Statistics and Probability Letters 100 (2015) 1–11

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Strong convergence of robust equivariant nonparametric functional regression estimators

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

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

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

ARTICLE INFO

Article history:
Received 17 July 2014
Received in revised form 23 January 2015
Accepted 24 January 2015
Available online 3 February 2015

MSC:
62G35
62F12

Keywords:
Functional data
Kernel weights
M-location functionals
Robust estimation

ABSTRACT

Robust nonparametric equivariant M-estimators for the regression function have been extensively studied when the covariates are in $\mathbb{R}^k$. In this paper, we derive strong uniform convergence rates for kernel-based robust equivariant M-regression estimator when the covariates are functional.

1. Introduction

A common problem in statistics is to study the relationship between a random variable Y and a set of covariates X. In many applications, the covariates can be seen as functions recorded over a period of time and regarded as realizations of a stochastic process, often assumed to be in the L^2 space on a real interval. These variables are usually called functional variables in the literature. In this general framework, statistical models adapted to infinite-dimensional data have been recently studied. We refer to Ramsay and Silverman (2002, 2005), Ferraty and Vieu (2006) and Ferraty and Romain (2011) for a description of different procedures for functional data. In particular, linear nonparametric regression estimators in the functional setting, that is, estimators based on a weighted average of the response variables, have been considered, among others, by Ferraty and Vieu (2004) and Ferraty et al. (2006) who also considered estimators of the conditional quantiles. Burba et al. (2009) studied k-nearest neighbor regression estimators while Ferraty et al. (2010) obtained almost complete uniform convergence results (with rates) for kernel-type estimators.

However, in the functional case the literature on robust proposals for nonparametric regression estimation is sparse. Cadre (2001) studied estimation procedures for the L^1 median estimators for a random variable on a Banach space while Azzedine et al. (2008) studied nonparametric robust estimation methods based on the M-estimators introduced by Huber (1964), when the scale is known.

* Correspondence to: Departamento de Matemáticas, FCEyN, UBA, Ciudad Universitaria, Pabellón 1, Buenos Aires, C1428EHA, Argentina. Fax: +54 11 45763335.

E-mail addresses: gboente@dm.uba.ar (G. Boente), avahnovan@gmail.com (A. Vahnovan).

http://dx.doi.org/10.1016/j.spl.2015.01.028
0167-7152/

In this paper, we consider the case in which the scale is unknown since in most practical situations, scale is unknown and needs to be estimated. As in location models, M-smoothers are shift equivariant. However, even if the mean and the median are scale equivariant, this property does not hold for M-location estimators unless a preliminary robust scale estimator is used to scale the residuals. The same holds for the robust nonparametric regression estimators considered in Azzedine et al. (2008). To ensure scale equivariance and robustness, a robust scale estimator needs to be used to decide which responses may be considered as atypical, so that their effect can be downweighted. In this sense, our contributions extend previous proposals in two directions. On one hand, we generalize the proposal given in the Euclidean case by Boente and Fraiman (1989) to provide robust equivariant estimators for the regression function in the functional case, that is, in the case where the covariates are in an infinite dimensional space. On the other hand, we extend the proposal given in Azzedine et al. (2008) to allow for an unknown scale and heteroscedastic models.

The paper is organized as follows. In Section 2, we state our notation, while in Section 3 we introduce the robust estimators to be considered. Section 4 contains the main results of this paper, that is, uniform convergence consistency and uniform convergence rates for the equivariant local M-estimators, over compact sets. These uniform convergence results are obtained either by giving conditions on the M-conditional location functional or by deriving similar results on the conditional empirical distribution function which extend those given in by Ferraty et al. (2010). Proofs are relegated to Appendix A and to the supplementary file available online (see Appendix B).

2. Basic definitions and notation

Throughout this paper, we consider independent and identically distributed observations (Y_i, X_i), $1 \leq i \leq n$ such that $Y_i \in \mathbb{R}$ and $X_i \in \mathcal{H}$ with the same distribution as (Y, X), where $(\mathcal{H}, d)$ is semi-metric functional space, that is, d satisfies the metric properties but $d(x, y) = 0$ does not imply $x = y$. We say that the observations satisfy a nonparametric functional regression model if (Y, X) is such that

$$Y = r(X) + U \tag{1}$$

where $r : \mathcal{H} \to \mathbb{R}$ is a smooth operator not necessarily linear. Throughout this paper, we will not require any moment conditions on the errors distribution. Usually in a robust framework, the error U is such that $U = \sigma(X)u$, where u is independent of X and with distribution F_0 symmetric around 0, that is, we assume that the errors u have scale equal to 1 to identify the function σ. When second moment exists, as it is the case of the classical approach, these conditions imply that $\mathbb{E}(U|X) = 0$ and $\mathrm{VAR}(U|X) = \sigma^2(X)$, which means that, in this situation, r and σ represent the conditional mean and standard deviation of the responses given the covariates, respectively. Hence, when $\mathbb{E}|Y| < \infty$, the regression function $r(X)$ in (1), which in this case equals $\mathbb{E}(Y|X)$, can be estimated using the extension to the functional setting of the Nadaraya–Watson estimator (see, for example, Härdle, 1990). To be more precise, let K be a kernel function and $h = h_n$ a sequence of strictly positive real numbers. Denote as

