	0.0032	−0.0001	−0.0005	−0.0005	−0.0011	−0.0016
	C_2	1.9680	1.9642	0.0008	0.0019	0.0010	−1.9839	−1.9855	0.0015	0.0015	0.0027
	C_3	0.0060	0.0011	0.0035	0.0094	0.0087	−0.0116	−0.0037	−0.0045	−0.0056	−0.0121
MAD	C_0	0.0017	0.0046	0.0074	0.0157	0.0150	0.0017	0.0045	0.0072	0.0147	0.0140
	C_1	0.0029	0.0051	0.0078	0.0159	0.0146	0.0028	0.0046	0.0072	0.0153	0.0135
	C_2	0.2105	0.2162	0.0108	0.0347	0.0375	0.2049	0.2124	0.0109	0.0356	0.0350
	C_3	0.1525	0.0192	0.0575	0.0910	0.0921	0.1517	0.0185	0.0555	0.0884	0.0929

residuals $\widehat{V}_i$ leads only to a small improvement on the predicted mean square errors specially for the least squares estimators, under C_0 and C_1. On the other hand, the estimates based on the robust procedures show a much better performance after residuals smoothing, i.e., when using $\widehat{\widehat{g}}_{\widetilde{h}}$. Note that both under C_0 and C_1 similar values are obtained for the mean and median over replications of $1000 \times MSE(\widehat{\widehat{g}})$. However, under C_2 and C_3, large values of the mean are observed for the robust procedures while the medians over replications of the mean square errors of the estimates $\widehat{\widehat{g}}$ based on *GM*-, *LMS*- and *LTS*-estimators is considerably reduced. This is related to the fact that in some replications, small values of the bandwidths are obtained leading to the sensitivity of the procedure, since a large amount of atypical data arise in the neighbourhood of each datum X_i. For that reason, we also report in Table 3, the upper trimmed means of $1000 \times MSE(\widehat{\widehat{g}})$ with trimming 1% and 5%, i.e., we have eliminated the largest 1% and 5% values obtained over replications (which in fact are the harmful ones) before computing the mean. The obtained trimmed means show that the amount of replications with large values of the mean square error is less than 1%. They also illustrate the stability of all the robust estimators except the *M*-estimators which are very sensitive to C_2.

Note that, under C_2, the estimators computed smoothing the residuals $\widehat{V}_i$ based on *GM*-estimators provide trimmed means and median values of $1000 \times MSE(\widehat{\widehat{g}}_{\widetilde{h}})$ similar to those obtained under C_0, which are only around 2.5 times those obtained for the least squares estimator for clean data. Besides, under C_0, the procedures based on the least median of squares and the least trimmed regression estimators lead to larger values of the mean square errors than those obtained with the *GM*-estimator, showing the loss of efficiency of these procedures. Indeed, under C_0, for these two estimators the mean square error of $\widehat{\widehat{g}}_{\widetilde{h}}$ is almost 4 times that of the least squares estimators. It is worth noticing that contamination C_3 seems to affect more the estimates computed using a *GM*-regression estimator than the *M*-estimator, since the median and the 5% trimmed means are smaller for the latter. In this case, the advantage of the estimators $\widehat{\widehat{g}}_{\widetilde{h}}$ over $\widehat{\widehat{g}}_{\widehat{h}}$ is not so large for *M* and *GM*-estimators while it is more evident with the *LMS*-estimators.

Table 4 also highlights the lack of robustness of least squares under contamination. Effectively, the classical procedure based on least squares tries to compromise between outlying and non-outlying observations and this is reflected on the values of $MSPE_{\text{CLEAN}}$ and $MSPE_{\text{OUT}}$. Note that, under C_2, the *GM*-estimators combined with residual smoothing lead to the smaller $MSPE_{\text{CLEAN}}$ showing its good prediction capability. Besides, when contaminating with leverage points, similar large values of $MSPE_{\text{OUT}}$ are obtained for the *GM*, *LMS* and *LTS*-estimators showing their ability to detect these atypical observations.

Even though a careful analysis of the bandwidth behaviour is beyond the scope of the paper, to study the performance of the selectors under the considered contaminations, Table 5 reports the median over replications of the data-driven bandwidths obtained for all the estimators, while Fig. 1 plots the density estimators of the smoothing parameters obtained for the classical and the *GM*-estimators. The solid black lines correspond to densities of the bandwidths obtained for clean data, while the solid and dashed grey lines to those obtained under C_1 and C_2, respectively. Finally, the dashed–dotted light grey lines correspond to C_3. As expected, the lack of robustness of classical cross-validation under contaminations with leverage points or with large responses is observed in the plots. Indeed, under C_2 and C_3 the classical data-driven bandwidth leads to over-smoothing. Its is worth noticing that the data-driven bandwidths $\widehat{h}$ obtained minimizing $RCV(h)$ are larger than those obtained minimizing $RCV_2(h)$ since, as expected, the residuals $\widehat{V}_i = Y_i - \widehat{\boldsymbol{\beta}}_{\widehat{h}}^{\top} \mathbf{Z}_i$ have lower variability that the original responses. In fact, when using *GM*-regression estimators, the bandwidth $\widehat{h}$ is slightly affected by C_2, while the selector based on $RCV_2(h)$ leads to similar data-driven bandwidths for clean and for contaminated data. Only under C_3 the

Table 3
Mean, upper trimmed means and median over replications of $1000 \times MSE(\widehat{\widehat{g}})$.

		LS	MT	GM	LMS	LTS	LS	MT	GM	LMS	LTS
		Mean					Median				
$\widehat{\widehat{g}}_{\widehat{h}}$	C_0	0.0605	0.2533	0.2852	0.3673	0.3166	0.0568	0.2373	0.2596	0.3256	0.2850
	C_1	0.1325	0.2652	0.2990	0.3917	0.3316	0.1189	0.2475	0.2677	0.3387	0.2996
	C_2	261.1257	380.8371	0.3820	306.7654	46.9271	150.3031	233.9465	0.3549	0.4945	0.4578
	C_3	320.8180	2.0200	12.7878	47.1658	7.2635	278.4263	0.3488	0.5433	1.1713	0.9157
$\widehat{\widehat{g}}_{\widetilde{h}}$	C_0	0.0600	0.1477	0.1632	0.2576	0.2386	0.0549	0.1140	0.1299	0.1971	0.1839
	C_1	0.1382	0.1655	0.1831	0.2789	0.2683	0.1208	0.1395	0.1575	0.2158	0.2118
	C_2	272.2864	363.4600	288.1509	112.6632	198.1059	176.1610	216.7916	0.2004	0.3781	0.3887
	C_3	326.7110	7.3351	4.8904	8.5787	3.9349	288.4221	0.2227	0.4859	0.7220	0.7867
		Trimmed Mean 1%					Trimmed Mean 5%				
$\widehat{\widehat{g}}_{\widehat{h}}$	C_0	0.0596	0.2488	0.2781	0.3499	0.3025	0.0572	0.2358	0.2620	0.3276	0.2838
	C_1	0.1284	0.2598	0.2900	0.3722	0.3199	0.1214	0.2464	0.2725	0.3479	0.3009
	C_2	237.9214	347.5331	0.3732	0.6193	0.5659	201.8992	290.1651	0.3524	0.5267	0.4947
	C_3	311.7964	0.3980	3.5321	33.5273	2.7940	291.6824	0.3649	0.6880	8.4066	1.8664
$\widehat{\widehat{g}}_{\widetilde{h}}$	C_0	0.0590	0.1436	0.1578	0.2435	0.2268	0.0565	0.1326	0.1464	0.2186	0.2052
	C_1	0.1338	0.1612	0.1774	0.2603	0.2467	0.1258	0.1504	0.1661	0.2358	0.2263
	C_2	253.2630	334.6200	0.2266	0.5317	0.5344	219.2911	283.1691	0.2104	0.4413	0.4464
	C_3	317.1967	1.1774	0.8311	2.3398	1.8756	296.6149	0.2573	0.6425	1.3196	1.2869

Table 4
Mean over replications of $1000 \times MSPE_{\text{CLEAN}}$ and $MSPE_{\text{OUT}}$ when using as regression estimators $\widehat{\widehat{g}}_{\widehat{h}}$ and $\widehat{\widehat{g}}_{\widetilde{h}}$.

		$1000 \times MSPE_{\text{CLEAN}}$					$MSPE_{\text{OUT}}$				
		LS	MT	GM	LMS	LTS	LS	MT	GM	LMS	LTS
$\widehat{\widehat{g}}_{\widehat{h}}$	C_0	0.0633	0.0516	0.0653	0.1610	0.1553					
	C_1	0.0904	0.0531	0.0666	0.1777	0.1496	0.0017	0.0014	0.0014	0.0018	0.0016
	C_2	2369.9763	764.7225	0.0883	3.6060	0.5611	2.8483	1.0402	481.0621	481.0944	481.9191
	C_3	105.2175	0.2771	1.5532	4.4791	3.2752	7.8885	7.0835	6.3120	6.9194	6.9867
$\widehat{\widehat{g}}_{\widetilde{h}}$	C_0	0.0201	0.0369	0.0484	0.1458	0.1417					
	C_1	0.0295	0.0393	0.0507	0.1625	0.1347	0.0011	0.0014	0.0014	0.0018	0.0015
	C_2	771.8536	766.3227	4.3219	2.5493	0.5426	1.0522	1.0424	480.9444	481.5092	481.7513
	C_3	34.8178	0.4153	1.3217	3.8085	3.3067	4.8365	7.0601	6.3289	6.9977	6.9937

Table 5
Median over replications of the data-driven bandwidths.

		LS	MT	GM	LMS	LTS
$\widehat{\widehat{g}}_{\widehat{h}}$	C_0	0.040	0.148	0.166	0.184	0.148
	C_1	0.058	0.148	0.166	0.184	0.166
	C_2	0.346	0.220	0.220	0.238	0.238
	C_3	0.346	0.202	0.202	0.076	0.094
$\widehat{\widehat{g}}_{\widetilde{h}}$	C_0	0.040	0.076	0.076	0.094	0.094
	C_1	0.058	0.076	0.094	0.094	0.094
	C_2	0.292	0.220	0.094	0.148	0.130
	C_3	0.310	0.094	0.148	0.166	0.166

density estimators show that some large values of the bandwidth may be obtained for the *GM*-estimator. A small sensitivity of the robust data-driven procedure given by $RCV(h)$ is also observed under C_3, where the density of $\widehat{h}$ has two modes indicating that the procedure can lead to some small or large values of the data-driven smoothing parameter. However, the procedure is much more stable than the classical one. Note also that as shown in Table 5, these conclusions are also valid for the least median and least trimmed estimators, while the bandwidths $\widetilde{h}$ obtained with the *M*-regression estimator are sensitive to C_2 due to the effect of leverage points on the regression parameter estimator $\widehat{\boldsymbol{\beta}}$. This fact also explains the behaviour of the mean square errors and predicted errors reported in Tables 3 and 4, respectively. In summary, Fig. 1 and Tables 3–5 show that the data-driven bandwidth $\widetilde{h}$ based on the robust criterion $RCV_2(h)$ related to the *M*-smoothers computed over the residuals of the *GM*-estimator is much more stable and a preferable choice.

Based on the obtained results, we recommend using a *GM*-estimator combined with an *M*-smoother of the residuals and the robust cross-validation criterion $RCV_2(h)$.

5. A real data set

In this section, we present the analysis of a real data, the spectrometric data set, described and analysed in Ferraty and Vieu [27], Aneiros-Pérez and Vieu [3] and Shang [49]. In particular, these last authors show that the semi-functional partial

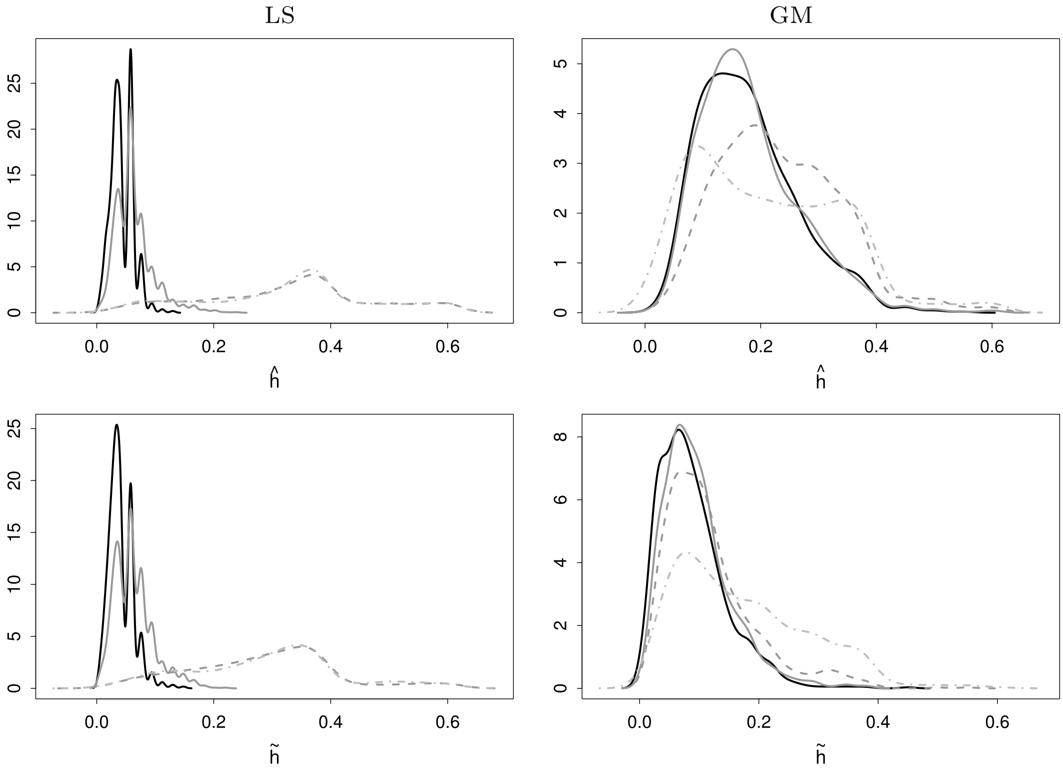

Fig. 1. Density estimators of the obtained bandwidths. Upper plots correspond to the bandwidth $\widehat{h}$, while lower ones to $\widetilde{h}$. The densities corresponding to C_0 to C_3 are given in solid black lines, solid and dashed grey lines and dashed–dotted light grey lines, respectively.

linear regression model gives more accurate forecasts than the functional nonparametric regression, since it uses additional information about protein and moisture contents. The data set was obtained from http://lib.stat.cmu.edu/datasets/tecator. Each food sample contains finely chopped pure meat with different percentages of the fat, protein and moisture contents. For each sample, we observe one spectrometric curve, denoted X_i, which corresponds to the absorbance measured at a grid of 100 wavelengths ranging from 850 nm to 1050 nm in the step. The fat, protein and moisture contents, measured in percent, are determined by analytic chemistry. More details on the data can be found in Ferraty and Vieu [27]. The aim of this analysis is not to achieve a full case study but to show how our method can be used to detect outliers for this real data set.