$$w_i(x) = K_i(x) \left(\sum_{i=1}^{n} K_i(x) \right)^{-1}, \tag{2}$$

where $K_i(x) = K\left(d(x, X_i)/h\right)$. Then, the classical regression estimator is defined as

$$\widehat{r}(x) = \sum_{i=1}^{n} w_i(x) Y_i. \tag{3}$$

Under regularity conditions, Ferraty and Vieu (2006) obtained convergence rates for the estimator $\widehat{r}(x)$, while Ferraty et al. (2010) derived uniform consistency results with rates for the estimator of the so-called generalized regression function $r_\varphi(x) = \mathbb{E}(\varphi(Y)|X = x)$ where φ is a known real Borel measurable function. As mentioned therein, this convergence is related to the Kolmogorov's ϵ-entropy of $S_{\mathcal{H}}$ and the function ϕ that controls the small ball probability of the functional variable X.

The conditional cumulative distribution function of Y given $X = x$ is defined, for each $x \in \mathcal{H}$, as $F(y|X = x) = \mathbb{P}(Y \leq y|X = x)$, for any $y \in \mathbb{R}$. As in Ferraty et al. (2010), we will assume that there is a regular version of the conditional distribution. An estimator of the conditional distribution function, can be obtained noting that $F(y|X = x) = \mathbb{E}(\mathbb{I}_{(-\infty,y]}(Y)|X = x)$, that is, taking $\varphi(Y) = \mathbb{I}_{(-\infty,y]}(Y)$ in the generalized regression function $r_\varphi(x)$ and using (3). Hence, the kernel estimator $\widehat{F}(y|X = x)$ of $F(y|X = x)$ equals

$$\widehat{F}(y|X = x) = \sum_{i=1}^{n} w_i(x) \mathbb{I}_{(-\infty,y]}(Y_i), \tag{4}$$

where $w_i(x)$ are defined in (2). Among other results, in Section 4, we obtain uniform strong convergence rates for $\widehat{F}(y|X = x)$ over $\mathbb{R} \times S_{\mathcal{H}}$ with $S_{\mathcal{H}} \subset \mathcal{H}$ a compact set, generalizing the results in Ferraty et al. (2010).

From now on, $m(x)$ stands for the median of the conditional distribution function, that is, $m(x) = \inf\left\{y \in \mathbb{R} : F(y|X = x) \geq \frac{1}{2}\right\}$. If $F(\cdot|X = x)$ is a strictly increasing distribution function, then the conditional median exists and is unique. Moreover, it can be defined as $m(x) = (F_Y^x)^{-1}(1/2)$, where $F_Y^x(y) = F(y|X = x)$. An estimator $\widehat{m}(x)$ of the conditional median is easily obtained as the median of $\widehat{F}(y|X = x)$.

3. The robust equivariant estimators and its related functional

The estimators defined in (3) are sensitive to atypical observations, since they are based on averaging the responses. The effect of an outlier Y_i will be larger as the distance of the related covariate X_i to the point x is smaller. In this sense, atypical data in the responses in nonparametric regression may lead to a complete distorted estimation. As mentioned by Härdle (1990) '*From a data-analytic viewpoint, a nonrobust behavior of the smoother is sometimes undesirable.... Any erratic behavior of the nonparametric pilot estimate will cause biased parametric formulations*'. This effect observed for covariates in $\mathbb{R}^k$ still appears for functional covariates. In this sense, robust estimates in a functional non-parametric setting can thus be defined as insensitive to a single wild spike outlier. To provide resistant estimators to large residuals when dealing with functional covariates, Azzedine et al. (2008) extended the robust nonparametric kernel regression estimator defined by Collomb and Härdle (1986) when $X \in \mathbb{R}^k$ to the infinite-dimensional setting. Some asymptotic results can be found in Crambes et al. (2008) and Attouch et al. (2009). The proposal given in Azzedine et al. (2008) assumes that scale is known and thus, has two main drawbacks. As discussed in the Introduction, it cannot be directly applied in practice where scale is usually unknown. Besides, if scale estimation is avoided, the M-local estimators do not provide equivariant estimators.

Let (Y, X) be a random element in $\mathbb{R} \times \mathcal{H}$ and define

$$\lambda(x, a, \sigma) = \mathbb{E}\left(\psi\left(\frac{Y-a}{\sigma}\right)\Big|X = x\right), \tag{5}$$

where $\psi : \mathbb{R} \to \mathbb{R}$ is an odd, bounded and continuous function. We denote by $g(x)$ the solution of $\lambda(x, a, s(x)) = 0$ where $s(x)$ is a robust measure of the conditional scale. The conditional scale measure can be taken as the normalized conditional median of the absolute deviation from the conditional median, that is,

$$s(x) = \frac{1}{c_G}\text{MED}(|Y - m(x)|\ |X = x) = \text{MAD}_c(F_Y^x(\cdot)) \tag{6}$$

where $m(x)$ is the median of the conditional distribution as defined in Section 2 and c_G is a constant ensuring Fisher-consistency at a given distribution G. Note that $s(x)$ which corresponds to a robust measure of the conditional scale, usually equals $\sigma(x)$ up to a multiplicative constant, when $U = \sigma(X)u$ with u independent of X. For instance, the median of the absolute deviation is usually calibrated so that $\text{MAD}(\Phi) = 1$, taking $c_\Phi = \Phi^{-1}(3/4)$, where Φ states for the distribution function of a standard normal random variable. In this case, when the errors u have a Gaussian distribution, we have that $s(x) = \text{MAD}_c(F_Y^x(\cdot)) = \sigma(x)$.

To obtain estimators of $g(x)$ we plug-into (5) an estimator of $F_Y^x(y)$, which will be taken as $\widehat{F}(y|X = x)$. Denote by $\widehat{s}(x)$ a robust estimator of the conditional scale, for instance, $\widehat{s}(x) = \text{MAD}_c(\widehat{F}(\cdot|X = x))$, the scale measure defined in (6) evaluated in $\widehat{F}(y|X = x)$. With this notation, the robust nonparametric estimator of $g(x)$ is given by the solution $\widehat{g}(x)$ of $\widehat{\lambda}(x, a, \widehat{s}(x)) = 0$, where

$$\widehat{\lambda}(x, a, \sigma) = \int \psi\left(\frac{y-a}{\sigma}\right) d\widehat{F}(y|X = x) = \sum_{i=1}^{n} w_i(x)\psi\left(\frac{Y_i - a}{\sigma}\right). \tag{7}$$

Remark 3.1. (a) In the classical setting, the target is to estimate the conditional mean $\mathbb{E}(Y|X = x)$ and this quantity is obtained choosing $\psi(t) = t$ in (5). When considering $\psi(t) = \text{sgn}(t)$ the target is now the conditional median. For general score functions ψ, the target is the solution $g(x)$ of $\lambda(x, a, s(x)) = 0$. When ψ is a strictly increasing function, $g(x)$ is the so-called robust conditional location functional as introduced in Boente and Fraiman (1989) who noted that this functional provides a natural extension of the conditional expectation. As mentioned, for instance, in Collomb and Härdle (1986), the choice of a bounded function ψ in (7) suggests more stable prediction properties, in particular, when a small amount of outliers are present in the responses Y_i.

As in the finite-dimensional setting, under model (1), the researcher is seeking for consistent estimators of the regression function r without requiring moment conditions on the errors U_i in which case bounded score functions are helpful. In this case, when ψ is odd, the assumption that the conditional distribution of the error U given X is symmetric around 0 is needed if we want to guarantee that all robust location conditional estimators are estimating the same quantity. This property is usually known as Fisher-consistency and means that the target functional is the quantity of interest, in our case, the regression function r in model (1). Hence, the robust conditional location functional g is Fisher-consistent for errors with symmetric conditional distribution and for that reason, we will call $g(x)$ the regression function. In particular, if $U = \sigma(X)u$, with u independent of X and with symmetric distribution, g is Fisher-consistent. This result may be extended straightforward if the oddness of the score function and the symmetry assumption on the errors distribution are replaced by $\mathbb{E}_{F_0}(\psi(u/\sigma)) = 0$, for any $\sigma > 0$.

(b) From Theorem 2.1 of Boente and Fraiman (1989), if the score function ψ is a strictly increasing bounded continuous score function such that $\lim_{t\to-\infty}\psi(t) < 0 < \lim_{t\to+\infty}\psi(t)$, the robust location conditional functional $g(x)$ exists, is unique and measurable. Furthermore, its weak continuity was obtained in Theorem 2.2 therein. Therefore, by applying this functional to weak consistent estimators of the conditional distribution, we obtain consistent and asymptotically strongly robust estimators of the robust location conditional functional $g(x)$. These results can be applied in our functional framework too, since they only require the existence of a regular version of the conditional distribution. It is also clear that when ψ is an odd function as required in A4 below, the condition $\lim_{t\to-\infty}\psi(t) < 0 < \lim_{t\to+\infty}\psi(t)$ is fulfilled. Besides, in this case, the continuity of ψ entails that for each fixed x, the estimating equation $\widehat{\lambda}(x, a, \widehat{s}(x)) = 0$ also has a solution, which is unique if ψ is strictly increasing, since this property is inherited by $\widehat{\lambda}(x, a, \widehat{s}(x))$. As mentioned for the location model in Maronna et al. (2006), uniqueness may hold without requiring the strict monotonicity of ψ as is the situation when considering $\psi(t) = \mathrm{sgn}(t)$ or as ψ the Huber score function.