The considered model is a semi-functional partial linear model $Y = Z_1\beta_1 + Z_2\beta_2 + g(X) + \epsilon$, where Y is the percentage of fat content, Z_1 and Z_2 the corresponding percentages of protein content and moisture content, respectively, and X is the spectrometric curve. The data set consists on $n = 215$ independent observations $(Y_i, Z_{i1}, Z_{i2}, X_i)$, $i = 1, \ldots, n$, of $\{(Y, Z_1, Z_2, X)\}$. As in Aneiros-Pérez and Vieu [3], the sample was divided into two subsets: a training subset $\mathcal{I}$ used to select some parameters of the estimates and the testing one $\mathcal{J}$ to verify the quality of prediction.

We considered as proximity measure the semi-metric

$$d_m(X, X^*) = \left[\int_0^1 \{X^{(m)}(t) - X^{*(m)}(t)\}^2 dt\right]^{1/2},$$

where $X^{(m)}$ stands for the m-derivative of X. In the results to be reported we choose $m = 1$, which corresponds to the optimal choice of m as reported also in Aneiros-Pérez and Vieu [3] and which highlights ranges of wavelengths with different large variations.

Before choosing the training and testing set, we looked for the performance of the trajectories, to detect possible isolated trajectories. We performed two studies. We used the functional boxplot (Sun and Genton [52]) and also, as in Gervini [28], the boxplot of the distances given by the semi-metric d_m. More precisely, we computed $\kappa_i = d(X_i, X_{i,[\alpha n]})$, where $X_{i,[\alpha n]}$ is the αth neighbour to X_i with $\alpha = 0.5$ and due to the asymmetry of κ_i, we performed an skewed-adjusted boxplot [37]. Figs. 2 and 3 plot the trajectories X_i and the first derivative of X_i. The left panel of Fig. 2 shows the functional boxplot with the five detected outliers in dotted lines with circles. On the other hand, the right panel of Fig. 2 depicts all the trajectories together with the outliers detected either by the functional boxplot or using the adjusted boxplot of κ_i. To distinguish the different types of atypical observations, the five outliers detected by the functional boxplot are given in dotted lines with circles, while the ten trajectories detected using the adjusted boxplot over κ_i are given solid black lines. These curves, corresponding to the observations labelled 34, 35, 43, 44, 45, 129, 140, 172, 186 and 215, are also shown in black lines in Fig. 3. Taking into account that the local kernel M-estimators are robust with respect to outliers in the responses but they cannot deal with

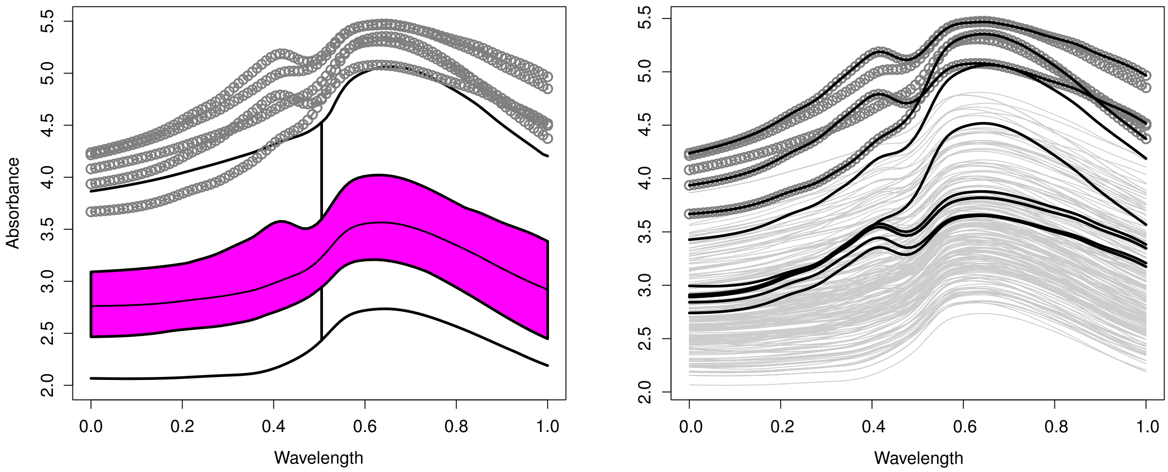

Fig. 2. In the left panel the functional boxplot of the spectrometric curves X_i is given with the detected outliers detected in dotted lines with circles. In the right panel, all the trajectories are plotted in light grey, we have highlighted in solid black lines the outliers detected by the adjusted boxplot of κ_i and with circles those detected by the functional boxplot. The observations labelled as 186 an 215 have identical trajectories.

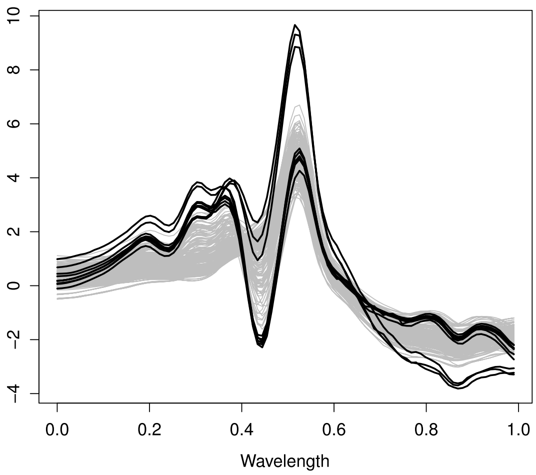

Fig. 3. First derivatives of spectrometric curves with the outliers detected by the adjusted boxplot of κ_i, in solid black lines.

atypical data on X_i, we removed the nine observations which appear to be isolated with respect to the semi-metric used. From the remaining 205 observations, the training set correspond to the first 155 observations, while the training set to the 50 remaining ones.

The estimators considered are the classical least squares kernel estimator introduced in Aneiros-Pérez and Vieu [3] and the robust GM-estimator computed using as score function on the residuals the Huber function with tuning constant $c_1 = 1.6$ and as weight function the bisquare weight function with constant $c_2 = \chi^2_{2,0.95}$. The weights w_2 which control the leverage of $\widehat{\mathbf{u}}_i$ were computed over the robust Mahalanobis distances $\{(\widehat{\mathbf{u}}_i - \widehat{\boldsymbol{\mu}})^\top \widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\mathbf{u}}_i - \widehat{\boldsymbol{\mu}})\}^{1/2}$, where $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ are S-estimators of $\widehat{\mathbf{u}}_i$. The kernel used in the smoothing step was the same as in our simulation study, i.e., $K(u) = (35/16)(1-u^2)^3 I_{[0,1]}(u)$. The data-driven bandwidths were computed through a K-fold cross-validation procedure, with $K = 5$, as described in Section 4.2. As in the simulation study, for the GM-estimators, we use the robust K-fold method (15) with μ_n the median and σ_n the MAD, while for the classical estimators we use the least squares K-fold procedure defined in (16).

We summarize the obtained results using the mean and median square prediction error over the testing subset defined as follows

$$MSPE = \frac{1}{\#\mathcal{J}} \frac{\sum_{j\in\mathcal{J}} (Y_j - \widehat{Y}_j)^2}{\text{MAD}^2_{\mathcal{J}}(Y)} \qquad MedPE = \frac{\text{median}_{\mathcal{J}}(Y_j - \widehat{Y}_j)^2}{\text{MAD}^2_{\mathcal{J}}(Y)},$$

Table 6
Analysis of the spectrometric data set.

	$\widehat{\beta}_1$	$\widehat{\beta}_2$	*MSPE*	*MedPE*
LS	−0.7777	−0.6764	0.0034	0.0018
GM	−1.0189	−0.8560	0.0078	0.0007

Table 7
Analysis of the spectrometric data set after identifying the outliers.

	$\widehat{\beta}_1$	$\widehat{\beta}_2$	*MSPE*	$MSPE_{\text{CLEAN}}$	*MedPE*
LS	−0.7777	−0.6764	0.0034	0.0022	0.0018
$LS^{-\text{OUT}}$	−1.0534	−0.8320	0.0091	0.0013	0.0006
GM	−1.0189	−0.8560	0.0078	0.0012	0.0007

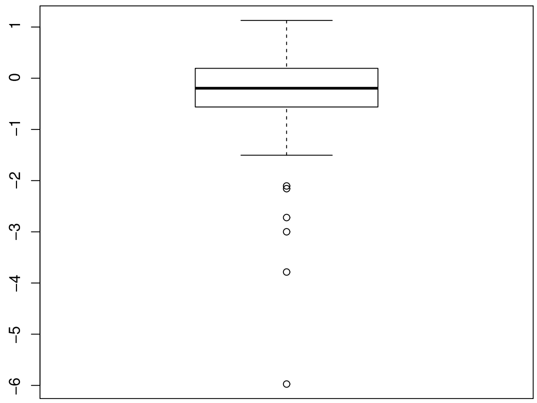

Fig. 4. Residuals boxplot on the testing set.

where $\#\mathcal{J}$ stands for the number of elements of the testing set $\mathcal{J}$. After detecting suspicious observations on the testing set, we also computed

$$MSPE_{\text{CLEAN}} = \frac{1}{\#\mathcal{J} - \sum_{i\in\mathcal{J}} \gamma_i} \frac{\sum_{j\in\mathcal{J}} (1-\gamma_j)(Y_j - \widehat{Y}_j)^2}{\text{MAD}^2_{\mathcal{J}}(Y)},$$

where $\gamma_i = 1$ if $(Y_i, \mathbf{Z}_i^\top, X_i)^\top$ corresponds to a detected atypical data and $\gamma_i = 0$ elsewhere, to evaluate the ability of the procedure to predict the non-outlying observations. Table 6 reports the obtained values for the parameter estimates as well as the mean and median square prediction errors obtained with the least squares and *GM*-estimators.

The observed differences between the mean and median square prediction error suggest the presence of some possible atypical observations in the training and/or testing sets. This also explains the better fit obtained by the *GM*-estimator when measuring the error with the median square prediction error. To detect the atypical observations on both subsets we performed a residual analysis, where the residuals were computed using the *GM*-estimators. As an illustration, Fig. 4 gives the boxplot of the residuals on the testing set, where 6 residuals have large negative values. The $MSPE_{\text{CLEAN}}$ was computed taking into account these observations and is reported in Table 7 together with the results obtained when the least squares method is applied to the training set without the identified atypical observations. We denote this last procedure as $LS^{-\text{OUT}}$ in Table 7. The obtained results show that, as is well known, the least squares procedure tends to compromise between outlying and non-outlying data leading to larger values of the $MSPE_{\text{CLEAN}}$. After removing the suspicious atypical observations, the classical procedure leads to results similar to those obtained with the robust method. This confirms the usefulness of robust estimators both to provide reliable inference methods and to identify potential outliers.

Acknowledgments

The authors wish to thank the Associate Editor and two anonymous referees for valuable comments which led to an improved version of the original paper. This research was partially supported by Grants PIP 112-201101-00339 from CONICET, PICT 2014-0351 from ANPCYT and 20120130100279BA from the Universidad de Buenos Aires at Buenos Aires, Argentina.

Appendix

Proof of Lemma 3.1. As in Bianco and Boente [9] it is enough to prove (a). Using that $\mathcal{H}$ is a separable and complete space, we get that for any $\epsilon > 0$, there exist compact sets $\mathcal{K}_1 \subset \mathbb{R}^{p+1}$ and $\mathcal{S}_{\mathcal{H}} \subset \mathcal{H}$ such that, if $\mathcal{K} = \mathcal{K}_1 \times \mathcal{S}_{\mathcal{H}}$, then $\Pr(\mathcal{K}) > 1 - (\epsilon/8\|f\|_\infty)$. Note that $|\mathrm{E}_{Q_n}(f) - \mathrm{E}_{P_n}(f)| \le A_{1n} + A_{2n}$, where

$$A_{1n} = \frac{1}{n}\sum_{i=1}^{n} |f(r_i + \widehat{\eta}_0(X_i), \mathbf{u}_i + \widehat{\boldsymbol{\eta}}(X_i)) - f(r_i, \mathbf{u}_i)| \, \mathbf{1}_{\mathcal{K}}(r_i, \mathbf{u}_i, X_i),$$
$$A_{2n} = 2\|f\|_\infty \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\mathcal{K}^c}(r_i, \mathbf{u}_i, X_i).$$

From (6) and the Strong Law of Large Numbers, we have that there exists a set $\mathcal{N} \subset \Omega$ such that $\Pr(\mathcal{N}) = 0$ and such that for any $\omega \notin \mathcal{N}$

$$\sup_{x\in\mathcal{S}_{\mathcal{H}}} |\widehat{\eta}_0(x)| + \sup_{x\in\mathcal{S}_{\mathcal{H}}} \|\widehat{\boldsymbol{\eta}}(x)\| \to 0 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\mathcal{K}^c}(r_i, \mathbf{u}_i, x_i) \to \Pr(\mathcal{K}^c). \tag{A.1}$$

Hence, for n large enough $A_{2n} \le \epsilon/2$ for $\omega \notin N$.

Denote by $\mathcal{C}_1$ the closure of a neighbourhood of radius 1 of $\mathcal{K}_1$. The uniform continuity of f on $\mathcal{C}_1$ implies that there exists δ such that $\max_{1\le j\le p+1} |u_j - v_j| < \delta$, $\mathbf{u}, \mathbf{v} \in \mathcal{C}_1$ entails $|f(\mathbf{u}) - f(\mathbf{v})| < \epsilon/2$. Hence, from (A.1) we have that, for $\omega \notin \mathcal{N}$ and n large enough, $\max_{0\le j\le p} \sup_{x\in\mathcal{S}_{\mathcal{H}}} |\widehat{\eta}_j(x)| < \delta$. Therefore, for $1 \le i \le n$, we have that

$$|f(r_i + \widehat{\eta}_0(X_i), \mathbf{u}_i + \widehat{\eta}(X_i)) - f(r_i, \mathbf{u}_i)| \, \mathbf{1}_{\mathcal{K}}(r_i, \mathbf{u}_i, X_i) < \frac{\epsilon}{2},$$

implying $A_{1n} < \epsilon/2$. Hence, $|\mathrm{E}_{Q_n}(f) - \mathrm{E}_{P_n}(f)| < \epsilon$ for n large enough and $\omega \notin \mathcal{N}$, concluding the proof. □

Proof of Theorem 3.1. Let $P_n(A) = (1/n)\sum_{i=1}^{n} \mathbf{1}_A(r_i, \mathbf{u}_i)$. From (7) and Lemma 3.1, we have that $|\mathrm{E}_{\widehat{P}_n}(f) - \mathrm{E}_{P_n}(f)| \xrightarrow{a.s.} 0$, additionally, as $\pi(P_n, P) \xrightarrow{a.s.} 0$ we get that $\pi(\widehat{P}_n, P) \xrightarrow{a.s.} 0$.

Using that $\boldsymbol{\beta}(H)$ is continuous at P, we obtain that $\boldsymbol{\beta}(\widehat{P}_n) \xrightarrow{a.s.} \boldsymbol{\beta}(P)$. The result follows now from the fact that $\widehat{\boldsymbol{\beta}}_{\mathrm{R}} = \boldsymbol{\beta}(\widehat{P}_n)$ and $\boldsymbol{\beta}(P) = \boldsymbol{\beta}$ since $r_i = \boldsymbol{\beta}^{\top}\mathbf{u}_i + \epsilon_i$. □

The proof of Proposition 3.2 uses arguments similar to those considered in [14]. We include its proof for completeness.

Proof of Proposition 3.2. For notation simplicity, denote $Z_{i0} = Y_i$, $u_{i0} = r_i$ and $W_i(x) = K_i(x)/\{n\mathrm{E}K_1(x)\}$. Then, using that $\widehat{\phi}_j(x)$ satisfies $\widehat{\lambda}_j(x, a, \widehat{s}_j) = 0$, we get that, for $0 \le j \le p$,

$$\sum_{i=1}^{n} W_i(x)\psi\left(\frac{Z_{ij} - \phi_j(x)}{\widehat{s}_j}\right) = \left(\frac{\widehat{\phi}_j(x) - \phi_j(x)}{\widehat{s}_j}\right)\left[\sum_{i=1}^{n} W_i(x)\psi'\left(\frac{Z_{ij} - \phi_j(x)}{\widehat{s}_j}\right) - \frac{1}{2}\sum_{i=1}^{n} W_i(x)\psi''\left(\frac{Z_{ij} - \widetilde{\xi}_j(x)}{\widehat{s}_j}\right)\left\{\frac{\widehat{\phi}_j(x) - \phi_j(x)}{\widehat{s}_j}\right\}\right],$$

where $\widetilde{\xi}_j(x)$ is an intermediate point between $\widehat{\phi}_j(x)$ and $\phi_j(x)$. Hence,

$$\widehat{\phi}_j(x) - \phi(x) = \widehat{s}_j \frac{1}{\widehat{A}_j(x)} \sum_{i=1}^{n} W_i(x)\psi\left(\frac{Z_{ij} - \phi_j(x)}{\widehat{s}_j}\right), \tag{A.2}$$

where

$$\widehat{A}_j(x) = \sum_{i=1}^{n} W_i(x)\psi'\left(\frac{Z_{ij} - \phi_j(x)}{\widehat{s}_j}\right) - \frac{1}{2}\sum_{i=1}^{n} W_i(x)\psi''\left(\frac{Z_{ij} - \widetilde{\xi}_j(x)}{\widehat{s}_j}\right)\left\{\frac{\widehat{\phi}_j(x) - \phi(x)}{\widehat{s}_j}\right\}.$$

Denote as $\widetilde{R}_0(x) = \sum_{i=1}^{n} K_i(x)/\{n\,\mathrm{E}K_1(x)\} = \sum_{i=1}^{n} W_i(x)$. Then, $\mathrm{E}\widetilde{R}_0(x) = 1$ which together with the fact that $n\phi(h)/\ln n \to \infty$ entails that $\sup_{x\in S_{\mathcal{H}}} |\widetilde{R}_0(x) - \mathrm{E}\widetilde{R}_0(x)| \xrightarrow{a.co.} 0$, so $\sup_{x\in\mathcal{S}_{\mathcal{H}}} |\widetilde{R}_0(x)| \le 2$ almost surely. Hence, using that the kernel is non-negative and (12), we obtain the bound

$$\sup_{x\in\mathcal{S}_{\mathcal{H}}} \left|\frac{1}{2}\sum_{i=1}^{n} W_i(x)\psi''\left(\frac{Z_{ij} - \widetilde{\xi}_j(x)}{\widehat{s}_j}\right)\{\widehat{\phi}_j(x) - \phi_j(x)\}\right| \le \|\psi''\|_\infty \sup_{x\in\mathcal{S}_{\mathcal{H}}} |\widehat{\phi}_j(x) - \phi_j(x)| = O_{a.co.}(h^\eta + \theta_n),$$

hence,

$$\sup_{x\in\mathcal{S}_{\mathcal{H}}}\left|\widehat{A}_j(x)-\sum_{i=1}^n W_i(x)\psi'\left(\frac{Z_{ij}-\phi_j(x)}{\widehat{s}_j}\right)\right| = O_{a.co.}(h^\eta+\theta_n).$$

Note that, since $u_{ij}=Z_{ij}-\phi_j(X_i)$ is independent of X_i and has a symmetric distribution around 0, $\mathrm{E}\left\{\psi'\left(u_{ij}/\sigma\right)|X_i=x\right\}=\nu_j(\sigma)$ and $\mathrm{E}\left\{\psi\left(u_{ij}/\sigma\right)|X_i=x\right\}=\mathrm{E}\psi\left(u_{ij}/\sigma\right)=0$. Moreover, using that $Z_{ij}-\phi_j(x)=u_{ij}+\phi_j(X_i)-\phi_j(x)$ for any σ, **H2**, **N4**, **N5** and similar arguments to those considered in the proof of Theorem 4.1 in Boente and Vahnovan [16] (see also Ferraty et al. [22]), we obtain that for $0<a<b$

$$\sup_{x\in\mathcal{S}_{\mathcal{H}}}\sup_{\sigma\in[a,b]}\left|\sum_{i=1}^n W_i(x)\psi'\left(\frac{Z_{ij}-\phi_j(x)}{\sigma}\right)-\nu_j(\sigma)\right| = O_{a.co.}(h^\eta+\theta_n),$$

$$\sup_{x\in\mathcal{S}_{\mathcal{H}}}\sup_{\sigma\in[a,b]}\left|\sum_{i=1}^n W_i(x)\psi\left(\frac{Z_{ij}-\phi_j(x)}{\sigma}\right)\right| = O_{a.co.}(h^\eta+\theta_n).$$

Using that $\widehat{s}_j\xrightarrow{a.s.}s_j$, we get that $\sup_{x\in\mathcal{S}_{\mathcal{H}}}|\widehat{A}_j(x)-\nu_j(\widehat{s}_j)|=O_{a.s.}(h^\eta+\theta_n)$. It is worth noting that the fact that $\nu_j(s_j)\neq 0$ together with the strong consistency of $\widehat{s}_j$ entails that $\nu_j(\widehat{s}_j)\neq 0$ almost surely, for n large enough. Thus, if we denote as $\widehat{B}_j(x)=\widehat{A}_j(x)^{-1}-\nu_j(\widehat{s}_j)^{-1}$, we have that

$$\sup_{x\in\mathcal{S}_{\mathcal{H}}}|\widehat{B}_j(x)| = O_{a.s.}(h^\eta+\theta_n), \tag{A.3}$$

$$\sup_{x\in\mathcal{S}_{\mathcal{H}}}\left|\sum_{i=1}^n W_i(x)\psi\left(\frac{Z_{ij}-\phi_j(x)}{\widehat{s}_j}\right)\right| = O_{a.co.}(h^\eta+\theta_n). \tag{A.4}$$

Note that (A.2) entails that

$$\widehat{\phi}_j(x)-\phi(x)=\frac{\widehat{s}_j}{\nu_j(\widehat{s}_j)}\sum_{i=1}^n W_i(x)\psi\left(\frac{Z_{ij}-\phi_j(x)}{\widehat{s}_j}\right)+\widehat{B}_j(x)\,\widehat{s}_j\sum_{i=1}^n W_i(x)\psi\left(\frac{Z_{ij}-\phi_j(x)}{\widehat{s}_j}\right),$$

which together with (A.3) and (A.4) leads to

$$\sup_{x\in\mathcal{S}_{\mathcal{H}}}\left|\widehat{\phi}_j(x)-\phi_j(x)-\frac{\widehat{s}_j}{\nu_j(\widehat{s}_j)}\sum_{i=1}^n W_i(x)\psi\left(\frac{Z_{ij}-\phi_j(x)}{\widehat{s}_j}\right)\right| = O_{a.s.}(h^{2\eta}+\theta_n^2),$$

concluding the proof. □

The following lemmas will be used in the proof of Theorem 3.2. Even if their proofs follow the same steps as that of Lemma A.1 in [14], we include them due to the differences appearing in the functional setting.

Lemma A.1. *Assume that (9) and **H1** to **H3**, **H5**, **N1**, **N2**, **N4** and **N5** and **N8** hold. Moreover, assume that $\Pr(X\in\mathcal{S}_{\mathcal{H}})=1$ with $\mathcal{S}_{\mathcal{H}}$ a compact set and that the sequence $h=h_n$ is such that $nh^{4\eta}\to 0$ as $n\to\infty$. Let $\widehat{s}_\ell$ and $\widehat{\sigma}$ be consistent estimators of s_ℓ and σ_0, respectively. Denote as $W_i(x)=K_i(x)/\{n\mathrm{E}K_1(x)\}$ with $K_i(x)=K\left(d(x,X_i)/h\right)$ and $u_{i0}=r_i$. For any fixed $0\le\ell\le p$, $1\le j\le p$, let $\widehat{R}(\sigma,s)=(1/n)\sum_{1\le i,m\le n}H_{im}(\sigma,s)$ where*

$$H_{im}(\sigma,s)=W_m(X_i)\psi_1'\left(\frac{\epsilon_i}{\sigma}\right)w_2(\|\mathbf{u}_i\|)\,u_{ij}\psi'\left(\frac{u_{m\ell}}{s}\right)\{\phi_\ell(X_m)-\phi_\ell(X_i)\}.$$

Denote $\mathcal{I}=[\sigma_0/2,2\,\sigma_0]$ and $\mathcal{I}_\ell=[s_\ell/2,2\,s_\ell]$, then, we have that

(a) *there exists a constant $C>0$ not depending on n such that for all $n\ge n_0$*

$$\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_j}\Pr\{\sqrt{n}\,|\widehat{R}(\sigma,s)|>\epsilon\}\le Ch^{2\eta},$$

(b) $\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_\ell}\sqrt{n}\left|\widehat{R}(\sigma,s)\right|=o_{\mathbb{P}}(1)$,

(c) $\sqrt{n}\,\widehat{R}(\widehat{\sigma},\widehat{s}_\ell)\xrightarrow{p}0$.