4. Main results

Uniform convergence results and uniform convergence rates for the local M-estimators are derived in Sections 4.1 and 4.2 under some general assumptions that are described below. From now on, $\xrightarrow{a.co.}$ and $a.co.$ stand for almost complete convergence while $\xrightarrow{a.s.}$ stands for almost sure convergence.

As mentioned in Section 2, 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 developed by Kolmogorov (1956) to classify compact sets according to their massivity and which is an important tool to obtain uniform convergence results. Given a subset $S_{\mathcal{H}} \subset \mathcal{H}$ and $\epsilon > 0$, denote $N_\epsilon(S_{\mathcal{H}})$ the minimal number of open balls of radius ϵ needed to cover $S_{\mathcal{H}}$, that is, $N_\epsilon(S_{\mathcal{H}})$ is the smallest integer k such that there exists $x_1, \dots, x_k$ such that $S_{\mathcal{H}} \subset \cup_{j=1}^{k} \mathcal{V}(x_k, \epsilon)$. If no such n exists, then $N_\epsilon(S_{\mathcal{H}}) = \infty$. The quantity $\psi_{S_{\mathcal{H}}}(\epsilon) = \log(N_\epsilon(S_{\mathcal{H}}))$ is called Kolmogorov's ϵ-entropy of the set $S_{\mathcal{H}}$ and it provides a measure of the complexity of the set $S_{\mathcal{H}}$, in the sense that a high entropy means that much more information is needed to describe the set with an accuracy ϵ. As is well known, the set $S_{\mathcal{H}}$ is totally bounded if it has a finite entropy, so that compact sets are complete and with finite entropy sets. Moreover, under the assumption that the space is complete, the set $S_{\mathcal{H}}$ has finite entropy if and only if its closure is compact. As mentioned in Ferraty et al. (2010), the choice of the topological structure through the choice of the semi-metric plays a crucial role to obtain uniform asymptotic results over $S_{\mathcal{H}}$. Several examples of sets with finite Kolmogorov's ϵ-entropy are discussed, among others, in Section 2.2 of Ferraty et al. (2010).

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

As mentioned in Ferraty and Vieu (2006), 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 A1. As in Ferraty et al. (2010), we have to take into account the uniformity aspect to obtain uniform results.

We will consider the following set of assumptions:

A1. For all $x \in S_{\mathcal{H}}$ there exists $\phi : \mathbb{R} \to \mathbb{R}_{>0}$ such that $\phi(h) \to 0$ when $h \to 0$ and $0 < C\phi(h) \le \mathbb{P}(X \in B(x, h)) \le C'\phi(h)$.

A2. 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 $-\infty < \inf_{u\in[0,1]} K'(u) \le \sup_{u\in[0,1]} K'(u) = \|K'\|_\infty < 0$.
  (b) If $K(1) > 0$, there exist $C, C' > 0$ such that $C\,\mathbb{I}_{[0,1]}(u) < K(u) < C'\mathbb{I}_{[0,1]}(u)$.

A3. The functions ϕ and $\psi_{S_{\mathcal{H}}}$ are such that:
  (a) $\phi : \mathbb{R} \to \mathbb{R}_{>0}$ is differentiable with first derivative ϕ'. Moreover, there exists $C_\phi > 0$ and $\eta_0 > 0$, such that for all $\eta < \eta_0$, $\phi'(\eta) < C_\phi$. If $K(1) = 0$, the function ϕ satisfies also the following additional condition: there exist $C > 0$ and $\eta_0 > 0$ such that for any $0 < \eta < \eta_0$ we have that $\int_0^\eta \phi(u)du > C\eta\phi(\eta)$.
  (b) for n large enough,
$$\frac{(\log n)^2}{n\phi(h)} < \psi_{S_{\mathcal{H}}}\left(\frac{\log n}{n}\right) < \frac{n\phi(h)}{\log n}.$$

A4. $\psi : \mathbb{R} \to \mathbb{R}$ is an odd, strictly increasing, bounded and continuous differentiable function, with bounded derivative ψ' such that $\zeta(u) = u\psi'(u)$ is bounded.