Proof. (a) Distinguishing the situation $\ell=0$ and $\ell\neq 0$, it is easy to see that $\mathrm{E}H_{im}(\sigma,s)=0$ since $\mathbf{u}_i$ satisfy **N8**, the errors ϵ_i have a symmetric distribution, ψ' and ψ_1' are even function and $(\epsilon_i,\mathbf{u}_i)$ is independent of X_i. Then, $\mathrm{E}\widehat{R}(\sigma,s)=0$ for all $\sigma,s>0$, so Markov's inequality entails that

$$\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_\ell}\Pr\{\sqrt{n}|\widehat{R}(\sigma,s)|>\epsilon\}\le\frac{n}{\epsilon^2}\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_\ell}\mathrm{var}\left\{\widehat{R}(\sigma,s)\right\}.$$

Then, (a) follows if we show that for some constant $C_1 > 0$ and n large enough

$$\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_\ell} \operatorname{var}\left\{\widehat{R}(\sigma,s)\right\} \le C_1 \frac{h^{2\eta}}{n}. \tag{A.5}$$

Denote $\lambda_1 = \mathrm{E}\left\{\psi_1'(\epsilon_1/\sigma)^2\right\}\mathrm{E}\left\{w_2^2(\|\mathbf{u}_1\|)\,u_{1j}^2\right\}\max\left[\left\{\mathrm{E}\psi'(u_{2\ell}/s)\right\}^2, \mathrm{E}\left\{\psi'(u_{2\ell}/s)^2\right\}\right]$ and note that, for $i\neq i'$, $i=m$ or $i'=m'$, $\operatorname{cov}\{H_{im}(\sigma,s),H_{i'm'}(\sigma,s)\}=0$. Then, using the fact that ϕ_ℓ is Lipschitz of order η and the kernel is non-negative with bounded support, we get that

$$\operatorname{cov}\{H_{im}(\sigma,s),H_{im'}(\sigma,s)\} \le \lambda_1\, h^{2\eta}\,\mathrm{E}\left\{W_m(X_i)W_{m'}(X_i)\right\}.$$

Recall that since $\mathrm{E}\sum_{m=1}^n W_m(x) = 1$ and $n\phi(h)/\ln n \to \infty$, we have that $\sup_{x\in S_{\mathcal{H}}}|\sum_{m=1}^n W_m(x)-1| \xrightarrow{a.co.} 0$, so that $\sup_{x\in\mathcal{S}_{\mathcal{H}}}|\sum_{m=1}^n W_m(x)|\le 2$ almost surely. Then, we get the bound

$$\begin{aligned}\operatorname{var}\{\widehat{R}(\sigma,s)\} &= \frac{1}{n^2}\sum_{i=1}^n\sum_{m=1}^n\sum_{m'=1}^n \operatorname{cov}\{H_{im}(\sigma,s),H_{im'}(\sigma,s)\} \le \frac{\lambda_1\,h^{2\eta}}{n^2}\sum_{i=1}^n \mathrm{E}\left\{\sum_{m=1}^n W_m(X_i)\sum_{m'=1}^n W_{m'}(X_i)\right\}\\ &\le \frac{\lambda_1\,h^{2\eta}}{n^2}\sum_{i=1}^n \mathrm{E}\left[\left\{\sum_{m=1}^n W_m(X_i)\right\}^2\right] \le \frac{\lambda_1\,h^{2\eta}}{n}\mathrm{E}\left[\sup_{x\in\mathcal{S}_{\mathcal{H}}}\left\{\sum_{m=1}^n W_m(x)\right\}^2\right] \le 4\,\lambda_1\frac{h^{2\eta}}{n},\end{aligned}$$

concluding the proof of (A.5).

(b) For any fixed ρ, denote as $N_{\ell,\rho}$ the minimum number of intervals $\mathcal{I}_{\ell,k}=[a_k^{(\ell)}-\rho, a_k^{(\ell)}+\rho]$ needed to cover $\mathcal{I}_\ell$ with $a_k^{(\ell)}\in\mathcal{I}_\ell$. Similarly, let N_ρ be the minimum number of intervals $\mathcal{I}_k=[a_k-\rho,a_k+\rho]$ needed to cover $\mathcal{I}$ with $a_k\in\mathcal{I}$. Then, we have that $\mathcal{I}_\ell\subset\bigcup_{k=1}^{N_{\ell,\rho}}\mathcal{I}_{\ell,k}$ and $\mathcal{I}\subset\bigcup_{k=1}^{N_\rho}\mathcal{I}_k$. As it is well known, $N_{\ell,\rho}\le A/\rho$ and $N_\rho\le A/\rho$, where the constant A does not depend on ρ. Then,

$$\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_\ell}|\widehat{R}(\sigma,s)| \le \max_{\substack{1\le k\le N_\rho\\1\le k_\ell\le N_{\ell,\rho}}}|\widehat{R}(a_k,a_{k_\ell}^{(\ell)})| + \max_{\substack{1\le k\le N_\rho\\1\le k_\ell\le N_{j,\rho}}}\sup_{\substack{\sigma\in\mathcal{I}\cap\mathcal{I}_k\\ s\in\mathcal{I}_\ell\cap\mathcal{I}_{\ell,k_\ell}}}\left|\widehat{R}(\sigma,s)-\widehat{R}(a_k,a_{k_\ell}^{(\ell)})\right|. \tag{A.6}$$

Let us begin bounding $\widehat{R}(\sigma,s)-\widehat{R}(a_k,a_{k_\ell}^{(\ell)})$. Denote as $V_{im}=W_m(X_i)w_2(\|\mathbf{u}_i\|)\,u_{ij}\{\phi_\ell(X_m)-\phi_\ell(X_i)\}$, then we have that

$$\begin{aligned}H_{im}(\sigma,s)-H_{im}(a_k,a_{k_\ell}^{(\ell)}) &= V_{im}\left\{\psi_1'\left(\frac{\epsilon_i}{\sigma}\right)\psi'\left(\frac{u_{m\ell}}{s}\right)-\psi_1'\left(\frac{\epsilon_i}{a_k}\right)\psi'\left(\frac{u_{m\ell}}{a_{k_\ell}^{(\ell)}}\right)\right\}\\ &= V_{im}\left[\psi_1'\left(\frac{\epsilon_i}{\sigma}\right)\left\{\psi'\left(\frac{u_{m\ell}}{s}\right)-\psi'\left(\frac{u_{m\ell}}{a_{k_\ell}^{(\ell)}}\right)\right\}+\psi'\left(\frac{u_{m\ell}}{a_{k_\ell}^{(\ell)}}\right)\left\{\psi_1'\left(\frac{\epsilon_i}{\sigma}\right)-\psi_1'\left(\frac{\epsilon_i}{a_k}\right)\right\}\right].\end{aligned}$$

Using that for $\sigma\in\mathcal{I}\cap\mathcal{I}_k$ and $s\in\mathcal{I}_\ell\cap\mathcal{I}_{\ell,k_\ell}$, we have that

$$\begin{aligned}\left|\psi_1'\left(\frac{\epsilon_i}{\sigma}\right)-\psi_1'\left(\frac{\epsilon_i}{a_k}\right)\right| &\le \|\zeta_1\|_\infty\frac{2}{\sigma_0}|\sigma-a_k| \le \rho\,\|\zeta_1\|_\infty\frac{2}{\sigma_0},\\ \left|\psi'\left(\frac{u_{m\ell}}{s}\right)-\psi'\left(\frac{u_{m\ell}}{a_{k_\ell}^{(\ell)}}\right)\right| &\le \|\zeta\|_\infty\frac{2}{s_\ell}|s-a_{k_\ell}^{(\ell)}| \le \rho\,\|\zeta\|_\infty\frac{2}{s_\ell}.\end{aligned}$$

Hence, the boundedness of ψ', ψ_1' and ψ_2 and the fact that ϕ_ℓ is Lipschitz of order η and the kernel is non-negative with bounded support, leads to

$$\left|H_{im}(\sigma,s)-H_{im}(a_k,a_{k_\ell}^{(\ell)})\right| \le W_m(X_i)h^\eta\rho\|\psi_2\|_\infty\left(\|\psi_1'\|_\infty\|\zeta\|_\infty\frac{2}{s_\ell}+\|\psi'\|_\infty\|\zeta\|_\infty\frac{2}{\sigma_0}\right) = B\,h^\eta\rho\,W_m(X_i),$$

which together with the fact that $\sup_{x\in\mathcal{S}_{\mathcal{H}}}|\sum_{m=1}^n W_m(x)|\le 2$ almost surely, entails

$$\left|\widehat{R}(\sigma,s)-\widehat{R}(a_k,a_{k_\ell}^{(\ell)})\right| \le \frac{1}{n}\sum_{1\le i,m\le n}\left|H_{im}(\sigma,s)-H_{im}(a_k,a_{k_\ell}^{(\ell)})\right| \le B\,h^\eta\rho\frac{1}{n}\sum_{1\le i\le n}\sum_{m=1}^n W_m(X_i) \le 2B\,h^\eta\rho.$$

Let $B_1=\epsilon/(4B)$ and choose $\rho=B_1 n^{-1/4}$ then using that $nh^{4\eta}\to 0$ we have that for n large enough $nh^{4\eta}\le 1$, so

$$\sqrt{n}\left|\widehat{R}(\sigma,s)-\widehat{R}(a_k,a_{k_\ell}^{(\ell)})\right| \le 2B\,n^{1/2}\,h^\eta\rho \le \frac{\epsilon}{2}n^{1/4}\,h^\eta \le \frac{\epsilon}{2}, \tag{A.7}$$

almost surely. Hence, from (A.6) and (A.7) and the fact that $N_\rho \le An^{1/4}$ and $N_{\ell,\rho} \le An^{1/4}$, for some constant A, we obtain

$$\Pr\left\{\sqrt{n}\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_\ell}|\widehat{R}(\sigma,s)|>\epsilon\right\}\le \Pr\left\{\sqrt{n}\max_{\substack{1\le k\le N_\rho\\ 1\le k_\ell\le N_{\ell,\rho}}}|\widehat{R}(a_k,a_{k_\ell}^{(\ell)})|>\frac{\epsilon}{2}\right\}$$
$$\le N_\rho N_{\ell,\rho}\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_\ell}\Pr\{\sqrt{n}|\widehat{R}(\sigma,s)|>\epsilon\}\le A^2Cn^{1/2}h^{2\eta}=A^2C\left(nh^{4\eta}\right)^{1/2},$$

concluding the proof of (b) since $nh^{4\eta}\to 0$.

(c) follows immediately from (b) using the consistency of $\widehat{\sigma}$ and $\widehat{s}_\ell$. □

Lemma A.2. *Let $\chi:\mathbb{R}\to\mathbb{R}$ and $\upsilon:\mathbb{R}\to\mathbb{R}$ be bounded and differentiable functions with bounded derivatives such that χ is an even function, υ is an odd function and $\chi_1(t)=t\chi'(t)$ and $\upsilon_1(t)=t\upsilon'(t)$ are bounded. Assume that model (9), **H1** to **H3**, **H5**, **N2** and **N8** hold and that $n^{1/2}\phi(h)\to 0$. Moreover, assume that $\Pr(X\in\mathcal{S}_{\mathcal{H}})=1$ with $\mathcal{S}_{\mathcal{H}}$ a compact set. Denote $u_{i0}=r_i$ and $W_i(x)=K_i(x)/\{n\mathrm{E}K_1(x)\}$ with $K_i(x)=K(d(x,X_i)/h)$. For any fixed $0\le\ell\le p$, $1\le j\le p$, let $\widehat{R}(\sigma,s)=(1/n)\sum_{1\le i,m\le n}H_{im}(\sigma,s)$ where*

$$H_{im}(\sigma,s)=W_m(X_i)\chi\left(\frac{\epsilon_i}{\sigma}\right)w_2(\|\mathbf{u}_i\|)\,u_{ij}\,\upsilon\left(\frac{u_{m\ell}}{s}\right).$$

Let $\mathcal{I}=[\sigma_0/2,2\sigma_0]$ and $\mathcal{I}_\ell=[s_\ell/2,2s_\ell]$, then,

(a) *For $\ell=j$ or $\ell=0$, $\sup_{\sigma\in\mathcal{I}}\sup_{s\in\mathcal{I}_\ell}n^{1/2}|\mathrm{E}\widehat{R}(\sigma,s)|\to 0$, while, for $\ell\ne j$, $\ell\ne 0$ $\mathrm{E}\widehat{R}(\sigma,s)=0$. In particular, for $\ell=0$, $\sup_{\sigma\in\mathcal{I}}\sup_{s\in\mathcal{I}_\ell}n^{1/2}|\mathrm{E}\widehat{R}(\sigma,s)|\to 0$, for any $1\le j\le p$.*

(b) *There exists a constant $C>0$ not depending on n such that, for all $n\ge n_0$,*

$$\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_j}\Pr\{n^\tau|\widehat{R}(\sigma,s)-\mathrm{E}\widehat{R}(\sigma,s)|>\epsilon\}\le\begin{cases}C\dfrac{n^{2\tau}}{n^2\phi(h)} & \text{for } \ell\ne 0\\ C\dfrac{n^{2\tau}}{n^2\phi^2(h)} & \text{for } \ell=0.\end{cases}$$

(c) *For any fixed (σ,s), $n^{1/2}\widehat{R}(\sigma,s)\xrightarrow{p}0$.*

(d) *$\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_\ell}n^{1/4}\left|\widehat{R}(\sigma,s)\right|=o_{\mathbb{P}}(1)$. Hence, if $\widehat{s}_\ell$ and $\widehat{\sigma}$ are consistent estimators of s_ℓ and σ_0, respectively, we have that $n^{1/4}\widehat{R}(\widehat{\sigma},\widehat{s}_\ell)\xrightarrow{p}0$.*

Proof. (a) Note that $\mathrm{E}H_{im}(\sigma,s)=0$ when $m\ne i$, for any ℓ,j. Besides, if $m=i$ and $\ell\ne j$, $\ell\ne 0$ we also have that $\mathrm{E}H_{im}(\sigma,s)=0$ from **N8** and the oddness of υ, so that $\mathrm{E}\widehat{R}(\sigma,s)=0$ for $\ell\ne j$, $\ell\ne 0$.