A5. The sequence $h = h_n$ is such that $h_n \to 0$, $n\phi(h_n) \to \infty$ and $n\phi(h_n)/\log n \to \infty$ as $n \to \infty$.

The requirement $\phi(h) > 0$ in A1 is a generalization to the functional setting of the assumption that the random vector $X \in \mathbb{R}^k$ has a positive density over compact sets. Note that we further assume that $\phi(h) \to 0$ when $h \to 0$, which entails $\mathbb{P}(X = x) = 0$ for all $x \in \mathcal{H}$. Moreover, assumption A5 also generalizes the bandwidth requirements from the finite-dimensional setting to the infinite-dimensional one to obtain uniform rates of convergence. Detailed comments on assumptions A1, A3(b) and, in particular, on the requirements on the bandwidth parameter are relegated to Section 4.3. Assumptions A2 and A3(a) are standard conditions on the kernel weight and the function ϕ to deal with functional covariates, while A4 is a usual requirement when considering M-location functionals.

4.1. Uniform strong convergence results

In this section, we will obtain uniform convergence results of the regression estimator $\widehat{g}$ defined as a solution of $\widehat{\lambda}(x, a, \widehat{s}(x)) = 0$, over sets compact sets. Theorems 4.1 and 4.3 provide uniform convergence for the robust regression estimators $\widehat{g}$. The main difference between Theorems 4.3 and 4.1 is that the latter bypass the estimators of the conditional distribution function requiring only smoothness to the function $\lambda(\cdot, a, \sigma)$, for each fixed a and σ.

Theorem 4.1. *Let $S_{\mathcal{H}} \subset \mathcal{H}$ be a compact set such that $\lambda(\cdot, a, \sigma) : \mathcal{H} \to \mathbb{R}$ is a continuous function on $S_{\mathcal{H}}$. Assume that* A1 *to* A5 *hold and that, for all $\sigma > 0$, $g(x)$ is the unique solution of $\lambda(x, a, \sigma) = 0$. Let $\widehat{s}(x)$ be a robust scale estimator such that with probability* 1*, there exist real constants $0 < A < B$ and $A < \widehat{s}(x) < B$ for all $x \in S_{\mathcal{H}}$ and $n \geq n_0$. Then, $\sup_{x\in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| \xrightarrow{a.s.} 0$ as $n \to \infty$.*

Theorem 4.2 generalizes the result obtained in Lemma 6.5 of Ferraty and Vieu (2006) and is the functional counterpart of Theorem 3.1 in Boente and Fraiman (1991). The following additional conditions are needed.

A6. $F(y|X = x)$ is symmetric around $g(x)$.
A7. Let $S_{\mathcal{H}}$ be a compact subset of $\mathcal{H}$ such that
  (a) for any fixed $y \in \mathbb{R}$, the function $F(y|X = x)$ is continuous on $S_{\mathcal{H}}$.
  (b) the following equicontinuity condition holds:
$$\forall \epsilon > 0 \, \exists \, \delta > 0 : \; |u - v| < \delta \;\Rightarrow\; \sup_{x\in S_{\mathcal{H}}} |F(u|X = x) - F(v|X = x)| < \epsilon$$
.
A8. $F(y|X = x)$ has a unique median $m(x)$.

Theorem 4.2. *Let $S_{\mathcal{H}} \subset \mathcal{H}$ be a compact set. Under* A1 *to* A3, A5 *and* A7*, we have that $\sup_{x\in S_{\mathcal{H}}} \sup_{y\in\mathbb{R}} |\widehat{F}(y|X = x) - F(y|X = x)| \xrightarrow{a.s.} 0$.*

Theorem 4.3. *Assume that* A1 *to* A3 *and* A5 *to* A7 *hold for a given compact set $S_{\mathcal{H}} \subset \mathcal{H}$. Moreover, let $\widehat{s}(x)$ be a robust scale estimator such that with probability* 1*, there exist real constants $0 < A < B$ and $A < \widehat{s}(x) < B$ for all $x \in S_{\mathcal{H}}$ and $n \geq n_0$. Then, we have that*

(a) $\sup_{x\in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| \xrightarrow{a.s.} 0$ *as $n \to \infty$, if* A4 *also holds.*
(b) $\sup_{x\in S_{\mathcal{H}}} |\widehat{m}(x) - m(x)| \xrightarrow{a.s.} 0$ *when $n \to \infty$, if in addition* A8 *holds.*