When $\ell=j$ or $\ell=0$, using that $\mathrm{E}H_{im}(\sigma,s)=0$ when $m\ne i$, we get that

$$\mathrm{E}\widehat{R}(\sigma,s)=\frac{1}{n}\sum_{i=1}^n\mathrm{E}H_{ii}(\sigma,s)=\mathrm{E}H_{22}(\sigma,s)=\mathrm{E}W_2(X_2)\mathrm{E}\left\{\chi\left(\frac{\epsilon_2}{\sigma}\right)w_2(\|\mathbf{u}_2\|)\,u_{2j}\upsilon\left(\frac{u_{2\ell}}{s}\right)\right\},$$

where $W_2(X_2)=K(0)/\{n\mathrm{E}K_1(X_2)\}$ and $u_{2\ell}=\boldsymbol{\beta}^\top\mathbf{u}_2+\epsilon_2$ if $\ell=0$. Recall that Lemma 4.3 and 4.4 in Ferraty and Vieu [27], **H1** and **H2** imply that there are constants $0<C<C'<\infty$ such that

$$C\phi(h)<\mathrm{E}K_1(x)<C'\phi(h)\quad\text{for any } x\in S_{\mathcal{H}}. \tag{A.8}$$

Therefore, when $\ell=j$ or $\ell=0$, using that χ, ψ_2 and υ are bounded, we get the following bound for $\mathrm{E}H_{22}(\sigma,s)$

$$|\mathrm{E}H_{22}(\sigma,s)|=\left|\frac{K(0)}{n}\mathrm{E}\left\{\frac{1}{\mathrm{E}K_1(X_2)}\right\}\mathrm{E}\left\{\chi\left(\frac{\epsilon_1}{\sigma}\right)w_2(\|\mathbf{u}_1\|)\,u_{1j}\upsilon\left(\frac{u_{1j}}{s}\right)\right\}\right|\le\frac{C}{n\phi(h)},$$

which entails that

$$\sup_{\sigma\in\mathcal{I}}\sup_{s\in\mathcal{I}_\ell}n^{1/2}|\mathrm{E}\widehat{R}(\sigma,s)|\le\frac{C}{n^{\frac{1}{2}}\phi(h)}\to 0,$$

since $n^{1/2}\phi(h)\to\infty$.

(b) As in Lemma A.1, Markov's inequality implies that

$$\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_\ell}\Pr\left\{n^\tau|\widehat{R}(\sigma,s)-\mathrm{E}\widehat{R}(\sigma,s)|>\epsilon\right\}\le\frac{n^{2\tau}}{\epsilon^2}\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_\ell}\mathrm{var}\left\{\widehat{R}(\sigma,s)\right\}.$$

Then, (b) follows if we show that for some constant $C_1 > 0$ and n large enough

$$\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_\ell} \text{var}\left\{\widehat{R}(\sigma,s)\right\} \le \begin{cases} C_1 \dfrac{1}{n^2\phi(h)} & \text{for } \ell \neq 0 \\ C_1 \dfrac{1}{n^2\phi^2(h)} & \text{for } \ell = 0. \end{cases} \tag{A.9}$$

Let us consider the situation, $\ell = 0$. In this case, we have that $\text{cov}\{H_{im}(\sigma,s), H_{i'm'}(\sigma,s)\} \neq 0$ only in the following three situations, $i = i'$ and $m = m'$ or $i = m$ and $i' = m'$ or $i = m'$ and $m = i'$. Then, if we denote as $\lambda_1 = \|\psi_2^2\|_\infty \|\chi^2\|_\infty \|\upsilon^2\|_\infty$ using that the kernel is non-negative with bounded support, $\text{E}H_{im}(\sigma,s) = 0$, for $m \neq i$ and that $W_i(X_i) = K(0)/\{n\text{E}K_1(x)\}$ together with (A.8), we get that

$$\begin{aligned}
\text{var}\{H_{im}(\sigma,s)\} &\le \text{E}H_{im}^2(\sigma,s) \le \lambda_1 \,\text{E}\left\{W_m^2(X_i)\right\},\\
|\text{cov}\{H_{ii}(\sigma,s), H_{i'i'}(\sigma,s)\}| &\le [\text{var}\{H_{ii}(\sigma,s)\}\ \text{var}\{H_{i'i'}(\sigma,s)\}]^{1/2} \le \lambda_1 \frac{C^{-1}K^2(0)}{n^2\phi^2(h)},\\
|\text{cov}\{H_{im}(\sigma,s), H_{mi}(\sigma,s)\}| &\le [\text{var}\{H_{im}(\sigma,s)\}\ \text{var}\{H_{mi}(\sigma,s)\}]^{1/2} \le \lambda_1 \,\text{E}\left\{W_m^2(X_i)\right\}.
\end{aligned}$$

Recall that since $\text{E}\sum_{m=1}^n W_m(x) = 1$ and $n\phi(h)/\ln n \to \infty$, we have that $\sup_{x\in\mathcal{S}_\mathcal{H}} |\sum_{m=1}^n W_m(x)| \le 2$ almost surely. Then, using again that $W_m(x) \le C^{-1}\|K\|_\infty/\{n\phi(h)\}$, we get the bound

$$\begin{aligned}
\text{var}\{\widehat{R}(\sigma,s)\} &= \frac{1}{n^2}\sum_{i=1}^n\sum_{m=1}^n \text{var}\{H_{im}(\sigma,s)\} + \frac{1}{n^2}\sum_{i=1}^n\sum_{m=1,m\neq i}^n \text{cov}\{H_{im}(\sigma,s), H_{mi}(\sigma,s)\}\\
&\quad + \frac{1}{n^2}\sum_{i=1}^n\sum_{i'=1,i'\neq i}^n \text{cov}\{H_{ii}(\sigma,s), H_{i'i'}(\sigma,s)\}\\
&\le 2\frac{\lambda_1}{n^2}\sum_{i=1}^n \text{E}\left\{\sum_{m=1}^n W_m^2(X_i)\right\} + \lambda_1\frac{C^{-1}K^2(0)}{n^2\phi^2(h)}\\
&\le 2\,C^{-1}\frac{\|K\|_\infty\,\lambda_1}{n^3\phi(h)}\sum_{i=1}^n \text{E}\left\{\sum_{m=1}^n W_m(X_i)\right\} + \frac{C^{-1}K^2(0)\,\lambda_1}{n^2\phi^2(h)}\\
&\le \frac{4\,C^{-1}\|K\|_\infty\,\lambda_1}{n^2\phi(h)} + \frac{C^{-1}K^2(0)\,\lambda_1}{n^2\phi^2(h)} \le \frac{C^{-1}\lambda_1\{4\,\|K\|_\infty + K^2(0)\}}{n^2\phi^2(h)},
\end{aligned}$$

since $\phi(h) < 1$, concluding the proof of (A.9), when $\ell = 0$.

When $\ell \neq 0$, we distinguish the cases $\ell = j$ and $\ell \neq j$.

For $\ell \neq j$, $\text{E}H_{im}(\sigma,s) = 0$, so $\text{cov}\{H_{im}(\sigma,s), H_{i'm'}(\sigma,s)\} = \text{E}\{H_{im}(\sigma,s)H_{i'm'}(\sigma,s)\}$. In this case, it is easy to see that $\text{cov}\{H_{im}(\sigma,s), H_{i'm'}(\sigma,s)\} \neq 0$ only when $i = i'$ and $m = m'$, in which case we have that

$$\text{cov}\{H_{im}(\sigma,s), H_{i'm'}(\sigma,s)\} = \begin{cases} \lambda_1\text{E}\left\{W_m^2(X_i)\right\} & \text{for } i = i',\ m = m' \text{ and } m \neq i,\\ \lambda_1\text{E}\left\{W_m^2(X_i)\right\} & \text{for } i = i' = m = m',\\ 0 & \text{otherwise.} \end{cases}$$

Hence, as above we have that

$$\text{var}\{\widehat{R}(\sigma,s)\} = \frac{1}{n^2}\sum_{i=1}^n\sum_{m=1}^n \text{var}\{H_{im}(\sigma,s)\} \le \lambda_1\frac{1}{n^2}\sum_{i=1}^n \text{E}\left\{\sum_{m=1}^n W_m^2(X_i)\right\} \le \lambda_1\frac{2\,\|K\|_\infty}{n^2\phi(h)}C^{-1},$$

as desired.

When $\ell = j$, $\text{cov}\{H_{im}(\sigma,s), H_{i'm'}(\sigma,s)\} = \text{E}\{H_{im}(\sigma,s)H_{i'm'}(\sigma,s)\}$, for $m \neq i$ and $m' \neq i'$. Straightforward calculations allow to see that $\text{cov}\{H_{im}(\sigma,s), H_{i'm'}(\sigma,s)\} \neq 0$ only when $(i,m) = (i',m')$ or $(i,m) = (m',i')$, in which case, it can be bounded as $\text{var}\{H_{im}(\sigma,s)\} \le \lambda_1\text{E}\left\{W_m^2(X_i)\right\}$ and $\text{cov}\{H_{im}(\sigma,s), H_{mi}(\sigma,s)\} = \lambda_2\text{E}\left\{W_m^2(X_i)\right\}$, for $i \neq m$, with $\lambda_2 = \{\text{E}\chi(\epsilon_1/\sigma)\}^2\left\{\text{E}\upsilon(u_{1j}/s)w_2(\|\mathbf{u}_1\|)\,u_{1j}\right\}^2$. Therefore,

$$\begin{aligned}
\text{var}\{\widehat{R}(\sigma,s)\} &= \frac{1}{n^2}\sum_{i=1}^n\sum_{m=1}^n \text{var}\{H_{im}(\sigma,s)\} + \frac{1}{n^2}\sum_{i=1}^n\sum_{m=1}^n \text{cov}\{H_{im}(\sigma,s), H_{mi}(\sigma,s)\}\\
&\le (\lambda_1+\lambda_2)\frac{1}{n^2}\sum_{i=1}^n \text{E}\left\{\sum_{m=1}^n W_m^2(X_i)\right\} \le (\lambda_1+\lambda_2)\frac{2\,\|K\|_\infty}{n^2\phi(h)},
\end{aligned}$$

concluding the proof of (b).

(c) Follows immediately from (a) and (b)

(d) As in Lemma A.1, for any fixed ρ, denote as $N_{\ell,\rho}$ the minimum number of intervals $\mathcal{I}_{\ell,k} = [a_k^{(\ell)} - \rho, a_k^{(\ell)} + \rho]$ needed to cover $\mathcal{I}_\ell$ with $a_k^{(\ell)} \in \mathcal{I}_\ell$. Similarly, let N_ρ be the minimum number of intervals $\mathcal{I}_k = [a_k - \rho, a_k + \rho]$ needed to cover $\mathcal{I}$ with $a_k \in \mathcal{I}$. Then, we have that $\mathcal{I}_\ell \subset \bigcup_{k=1}^{N_{\ell,\rho}} \mathcal{I}_{\ell,k}$ and $\mathcal{I} \subset \bigcup_{k=1}^{N_\rho} \mathcal{I}_k$ and as it is well known $N_{\ell,\rho} \le A/\rho$ and $N_\rho \le A/\rho$, where the constant A does not depend on ρ. Then,

$$\sup_{\sigma\in\mathcal{I},\, s\in\mathcal{I}_\ell} |\widehat{R}(\sigma,s) - \mathrm{E}\widehat{R}(\sigma,s)| \le \max_{\substack{1\le k\le N_\rho \\ 1\le k_\ell\le N_{\ell,\rho}}} |\widehat{R}(a_k, a_{k_\ell}^{(\ell)}) - \mathrm{E}\widehat{R}(a_k, a_{k_\ell}^{(\ell)})| + \max_{\substack{1\le k\le N_\rho \\ 1\le k_\ell\le N_{j,\rho}}} \sup_{\substack{\sigma\in\mathcal{I}\cap\mathcal{I}_k \\ s\in\mathcal{I}_\ell\cap\mathcal{I}_{\ell,k_\ell}}} \left|\widehat{R}(\sigma,s) - \widehat{R}(a_k, a_{k_\ell}^{(\ell)})\right|$$
$$+ \max_{\substack{1\le k\le N_\rho \\ 1\le k_\ell\le N_{j,\rho}}} \sup_{\substack{\sigma\in\mathcal{I}\cap\mathcal{I}_k \\ s\in\mathcal{I}_\ell\cap\mathcal{I}_{\ell,k_\ell}}} \left|\mathrm{E}\left\{\widehat{R}(\sigma,s) - \widehat{R}(a_k, a_{k_\ell}^{(\ell)})\right\}\right|. \tag{A.10}$$

Let us begin bounding $\widehat{R}(\sigma,s) - \widehat{R}(a_k, a_{k_\ell}^{(\ell)})$. Denote $V_{im} = W_m(X_i) w_2(\|\mathbf{u}_i\|)\, u_{ij}$, then we have that

$$H_{im}(\sigma,s) - H_{im}(a_k, a_{k_\ell}^{(\ell)}) = V_{im}\left\{\chi\left(\frac{\epsilon_i}{\sigma}\right)\upsilon\left(\frac{u_{m\ell}}{s}\right) - \chi\left(\frac{\epsilon_i}{a_k}\right)\upsilon\left(\frac{u_{m\ell}}{a_{k_\ell}^{(\ell)}}\right)\right\}$$
$$= V_{im}\left[\chi\left(\frac{\epsilon_i}{\sigma}\right)\left\{\upsilon\left(\frac{u_{m\ell}}{s}\right) - \upsilon\left(\frac{u_{m\ell}}{a_{k_\ell}^{(\ell)}}\right)\right\} + \upsilon\left(\frac{u_{m\ell}}{a_{k_\ell}^{(\ell)}}\right)\left\{\chi\left(\frac{\epsilon_i}{\sigma}\right) - \chi\left(\frac{\epsilon_i}{a_k}\right)\right\}\right].$$