Remark 4.1. (i) Lemma A.4 in Appendix A together with Theorem 4.2 implies that if we choose $\widehat{s}(x) = \text{MAD}_c(\widehat{F}(\cdot|X = x))$, the condition required to the scale estimator in Theorems 4.1 and 4.3 is fulfilled.
(ii) When estimating the scale function, assumption A6 ensures that $\lambda(x, g(x), \sigma) = 0$ for all $\sigma > 0$, as required in Theorem 4.1. Besides, this condition is also needed in Theorem 4.3 to derive consistency of the robust equivariant proposal (see also, Boente and Fraiman, 1989). Note that to ensure uniform consistency of the local median, we assume that $F(y|X = x)$ has a unique median $m(x)$ instead of A4.
(iii) It is worth noting that A7(a) and the compactness of $S_{\mathcal{H}}$ entail that for any $\epsilon > 0$, there exists $\delta > 0$ such that $\sup_{x\in S_{\mathcal{H}}} \sup_{d(x,u)<\delta} |F(y|X = x) - F(y|X = u)| < \epsilon$, which is used in the proof of Lemma A.3(a). Furthermore, A7(a) also implies that there exist real numbers a, b such that, for all $x \in S_{\mathcal{H}}$, $F(b|X = x) > 1 - \epsilon$ and $F(a|X = x) < \epsilon$ which allows to show that $s(x) = \text{MAD}_c(F_Y^x(\cdot))$ is bounded and bounded away from 0 for any $x \in S_{\mathcal{H}}$.
(iv) Assume that (Y, X) fulfills the model (1) with $U = \sigma(X)u$, where u is independent of X and has a symmetric distribution. Then, $F(y|X = x) = F_0((y - r(x))/\sigma(x))$, so A7(a) holds if F_0 is a continuous function and $r : \mathcal{H} \to \mathbb{R}$ and $\sigma : \mathcal{H} \to \mathbb{R}$ are continuous functions on $S_{\mathcal{H}}$. Moreover, if $\inf_{x\in S_{\mathcal{H}}} \sigma(x) > 0$ and F_0 is Lipschitz then A7(b) holds. It is worth noticing that if the distribution F_0 of u has unique median at 0 then A8 holds.

4.2. Uniform strong convergence rates

The uniform convergence rates of this section will require the following additional assumptions

A9. The Kolmogorov ϵ-entropy of $S_{\mathcal{H}}$ satisfies one of the following
  (a) $\sum_{n=1}^{\infty} n \exp\left\{(1 - \beta)\psi_{S_{\mathcal{H}}}(\log(n)/n)\right\} < \infty$ for some $\beta > 1$.
  (b) $\sum_{n=1}^{\infty} n^{1/2} \exp\left\{(1 - \beta)\psi_{S_{\mathcal{H}}}(\log(n)/n)\right\} < \infty$ for some $\beta > 1$.
A10. The function $F(y|X = x)$ is uniformly Lipschitz in a neighborhood $S_{\mathcal{H}}^{\epsilon}$ of $S_{\mathcal{H}}$, that is, there exist a constant $D > 0$ and $\eta_1 > 0$ such that for $x_1, x_2 \in S_{\mathcal{H}}^{\epsilon}$, we have $\sup_{y\in\mathbb{R}} |F(y|X = x_1) - F(y|X = x_2)| \leq D\, d^{\eta_1}(x_1, x_2)$.

Theorems 4.5 and 4.6 give almost complete convergence rates for the estimators of the empirical conditional distribution and for the local M-estimators of the regression function. Besides, Theorem 4.4 also provides uniform convergence rates for the robust regression estimators $\widehat{g}$. The main difference between Theorems 4.6 and 4.4 is that the latter requires a Lipschitz condition on the regression function instead of requiring a uniform Lipschitz condition on the conditional distribution function, which is more natural in the present setting. However, a stronger condition on the entropy of the set $S_{\mathcal{H}}$, assumption A9(a), is required. On the other hand, to obtain rates of convergence for the conditional distribution function A9(b) and A10 will be used. The proofs of the results in this Section are given in the supplementary material available online (see Appendix B).

Theorem 4.4. *Let $S_{\mathcal{H}} \subset \mathcal{H}$ be a compact set. Assume that* A1 *to* A5 *and* A9(a) *hold. Moreover, assume that σ and g are Lipschitz functions of order η_σ and η_g, respectively. If $\eta_{\min} = \min(\eta_\sigma, \eta_g)$ is such that $\eta_{\min} < 1$, assume furthermore that conditions (8) or (9) below hold*

$$h \left(\frac{n}{\log n}\right)^{1-\eta_{\min}} \leq C_{\eta_{\min}} \quad \textit{for all } n \geq 1 \tag{8}$$

$$\phi(h) \left(\frac{n}{\log n}\right)^{1-\eta_{\min}} \leq C_{\eta_{\min}} \quad \textit{for all } n \geq 1. \tag{9}$$