Note that for $\sigma \in \mathcal{I}\cap\mathcal{I}_k$ and $s \in \mathcal{I}_\ell \cap \mathcal{I}_{\ell,k_\ell}$, we have that

$$\left|\chi\left(\frac{\epsilon_i}{\sigma}\right) - \chi\left(\frac{\epsilon_i}{a_k}\right)\right| \le \|\chi_1\|_\infty \frac{2}{\sigma_0}|\sigma - a_k| \le \rho\,\|\chi_1\|_\infty \frac{2}{\sigma_0},$$
$$\left|\upsilon\left(\frac{u_{m\ell}}{s}\right) - \upsilon\left(\frac{u_{m\ell}}{a_{k_\ell}^{(\ell)}}\right)\right| \le \|\upsilon_1\|_\infty \frac{2}{s_\ell}|s - a_{k_\ell}^{(\ell)}| \le \rho\,\|\upsilon_1\|_\infty \frac{2}{s_\ell}.$$

Hence, using the boundedness of χ, υ and ψ_2 and the fact that the kernel is non-negative with bounded support, we obtain that

$$\left|H_{im}(\sigma,s) - H_{im}(a_k, a_{k_\ell}^{(\ell)})\right| \le W_m(X_i)\,\rho\,\|\psi_2\|_\infty\left(\|\chi\|_\infty\|\upsilon_1\|_\infty\frac{2}{s_\ell} + \|\upsilon\|_\infty\|\chi_1\|_\infty\frac{2}{\sigma_0}\right) = B\,\rho\,W_m(X_i)$$

which together with the fact that $\sup_{x\in\mathcal{S}_\mathcal{H}} |\sum_{m=1}^n W_m(x)| \le 2$ almost surely, entails

$$\left|\widehat{R}(\sigma,s) - \widehat{R}(a_k, a_{k_\ell}^{(\ell)})\right| \le \frac{1}{n}\sum_{1\le i,m\le n}\left|H_{im}(\sigma,s) - H_{im}(a_k, a_{k_\ell}^{(\ell)})\right| \le B\,\rho\frac{1}{n}\sum_{1\le i\le n}\sum_{m=1}^n W_m(X_i) \le 2B\,\rho.$$

Let $B_1 = \epsilon/(8B)$ and choose $\rho = B_1 n^{-1/4}$. Then, we have that, for n large enough,

$$n^{1/4}\left|\widehat{R}(\sigma,s) - \widehat{R}(a_k, a_{k_\ell}^{(\ell)})\right| \le 2B\,n^{1/4}\rho \le \frac{\epsilon}{4}, \tag{A.11}$$

almost surely. Hence, using (A.10) and (A.11) and the fact that $N_\rho \le An^{1/4}$ and $N_{\ell,\rho} \le An^{1/4}$, for some constant A, together with (b) we obtain

$$\Pr\left\{n^{1/4}\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_\ell}|\widehat{R}(\sigma,s) - \mathrm{E}\widehat{R}(\sigma,s)| > \epsilon\right\} \le \Pr\left\{n^{1/4}\max_{\substack{1\le k\le N_\rho \\ 1\le k_\ell\le N_{\ell,\rho}}}|\widehat{R}(a_k, a_{k_\ell}^{(\ell)}) - \mathrm{E}\widehat{R}(a_k, a_{k_\ell}^{(\ell)})| > \frac{\epsilon}{2}\right\}$$
$$\le N_\rho\, N_{\ell,\rho}\sup_{\sigma\in\mathcal{I},s\in\mathcal{I}_\ell}\Pr\{n^{1/4}|\widehat{R}(\sigma,s) - \mathrm{E}\widehat{R}(\sigma,s)| > \epsilon\}$$
$$\le \begin{cases} A^2Cn^{1/2}\dfrac{n^{1/2}}{n^2\phi(h)} = A^2C\dfrac{1}{n\phi(h)} & \text{for } \ell \ne 0 \\[2ex] A^2Cn^{1/2}\dfrac{n^{1/2}}{n^2\phi^2(h)} = A^2C\dfrac{1}{n\phi^2(h)} & \text{for } \ell = 0, \end{cases}$$

concluding the proof of (c) since $n^{1/2}\phi(h) \to \infty$. □

Proof of Theorem 3.2. As in Bianco and Boente [9], write

$$\mathbf{L}_n(\sigma,\mathbf{b}) = \frac{\sigma}{n}\sum_{i=1}^{n}\psi_1\left(\frac{r_i-\mathbf{u}_i^{\top}\mathbf{b}}{\sigma}\right) w_2(\|\mathbf{u}_i\|)\mathbf{u}_i,$$

$$\widehat{\mathbf{L}}_n(\sigma,\mathbf{b}) = \frac{\sigma}{n}\sum_{i=1}^{n}\psi_1\left(\frac{\widehat{r}_i-\widehat{\mathbf{u}}_i^{\top}\mathbf{b}}{\sigma}\right) w_2(\|\widehat{\mathbf{u}}_i\|)\widehat{\mathbf{u}}_i.$$

Using a first order Taylor expansion around $\widehat{\boldsymbol{\beta}}$, we get

$$\widehat{\mathbf{L}}_n(\sigma,\boldsymbol{\beta}) = \widehat{\mathbf{L}}_n(\sigma,\widehat{\boldsymbol{\beta}}) + \frac{1}{n}\sum_{i=1}^{n}\psi_1'\left(\frac{\widehat{r}_i-\widehat{\mathbf{u}}_i^{\top}\widetilde{\boldsymbol{\beta}}}{\sigma}\right) w_2(\|\widehat{\mathbf{u}}_i\|)\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i^{\top}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}),$$

with $\widetilde{\boldsymbol{\beta}}$ an intermediate point between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$. From $\widehat{\mathbf{L}}_n(\widehat{\sigma},\widehat{\boldsymbol{\beta}})=0$ we get that $(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \mathbf{A}_n^{-1}\widehat{\mathbf{L}}_n(\widehat{\sigma},\boldsymbol{\beta})$ where

$$\mathbf{A}_n = \frac{1}{n}\sum_{i=1}^{n}\psi_1'\left(\frac{\widehat{r}_i-\widehat{\mathbf{u}}_i^{\top}\widetilde{\boldsymbol{\beta}}}{\widehat{\sigma}}\right) w_2(\|\widehat{\mathbf{u}}_i\|)\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i^{\top}.$$

Lemma 3.2 together with the consistency of $\widehat{\boldsymbol{\beta}}$ imply that $\mathbf{A}_n \xrightarrow{p} \mathbf{A}$ and therefore, as

$$\widehat{\mathbf{L}}_n(\widehat{\sigma},\boldsymbol{\beta}) = \{\widehat{\mathbf{L}}_n(\widehat{\sigma},\boldsymbol{\beta})-\mathbf{L}_n(\widehat{\sigma},\boldsymbol{\beta})\} + \{\mathbf{L}_n(\widehat{\sigma},\boldsymbol{\beta})-\mathbf{L}_n(\sigma_0,\boldsymbol{\beta})\} + \mathbf{L}_n(\sigma_0,\boldsymbol{\beta}),$$

from **N3** it will be enough to show that

(a) $n^{1/2}\mathbf{L}_n(\sigma_0,\boldsymbol{\beta}) \xrightarrow{D} \mathcal{N}(0,\boldsymbol{\Sigma})$,
(b) $n^{1/2}\{\widehat{\mathbf{L}}_n(\widehat{\sigma},\boldsymbol{\beta})-\mathbf{L}_n(\widehat{\sigma},\boldsymbol{\beta})\} \xrightarrow{p} 0$,
(c) $n^{1/2}\{\mathbf{L}_n(\widehat{\sigma},\boldsymbol{\beta})-\mathbf{L}_n(\sigma_0,\boldsymbol{\beta})\} \xrightarrow{p} 0$.

(a) Follows immediately from the Central Limit Theorem, since $r_i-\mathbf{u}_i^{\top}\boldsymbol{\beta}=\epsilon_i$. To derive (c), write $\psi_{1,s}(t) = s\,\psi_1(t/s)$. Then, the fact that $r_i-\mathbf{u}_i^{\top}\boldsymbol{\beta}=\epsilon_i$ allows to write

$$n^{1/2}\{\mathbf{L}_n(\widehat{\sigma},\boldsymbol{\beta})-\mathbf{L}_n(\sigma_0,\boldsymbol{\beta})\} = n^{-1/2}\sum_{i=1}^{n}\left\{\psi_{1,\widehat{\sigma}}(\epsilon_i)-\psi_{1,\sigma_0}(\epsilon_i)\right\} w_2(\|\mathbf{u}_i\|)\mathbf{u}_i.$$

Using the boundedness of ψ_2, the maximal inequality for covering numbers together with the bound

$$|\psi_{1,s_1}(r)-\psi_{1,s_2}(r)| \le (\|\psi_1\|_\infty + \|\varphi_1\|_\infty)\,|s_1-s_2|,$$

we easily obtain (c).

It remains to prove (b). Denote ξ_i intermediate points between $r_i-\mathbf{u}_i^{\top}\widetilde{\boldsymbol{\beta}}$ and $\widehat{r}_i-\widehat{\mathbf{u}}_i^{\top}\widetilde{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\eta}}(x) = (\widehat{\eta}_1(x),\ldots,\widehat{\eta}_p(x))^{\top}$ with $\widehat{\eta}_j(x) = \widehat{\phi}_j(x)-\phi_j(x)$ for $0\le j\le p$. Using a second order Taylor expansion, we have that $\widehat{\mathbf{L}}_n(\widehat{\sigma},\boldsymbol{\beta}) = \mathbf{L}_n(\widehat{\sigma},\boldsymbol{\beta}) + \widehat{\mathbf{L}}_{n,1} + \widehat{\mathbf{L}}_{n,2} + \widehat{\mathbf{L}}_{n,3} + \widehat{\mathbf{L}}_{n,4} + \widehat{\mathbf{L}}_{n,5}$, where

$$\begin{aligned}
\widehat{\mathbf{L}}_{n,1} &= \frac{1}{n}\sum_{i=1}^{n}\psi_1'\left(\frac{r_i-\mathbf{u}_i^{\top}\boldsymbol{\beta}}{\widehat{\sigma}}\right)\left\{\boldsymbol{\beta}^{\top}\widehat{\boldsymbol{\eta}}(X_i)-\widehat{\eta}_0(X_i)\right\} w_2(\|\mathbf{u}_i\|)\mathbf{u}_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\psi_1'\left(\frac{\epsilon_i}{\widehat{\sigma}}\right)\left\{\boldsymbol{\beta}^{\top}\widehat{\boldsymbol{\eta}}(X_i)-\widehat{\eta}_0(X_i)\right\} w_2(\|\mathbf{u}_i\|)\mathbf{u}_i, \\
\widehat{\mathbf{L}}_{n,2} &= \frac{\widehat{\sigma}}{n}\sum_{i=1}^{n}\psi_1\left(\frac{r_i-\mathbf{u}_i^{\top}\boldsymbol{\beta}}{\widehat{\sigma}}\right)\{w_2(\|\widehat{\mathbf{u}}_i\|)\widehat{\mathbf{u}}_i-w_2(\|\mathbf{u}_i\|)\mathbf{u}_i\} = \frac{\widehat{\sigma}}{n}\sum_{i=1}^{n}\psi_1\left(\frac{\epsilon_i}{\widehat{\sigma}}\right)\{w_2(\|\widehat{\mathbf{u}}_i\|)\widehat{\mathbf{u}}_i-w_2(\|\mathbf{u}_i\|)\mathbf{u}_i\}, \\
\widehat{\mathbf{L}}_{n,3} &= \frac{\widehat{\sigma}}{n}\sum_{i=1}^{n}\left\{\psi_1\left(\frac{\widehat{r}_i-\widehat{\mathbf{u}}_i^{\top}\boldsymbol{\beta}}{\widehat{\sigma}}\right)-\psi_1\left(\frac{r_i-\mathbf{u}_i^{\top}\boldsymbol{\beta}}{\widehat{\sigma}}\right)\right\} w_2(\|\widehat{\mathbf{u}}_i\|)(\widehat{\mathbf{u}}_i-\mathbf{u}_i), \\
\widehat{\mathbf{L}}_{n,4} &= \frac{1}{2\,\widehat{\sigma}\,n}\sum_{i=1}^{n}\psi_1''\left(\frac{\xi_i}{\widehat{\sigma}}\right)\left\{\boldsymbol{\beta}^{\top}\widehat{\boldsymbol{\eta}}(X_i)-\widehat{\eta}_0(X_i)\right\}^2 w_2(\|\widehat{\mathbf{u}}_i\|)\mathbf{u}_i, \\
\widehat{\mathbf{L}}_{n,5} &= \frac{1}{n}\sum_{i=1}^{n}\psi_1'\left(\frac{r_i-\mathbf{u}_i^{\top}\boldsymbol{\beta}}{\widehat{\sigma}}\right)\left\{\boldsymbol{\beta}^{\top}\widehat{\boldsymbol{\eta}}(X_i)-\widehat{\eta}_0(X_i)\right\}\{w_2(\|\widehat{\mathbf{u}}_i\|)-w_2(\|\mathbf{u}_i\|)\}\,\mathbf{u}_i.
\end{aligned}$$

Note that

$$w_2(\|\widehat{\mathbf{u}}_i\|)-w_2(\|\mathbf{u}_i\|) = \{\psi_2(\|\widehat{\mathbf{u}}_i\|)-\psi_2(\|\mathbf{u}_i\|)\}\frac{1}{\|\mathbf{u}_i\|} - \frac{w_2(\|\widehat{\mathbf{u}}_i\|)}{\|\mathbf{u}_i\|}\{\|\widehat{\mathbf{u}}_i\|-\|\mathbf{u}_i\|\},$$

so **N2** entails $|w_2(\|\widehat{\mathbf{u}}_i\|) - w_2(\|\mathbf{u}_i\|)| \le C\|\widehat{\boldsymbol{\eta}}(X_i)\|/\|\mathbf{u}_i\|$, where $C = \|w_2\|_\infty + \|\psi_2'\|_\infty$. Using this bound together with the fact that $r_i - \widehat{r}_i = \widehat{\eta}_0(X_i)$ and $u_{ij} - \widehat{u}_{ij} = \widehat{\eta}_j(X_i)$, we get, as in Bianco and Boente [9], that

$$n^{1/2}\|\widehat{\mathbf{L}}_{n,3}\| \le p\,\|w_2\|_\infty \|\psi_1'\|_\infty\, n^{1/2}\left\{\max_{0\le j\le p}\sup_{x\in\mathcal{S}_{\mathcal{H}}}|\widehat{\eta}_j(x)|\right\}^2 (1+p\,\|\boldsymbol{\beta}\|),$$

$$n^{1/2}\|\widehat{\mathbf{L}}_{n,4}\| \le \frac{1}{2\,\widehat{\sigma}}\,\|\psi_1''\|_\infty n^{1/2}\left\{\max_{0\le j\le p}\sup_{x\in\mathcal{S}_{\mathcal{H}}}|\widehat{\eta}_j(x)|\right\}^2 (1+p\,\|\boldsymbol{\beta}\|)^2\left\{\|\psi_2\|_\infty + p\,\|w_2\|_\infty\max_{0\le j\le p}\sup_{x\in\mathcal{S}_{\mathcal{H}}}|\widehat{\eta}_j(x)|\right\},$$

$$n^{1/2}\|\widehat{\mathbf{L}}_{n,5}\| \le p\,C\|\psi_1'\|_\infty(1+p\|\boldsymbol{\beta}\|)n^{1/2}\left\{\max_{0\le j\le p}\sup_{x\in\mathcal{S}_{\mathcal{H}}}|\widehat{\eta}_j(x)|\right\}^2.$$

Hence, (14) and the consistency of $\widehat{\sigma}$, entails that $n^{1/2}\|\widehat{\mathbf{L}}_{n,j}\| \xrightarrow{p} 0$, for $3\le j\le 5$.

It remains to show that $n^{1/2}\widehat{\mathbf{L}}_{n,j} \xrightarrow{p} 0$ for $j = 1,2$. First note that since $\widehat{\mathbf{u}}_i - \mathbf{u}_i = -\widehat{\boldsymbol{\eta}}(X_i)$, $w_2(t) = \varphi_2(t^2)$ and $\|\widehat{\mathbf{u}}_i\|^2 - \|\mathbf{u}_i\|^2 = \|\widehat{\boldsymbol{\eta}}(X_i)\|^2 + 2\,\mathbf{u}_i^{\top}\widehat{\boldsymbol{\eta}}(X_i)$, we have that

$$\begin{aligned}
w_2(\|\widehat{\mathbf{u}}_i\|)\widehat{\mathbf{u}}_i - w_2(\|\mathbf{u}_i\|)\mathbf{u}_i &= \{w_2(\|\widehat{\mathbf{u}}_i\|) - w_2(\|\mathbf{u}_i\|)\}\,\mathbf{u}_i + \{w_2(\|\widehat{\mathbf{u}}_i\|) - w_2(\|\mathbf{u}_i\|)\}\,(\widehat{\mathbf{u}}_i - \mathbf{u}_i) + w_2(\|\mathbf{u}_i\|)\,(\widehat{\mathbf{u}}_i - \mathbf{u}_i)\\
&= \left\{\varphi_2(\|\widehat{\mathbf{u}}_i\|^2) - \varphi_2(\|\mathbf{u}_i\|^2)\right\}\mathbf{u}_i - \widehat{\boldsymbol{\eta}}(X_i)\left\{\varphi_2(\|\widehat{\mathbf{u}}_i\|^2) - \varphi_2(\|\mathbf{u}_i\|^2)\right\} - \widehat{\boldsymbol{\eta}}(X_i)w_2(\|\mathbf{u}_i\|)\\
&= \varphi_2'(\|\mathbf{u}_i\|^2)\left(\|\widehat{\mathbf{u}}_i\|^2 - \|\mathbf{u}_i\|^2\right)\mathbf{u}_i + \varphi_2'(\xi_i)\left(\|\widehat{\mathbf{u}}_i\|^2 - \|\mathbf{u}_i\|^2\right)^2\mathbf{u}_i\\
&\quad - \widehat{\boldsymbol{\eta}}(X_i)\left\{\varphi_2(\|\widehat{\mathbf{u}}_i\|^2) - \varphi_2(\|\mathbf{u}_i\|^2)\right\} - \widehat{\boldsymbol{\eta}}(X_i)w_2(\|\mathbf{u}_i\|)\\
&= \varphi_2'(\|\mathbf{u}_i\|^2)\left\{\|\widehat{\boldsymbol{\eta}}(X_i)\|^2 + 2\,\mathbf{u}_i^{\top}\widehat{\boldsymbol{\eta}}(X_i)\right\}\mathbf{u}_i + \varphi_2'(\xi_i^2)\left\{\|\widehat{\boldsymbol{\eta}}(X_i)\|^2 + 2\,\mathbf{u}_i^{\top}\widehat{\boldsymbol{\eta}}(X_i)\right\}^2\mathbf{u}_i\\
&\quad - \widehat{\boldsymbol{\eta}}(X_i)\left\{\varphi_2(\|\widehat{\mathbf{u}}_i\|^2) - \varphi_2(\|\mathbf{u}_i\|^2)\right\} - \widehat{\boldsymbol{\eta}}(X_i)w_2(\|\mathbf{u}_i\|),
\end{aligned}$$

with ξ_i^2 an intermediate point between $\|\widehat{\mathbf{u}}_i\|^2$ and $\|\mathbf{u}_i\|^2$. Hence, $\widehat{\mathbf{L}}_{n,2} = 2\,\widehat{\mathbf{L}}_{n,2,1} - \widehat{\mathbf{L}}_{n,2,2} + \widehat{\mathbf{L}}_{n,3} - \widehat{\mathbf{L}}_{n,2,4}$ where

$$\widehat{\mathbf{L}}_{n,2,1} = \frac{\widehat{\sigma}}{n}\sum_{i=1}^{n}\psi_1\left(\frac{\epsilon_i}{\widehat{\sigma}}\right)\varphi_2'(\|\mathbf{u}_i\|^2)\mathbf{u}_i\,\mathbf{u}_i^{\top}\widehat{\boldsymbol{\eta}}(X_i),$$

$$\widehat{\mathbf{L}}_{n,2,2} = \frac{\widehat{\sigma}}{n}\sum_{i=1}^{n}\psi_1\left(\frac{\epsilon_i}{\widehat{\sigma}}\right)w_2(\|\mathbf{u}_i\|)\,\widehat{\boldsymbol{\eta}}(X_i),$$

$$\widehat{\mathbf{L}}_{n,2,3} = \frac{\widehat{\sigma}}{n}\sum_{i=1}^{n}\psi_1\left(\frac{\epsilon_i}{\widehat{\sigma}}\right)\left[\varphi_2'(\|\mathbf{u}_i\|^2)\|\widehat{\boldsymbol{\eta}}(X_i)\|^2 + \varphi_2'(\xi_i^2)\left\{\|\widehat{\boldsymbol{\eta}}(X_i)\|^2 + 2\,\mathbf{u}_i^{\top}\widehat{\boldsymbol{\eta}}(X_i)\right\}^2\right]\mathbf{u}_i,$$

$$\widehat{\mathbf{L}}_{n,2,4} = \frac{\widehat{\sigma}}{n}\sum_{i=1}^{n}\psi_1\left(\frac{\epsilon_i}{\widehat{\sigma}}\right)\widehat{\boldsymbol{\eta}}(X_i)\left\{\varphi_2(\|\widehat{\mathbf{u}}_i\|^2) - \varphi_2(\|\mathbf{u}_i\|^2)\right\}.$$

Using that φ_2' and $\zeta_2(t) = t\varphi_2'(t)$ are bounded, analogous arguments to those considered above, allow to show that $n^{1/2}\widehat{\mathbf{L}}_{n,2,j} \xrightarrow{p} 0$ for $j = 3,4$.

Therefore, we only have to show that $n^{1/2}\widehat{\mathbf{L}}_{n,1} \xrightarrow{p} 0$ and $n^{1/2}\widehat{\mathbf{L}}_{n,2,j} \xrightarrow{p} 0$ for $j = 1,2$. We will show that, for any $0\le\ell\le p$ and $1\le j\le p$,

$$\widehat{L}_{n,1,j,\ell} = n^{1/2}\frac{\widehat{\sigma}}{n}\sum_{i=1}^{n}\psi_1'\left(\frac{\epsilon_i}{\widehat{\sigma}}\right)w_2(\|\mathbf{u}_i\|)u_{ij}\,\widehat{\eta}_\ell(X_i) \xrightarrow{p} 0, \tag{A.12}$$

$$\widehat{L}_{n,2,1,j,\ell} = n^{1/2}\frac{\widehat{\sigma}}{n}\sum_{i=1}^{n}\psi_1\left(\frac{\epsilon_i}{\widehat{\sigma}}\right)\varphi_2'(\|\mathbf{u}_i\|^2)\,u_{ij}\,u_{i\ell}\,\widehat{\eta}_\ell(X_i) \xrightarrow{p} 0, \tag{A.13}$$

$$\widehat{L}_{n,2,2,j} = n^{1/2}\frac{\widehat{\sigma}}{n}\sum_{i=1}^{n}\psi_1\left(\frac{\epsilon_i}{\widehat{\sigma}}\right)w_2(\|\mathbf{u}_i\|)\,\widehat{\eta}_\ell(X_i) \xrightarrow{p} 0. \tag{A.14}$$

We will only prove (A.12), since the proofs of (A.13) and (A.14) can be obtained similarly using the fact that $\mathrm{E}\psi_1(\epsilon_i/\sigma) = 0$ for all $\sigma > 0$ and that $\varphi_2'(\|\mathbf{u}_i\|^2)\,u_{ij}\,u_{i\ell}$ is bounded.

Recall that $W_i(x) = K_i(x)/\{n\mathrm{E}K_1(x)\}$ with $K_i(x) = K(d(x,X_i)/h)$ and denote as

$$\widehat{\widehat{L}}_{n,1,j,\ell} = n^{1/2}\frac{\widehat{\sigma}}{n}\sum_{i=1}^{n}\psi_1'\left(\frac{\epsilon_i}{\widehat{\sigma}}\right)w_2(\|\mathbf{u}_i\|)u_{ij}\sum_{m=1}^{n}W_m(X_i)\,\psi\left(\frac{Z_{m\ell} - \phi_\ell(X_i)}{\widehat{s}_\ell}\right).$$

Proposition 3.2 entails that

$$\sup_{x\in\mathcal{S}_{\mathcal{H}}}\left|\widehat{\eta}_\ell(x)-\nu_\ell(\widehat{s}_\ell)^{-1}\widehat{s}_\ell\sum_{m=1}^n W_m(x)\,\psi\left(\frac{Z_{m\ell}-\phi_\ell(x)}{\widehat{s}_\ell}\right)\right| = O_{a.s.}(h^{2\eta}+\theta_n^2),$$

so, using that ψ_1' and ψ_2 are bounded we get that

$$\widehat{L}_{n,1,j,\ell}=\nu_\ell(\widehat{s}_\ell)^{-1}\widehat{s}_\ell\widehat{\widehat{L}}_{n,1,j,\ell}+\nu_\ell(\widehat{s}_\ell)^{-1}\widehat{s}_\ell\, n^{1/2}O_{a.s.}(h^{2\eta}+\theta_n^2)=\nu_\ell(\widehat{s}_\ell)^{-1}\widehat{s}_\ell\widehat{\widehat{L}}_{n,1}+o_{\mathbb{P}}(1),$$

since $nh^{4\eta}\to 0$ and $n^{1/2}\theta_n^2\to 0$. Note that since $\widehat{s}_j\xrightarrow{a.s.}s_j$, (A.12) follows if we show that $\widehat{\widehat{L}}_{n,1,j,\ell}\xrightarrow{p}0$. A Taylor's expansion of order two, lead to $\widehat{\widehat{L}}_{n,1,j,\ell}=\widehat{\sigma}\,n^{1/2}\left(\widehat{R}_{1n}+\widehat{s}_\ell^{-1}\widehat{R}_{2n}+\widehat{s}_\ell^{-2}\widehat{R}_{3n}\right)$, where

$$\widehat{R}_{1n}=\frac{1}{n}\sum_{i=1}^n\psi_1'\left(\frac{\epsilon_i}{\widehat{\sigma}}\right)w_2(\|\mathbf{u}_i\|)\,u_{ij}\sum_{m=1}^n W_m(X_i)\psi\left(\frac{u_{m\ell}}{\widehat{s}_\ell}\right),$$

$$\widehat{R}_{2n}=\frac{1}{n}\sum_{i=1}^n\psi_1'\left(\frac{\epsilon_i}{\widehat{\sigma}}\right)w_2(\|\mathbf{u}_i\|)\,u_{ij}\sum_{m=1}^n W_m(X_i)\psi'\left(\frac{u_{m\ell}}{\widehat{s}_\ell}\right)\{\phi_\ell(X_m)-\phi_\ell(X_i)\},$$

$$\widehat{R}_{3n}=\frac{1}{n}\sum_{i=1}^n\psi_1'\left(\frac{\epsilon_i}{\widehat{\sigma}}\right)w_2(\|\mathbf{u}_i\|)\,u_{ij}\sum_{m=1}^n W_m(X_i)\psi''\left(\frac{u_{m\ell}+\xi_{im}}{\widehat{s}_\ell}\right)\{\phi_\ell(X_m)-\phi_\ell(X_i)\}^2,$$

with ξ_{im} an intermediate point between 0 and $\phi_\ell(X_m)-\phi_\ell(X_i)$.