Let $g(x)$ be the unique solution of $\lambda(x, a, s(x)) = 0$ and let $\widehat{s}(x)$ be a robust scale estimator satisfying that with probability 1, *there exist real constants $0 < A < B$ such that $A < \widehat{s}(x) < B$ for all $x \in S_{\mathcal{H}}$ and $n \geq n_0$. Then, if $\widehat{g}(x)$ is a solution of (7) such that $\sup_{x\in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| \xrightarrow{a.s.} 0$, we have that $\sup_{x\in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| = O_{a.co.}\ (h^{\eta_{\min}} + \theta_n)$, where $\theta_n^2 = \psi_{S_{\mathcal{H}}}(\log(n)/n)/(n\phi(h))$.*

Theorem 4.5. *Let $S_{\mathcal{H}} \subset \mathcal{H}$ be a compact set and denote as $\theta_n^2 = \psi_{S_{\mathcal{H}}}(\log(n)/n)/(n\phi(h))$. Assume that* A1 *to* A3, A5, A7, A9(b) *and* A10 *hold. If η_1 in* A10 *is such that $\eta_1 < 1/2$, assume in addition that there exists a positive constant $C^\star$ such that*

$$\left(\frac{n}{\log n}\right)^{1-\eta_1} \phi(h) \leq C^\star. \tag{10}$$

Then, we have that $\sup_{y\in\mathbb{R}} \sup_{x\in S_{\mathcal{H}}} |\widehat{F}(y|X = x) - F(y|X = x)| = O_{a.co.}\ (h^{\eta_1} + \theta_n)$.

Remark 4.2. Assume that (Y, X) satisfy model (1) with $U = \sigma(X)u$ where $\sigma(x) = \sigma > 0$ for all x, i.e., that the nonparametric regression model is homoscedastic. Then, if F_0 is Lipschitz and r is Lipschitz of order η_1, A10 holds. In the heteroscedastic situation, to ensure that A10 holds we need to require that $\inf_{x\in S_{\mathcal{H}}} \sigma(x) > 0$, σ is Lipschitz of order η_1 and F_0 has a density f_0 such that $y f_0(y)$ is bounded.

It is worth mentioning that if instead of being concerned with uniform convergence rates for $x \in S_{\mathcal{H}}$ and $y \in \mathbb{R}$, we are only interested in obtaining uniform rates for $x \in S_{\mathcal{H}}$ as in Corollary 3 of Ferraty et al. (2010), it suffices that the Kolmogorov's ϵ-entropy of $S_{\mathcal{H}}$ satisfies $\sum_{n=1}^{\infty} \exp\left\{(1-\beta)\psi_{S_{\mathcal{H}}}(\log(n)/n)\right\} < \infty$, for some $\beta > 1$, instead of A9. The uniformity for $y \in \mathbb{R}$ needed to obtain uniform convergence rates for $\widehat{g}$ require stronger conditions on the entropy of the set $S_{\mathcal{H}}$.

Theorem 4.6. *Let $S_{\mathcal{H}} \subset \mathcal{H}$ be a compact set. Assume that* A1 *to* A7 , A9(b) *and* A10 *hold. Moreover, if η_1 in A10 is such that $\eta_1 < 1/2$, assume that (10) holds. Let $\widehat{s}(x)$ be a robust scale estimator such that with probability* 1, *there exist real constants $0 < A < B$ verifying $A < \widehat{s}(x) < B$ for all $x \in S_{\mathcal{H}}$ and $n \geq n_0$. Then, if $\theta_n^2 = \psi_{S_{\mathcal{H}}}(\log(n)/n)/(n\phi(h))$, we have that $\sup_{x\in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| = O_{a.co.}\ (h^{\eta_1} + \theta_n)$.*

4.3. General comments on the assumptions

In this section, we will briefly discuss the assumptions considered and in particular, how they influence the obtained convergence rates.

As mentioned for instance in Ferraty et al. (2010), assumptions A1, A3 and A5 are standard conditions to derive uniform consistency of the classical Nadaraya–Watson estimator in a functional setting, while some additional conditions on the entropy of the set $S_{\mathcal{H}}$ are needed to obtain uniform convergence rates.

The function ϕ defined in A1 controls uniformly the concentration of the probability measure of the functional variable on a small ball. It is a decreasing function of h and $\lim_{h\to 0} \phi(h) = 0$. Hence, the condition of having a bounded derivative given in A3(a), allows us to consider ϕ as a Lipschitz function around zero, that is, $\phi(h) \leq Ch$ which together with A5 implies that $nh/\log n \to \infty$. Sufficient conditions ensuring that A1 holds, i.e., that a concentration property for the probability measure holds uniformly in $S_{\mathcal{H}}$, are given in Section 7.2 of Ferraty et al. (2010).