Using that $n\,h^{4\eta}\to 0$, ϕ_ℓ is Lipschitz of order η, $\widehat{s}_\ell\xrightarrow{p}s_\ell>0$ and $\sup_{x\in\mathcal{S}_{\mathcal{H}}}\left|\sum_{i=1}^n W_i(x)\right|\le 2$ almost surely, we obtain that

$$n^{1/2}|\widehat{R}_{3n}|\le\|\psi\prime\|_\infty\|\psi_2\|_\infty\|\psi''\|_\infty\sup_{x\in\mathcal{S}_{\mathcal{H}}}\left|\sum_{i=1}^n W_i(x)\right|\left(n\,h^{4\eta}\right)^{1/2}\xrightarrow{a.s.}0.$$

On the other hand, Lemma A.1 implies that $n^{1/2}\widehat{R}_{2n}\xrightarrow{p}0$, since $\widehat{\sigma}\to\sigma_0$. To show that $n^{1/2}\widehat{R}_{1n}\xrightarrow{p}0$, define

$$H_{1,im}(\sigma,s)=W_m(X_i)\psi_1'\left(\frac{\epsilon_i}{\sigma}\right)w_2(\|\mathbf{u}_i\|)\,u_{ij}\psi\left(\frac{u_{m\ell}}{s}\right)$$

and denote $\widehat{R}_{1n}(\sigma,s)=(1/n)\sum_{i=1}^n\sum_{m=1}^n H_{1,im}(\sigma,s)$. Then, $\widehat{R}_{1n}=\widehat{R}_{1n}(\widehat{\sigma},\widehat{s}_\ell)$.

Since $n^{1/4}(\widehat{\sigma}-\sigma)=O_{\mathbb{P}}(1)$, we have that $\widehat{R}_{1n}=\widehat{R}_{1n}(\sigma_0,\widehat{s}_\ell)-(\widehat{\sigma}-\sigma_0)\ \{\xi/(\widehat{\sigma}\,\sigma_0)\}\ \widehat{R}_{12n}(\xi,\widehat{s}_\ell)$, where ξ is an intermediate point between σ_0 and $\widehat{\sigma}$ and

$$\widehat{R}_{12n}(\sigma,s)=\frac{1}{n}\sum_{i=1}^n\sum_{m=1}^n W_m(X_i)\psi_1''\left(\frac{\epsilon_i}{\sigma}\right)\frac{\epsilon_i}{\sigma}w_2(\|\mathbf{u}_i\|)\,u_{ij}\psi\left(\frac{u_{m\ell}}{s}\right).$$

Taking $\upsilon(t)=\psi(t)$ and $\chi(t)=t\psi_1''(t)$ in Lemma A.2, from the consistency of $\widehat{\sigma}$ and $\widehat{s}_\ell$ we get that $n^{1/4}\widehat{R}_{12n}(\xi,\widehat{s}_\ell)\xrightarrow{p}0$ which entails that $n^{1/2}\left\{\widehat{R}_{1n}-\widehat{R}_{1n}(\sigma_0,\widehat{s}_\ell)\right\}=o_{\mathbb{P}}(1)$, since $n^{1/4}(\widehat{\sigma}-\sigma)=O_{\mathbb{P}}(1)$.

On the other hand, $\widehat{R}_{1n}(\sigma_0,\widehat{s}_\ell)=\widehat{R}_{1n}(\sigma_0,s_\ell)-(\widehat{s}_\ell-s_\ell)\ \{\xi_\ell/(\widehat{s}_\ell\,s_\ell)\}\ \widehat{R}_{13n}(\sigma_0,\xi_\ell)$, where ξ_ℓ is an intermediate point between s_ℓ and $\widehat{s}_\ell$ and

$$\widehat{R}_{13n}(\sigma,s)=\frac{1}{n}\sum_{i=1}^n\sum_{m=1}^n W_m(X_i)\psi_1'\left(\frac{\epsilon_i}{\sigma}\right)w_2(\|\mathbf{u}_i\|)\,u_{ij}\psi'\left(\frac{u_{m\ell}}{s}\right)\frac{u_{m\ell}}{s}.$$

Lemma A.2 with $\upsilon(t)=t\psi'(t)$ and $\chi(t)=\psi_1'(t)$ implies that $n^{1/4}\widehat{R}_{13n}(\sigma,\xi_\ell)\xrightarrow{p}0$ which, together with the fact that $n^{1/4}(\widehat{s}_\ell-s_\ell)=O_{\mathbb{P}}(1)$, entails that $n^{1/2}\left\{\widehat{R}_{1n}(\sigma_0,\widehat{s}_\ell)-\widehat{R}_{1n}(\sigma_0,s_\ell)\right\}\xrightarrow{p}0$. Using again Lemma A.2 with $\upsilon(t)=\psi'(t)$ and $\chi(t)=\psi_1(t)$, we get that $n^{1/2}\widehat{R}_{1n}(\sigma_0,s_\ell)=o_{\mathbb{P}}(1)$ leading to $n^{1/2}\widehat{R}_{1n}\xrightarrow{p}0$ and concluding the proof. □

References

[1] A. Ait-Saïdi, F. Ferraty, R. Kassa, P. Vieu, Cross-validated estimations in the single functional index model, Statistics 42 (2008) 475–494.
[2] G. Aneiros-Pérez, N. Ling, P. Vieu, Error variance estimation in semi-functional partially linear regression models, J. Nonparametr. Stat. 27 (2015) 316–330.
[3] G. Aneiros-Pérez, P. Vieu, Semi-functional partial linear model, Statist. Probab. Lett. 11 (2006) 1102–1110.
[4] G. Aneiros-Pérez, P. Vieu, Nonparametric time series prediction: a semi-functional partial linear modelling, J. Multivariate Anal. 99 (2008) 834–857.
[5] G. Aneiros-Pérez, P. Vieu, Automatic estimation procedure in partial linear model with functional data, Statist. Papers 52 (2011) 751–771.
[6] G. Aneiros-Pérez, P. Vieu, Testing linearity in semi-parametric functional data analysis, Comput. Statist. 28 (2013) 413–434.

[7] N. Azzedine, A. Laksaci, E. Ould-Said, On robust nonparametric regression estimation for a functional regressor, Statist. Probab. Lett. 78 (2008) 3216–3221.
[8] K. Benhenni, F. Ferraty, M. Rachdi, P. Vieu, Local smoothing regression with functional data, Comput. Statist. 22 (2007) 353–369.
[9] A. Bianco, G. Boente, Robust estimators in semiparametric partly linear regression models, J. Statist. Plann. Inference 122 (2004) 229–252.
[10] A. Bianco, G. Boente, Robust estimators under a semiparametric partly linear autoregression model: asymptotic behavior and bandwidth selection, J. Time Series Anal. 28 (2007) 274–306.
[11] A. Bianco, G. Boente, E. Martinez, Robust tests in semiparametric partly linear models, Scand. J. Statist. 33 (2006) 435–450.
[12] G. Boente, R. Fraiman, Robust nonparametric regression estimation for dependent observations, Ann. Statist. 17 (1989) 1242–1256.
[13] G. Boente, R. Fraiman, J. Meloche, Robust plug-in bandwidth estimators in nonparametric regression, J. Statist. Plann. Inference 57 (1997) 109–142.
[14] G. Boente, J.C. Pardo-Fernández, Robust testing for superiority between two regression curves, Comput. Statist. Data Anal. 97 (2016) 151–168.
[15] G. Boente, D. Rodriguez, Robust bandwidth selection in semiparametric partly linear regression models: Monte Carlo study and influential analysis, Comput. Statist. Data Anal. 52 (2008) 2808–2828.
[16] G. Boente, A. Vahnovan, Strong convergence of robust equivariant nonparametric functional regression estimators, Statist. Probab. Lett. 100 (2015) 1–11.
[17] E. Cantoni, E. Ronchetti, Resistant selection of the smoothing parameter for smoothing splines, Stat. Comput. 11 (2) (2001) 141–146.
[18] G. Chagny, A. Roche, Adaptive estimation in the functional nonparametric regression model, J. Multivariate Anal. 146 (2016) 105–118.
[19] D. Chen, P. Hall, H.G. Müller, Single and multiple index functional regression models with nonparametric link, Ann. Statist. 39 (2011) 1720–1747.
[20] A. Cuevas, A partial overview of the theory of statistics with functional data, J. Statist. Plann. Inference 147 (2014) 1–23.
[21] F. Ferraty, A. Goia, E. Salinelli, P. Vieu, Functional projection pursuit regression, TEST 22 (2013) 293–320.
[22] F. Ferraty, A. Laksaci, A. Tadj, P. View, Rate of uniform consistency for nonparametric estimates with functional variables, J. Statist. Plann. Inference 140 (2010) 335–352.
[23] F. Ferraty, A. Laksaci, P. Vieu, Estimating some characteristics of the conditional distribution in nonparametric functional models, Stat. Inference Stoch. Process. 9 (2006) 47–76.
[24] F. Ferraty, A. Peuch, P. Vieu, Modèle à indice fonctionnel simple, C. R. Acad. Sci. Paris 336 (2003) 1025–1028.
[25] F. Ferraty, Y. Romain, The Oxford Handbook of Functional Data Analysis, Oxford University Press, 2010.
[26] F. Ferraty, P. Vieu, The functional nonparametric model and application to spectrometric data, Comput. Statist. 17 (2002) 545–564.
[27] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer, New York, 2006.
[28] D. Gervini, Outlier detection and trimmed estimation for general functional data, Statist. Sinica 22 (2012) 1639–1660.
[29] I. Ghement, M. Ruiz, R. Zamar, Robust estimation of error scale in nonparametric regression models, J. Statist. Plann. Inference 138 (2008) 3200–3216.
[30] A. Goia, P. Vieu, A partitioned Single Functional Index Model, Comput. Statist. 30 (2015) 673–692.
[31] A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, J. Multivariate Anal. 146 (2016) 1–6.
[32] W. González-Manteiga, G. Aneiros-Pérez, Testing in partial linear regression models with dependent errors, J. Nonparametr. Stat. 15 (2003) 93–111.
[33] W. Härdle, H. Liang, J. Gao, Partially Linear Models, Springer, New York, 2000.
[34] X. He, Z. Zhu, W. Fung, Estimation in a semiparametric model for longitudinal data with unspecified dependence structure, Biometrika 89 (2002) 579–590.
[35] G. Henry, D. Rodriguez, Robust estimators in partly linear regression models on Riemannian manifolds, Comm. Statist. Theory Methods (2016) http://dx.doi.org/10.1080/03610926.2013.775302, http://www.tandfonline.com/doi/full/10.1080/03610926.2013.775302 (in press).
[36] L. Horváth, P. Kokoszka, Inference for Functional Data with Applications, Springer, New York, 2012.
[37] M. Hubert, E. Vandervieren, An adjusted boxplot for skewed distributions, Comput. Statist. Data Anal. 52 (2008) 5186–5201.
[38] D. Leung, Cross-validation in nonparametric regression with outliers, Ann. Statist. 33 (2005) 2291–2310.
[39] D. Leung, F. Marriott, E. Wu, Bandwidth selection in robust smoothing, J. Nonparametr. Stat. 4 (1993) 333–339.
[40] O. Linton, Second order approximations in the partially linear regression model, Econometrica 63 (1995) 1079–1112.
[41] R. Maronna, D. Martin, V.J. Yohai, Robust Statistics: Theory and Methods, John Wiley & Sons, New York, 2006.
[42] M. Rachdi, P. Vieu, Nonparametric regression for functional data: automatic smoothing parameter selection, J. Statist. Plann. Inference 137 (2007) 2784–2801.
[43] J. Ramsay, B. Silverman, Applied Functional Data Analysis: Methods and Case Studies, Springer, New York, 2002.
[44] J. Ramsay, B. Silverman, Functional Data Analysis, Springer, New York, 2005.
[45] P. Robinson, Root$-n-$consistent semiparametric regression, Econometrica 56 (1988) 931–954.
[46] P. Rousseeuw, Least median of squares regression, J. Amer. Statist. Assoc. 79 (1984) 871–880.
[47] H.L. Shang, Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density, Comput. Statist. Data Anal. 67 (2013) 185–198.
[48] H.L. Shang, Bayesian bandwidth estimation for a functional nonparametric regression model with mixed types of regressors and unknown error density, J. Nonparametr. Stat. 26 (2014) 599–615.
[49] H.L. Shang, Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density, Comput. Statist. 29 (2014) 829–848.
[50] H. Shin, Partial functional linear regression, J. Statist. Plann. Inference 139 (2009) 3405–3418.
[51] P. Speckman, Kernel smoothing in partial linear models, J. R. Stat. Soc. Ser. B Stat. Methodol. 50 (1988) 413–436.
[52] Y. Sun, M.G. Genton, Functional Boxplots, J. Comput. Graph. Statist. 20 (2011) 316–334.
[53] A. Vahnovan, Estimadores robustos en modelos de regresión noparamétricos funcionales y en modelos semi-funcionales parcialmente lineales (Ph.D. thesis), National University of La Plata, 2013, Available at http://sedici.unlp.edu.ar/handle/10915/44589.
[54] F. Wang, D. Scott, The L_1 method for robust nonparametric regression, J. Amer. Statist. Assoc. 89 (1994) 65–76.
[55] D. Zhang, X. Lin, M.F. Sowers, Assessing the effects of reproductive hormone profiles on bone mineral density using functional two-stage mixed models, Biometrics 63 (2007) 351–362.