As discussed in Ferraty et al. (2006, 2010), besides to the rates h^{η_1} and $h^{\eta_{\min}}$ related to the smoothness of the target function appearing in Theorems 4.4 and 4.5, respectively, the rate $\theta_n = \sqrt{\psi_{S_{\mathcal{H}}}(\log(n)/n)/(n\phi(h))}$ is linked to topological properties, taking into account both the concentration of the probability measure of the random element X and the ϵ-entropy of the set $S_{\mathcal{H}}$. As mentioned by these authors, the notion of small ball is strongly related with the semi-metric d whose choice can increase the concentration of the probability measure to deal with the curse of the infinite-dimension simultaneously increasing the entropy of the set $S_{\mathcal{H}}$.

We will give some examples in which the assumptions on the entropy and the measure concentration are fulfilled. For further discussions we refer to Ferraty and Vieu (2006) and Ferraty et al. (2006, 2010).

(i) The first example corresponds to multivariate nonparametric regression, i.e., $X \in \mathbb{R}^k$. In this setting, one usually assumes that the covariate X has bounded density f_X that is strictly positive over compact sets, so that $\phi(h) = h^k$ and A5 corresponds to the standard requirements on the bandwidth. On the other hand, the ball of radius ρ in $\mathbb{R}^k$ can be covered by $(4\rho + \epsilon)^k/\epsilon^k$ balls of radius ϵ (see van de Geer, 2000). Hence, the ϵ-entropy of any compact set, $\psi_{S_{\mathcal{H}}}(\epsilon)$ is of order $\log(1/\epsilon)$ so that A9 is fulfilled for $\beta > 2$. Finally, A3 is satisfied in the Euclidean situation if $(\log n)^2 = O(nh^k)$. In this case, the term $\theta_n = \sqrt{\log n/(nh^k)}$ is the standard pointwise rate of convergence in the Euclidean setting. In particular, if $h = n^{-\tau}$, then, A5 and A3(b) hold if $0 < \tau < 1/k$.

(ii) Our second example is the Cameron Martin space and we refer to Li and Shao (2001) for the results to be mentioned below. From now on, $f_1(\epsilon) \approx f_2(\epsilon)$ means that f_1 and f_2 have the same order, i.e., for some constants $C_1, C_2 > 0$ and for ϵ small enough $C_1 f_1(\epsilon) \le f_2(\epsilon) \le C_2 f_1(\epsilon)$.

Let us consider the situation where $\mathcal{H}$ is a real separable Banach space with norm $\|\cdot\|$ and X is a centered Gaussian random element with law μ. Let H_μ be the reproducing Hilbert space generated by μ and denote as $\|\cdot\|_\mu$ the inner product norm induced on H_μ. The unit ball $\mathcal{B}_{H_\mu}(0,1) = \{x \in H_\mu : \|x\|_\mu \le 1\}$ is a compact set of $\mathcal{H}$ with the topology induced by $\|\cdot\|$. Moreover, its ϵ-entropy, $\psi_{\mathcal{B}_{H_\mu}(0,1)}(\epsilon)$, with respect to $\|\cdot\|$, is related to the centered small ball probability $\varphi_0(\epsilon) = \mu(\|x\| \le \epsilon) = \mathbb{P}(\|X\| \le \epsilon) = \mathbb{P}(X \in B(0,\epsilon))$. From Theorem 3.3 in Li and Shao (2001), we have that for $\alpha_1 > 0$ and $\alpha_2 \in \mathbb{R}$,

$$-\log(\varphi_0(\epsilon)) \approx \epsilon^{-\frac{2}{\alpha_1}} |\log(\epsilon)|^{\alpha_2} \quad \text{if and only if} \quad \psi_{\mathcal{B}_{H_\mu}(0,1)}(\epsilon) \approx \left[\epsilon^{-\frac{2}{\alpha_1}} |\log(\epsilon)|^{\alpha_2}\right]^{\frac{\alpha_1}{1+\alpha_1}}. \tag{11}$$

On the other hand, using Theorem 3.1 in Li and Shao (2001) we get that, for any $x \in \mathcal{B}_{H_\mu}(0,1)$,

$$\exp(-1/2)\, \varphi_0(h) \le \mathbb{P}(X \in B(x,h)) = \mathbb{P}(\|X - x\| \le h) \le \varphi_0(h),$$

so the choice $\phi(h) = \varphi_0(h)$ satisfies A1 and A3. Stochastic processes such that $-\log(\phi(h)) \approx h^{-2/\alpha_1}(\log(1/h))^{\alpha_2}$ for $\alpha_1 > 0$ are usually known as exponential-type processes and pointwise convergence results are discussed in Ferraty and Vieu (2006).

Straightforward calculations and (11) allow to see that if we consider an exponential-type process with $\alpha_1 > 1$ and we choose $S_{\mathcal{H}} = \mathcal{B}_{H_\mu}(0,1)$ and the bandwidth h_n such that $\phi(h) = \varphi_0(h) = n^{-A}$ where $A < (\alpha_1 - 1)/(\alpha_1 + 1)$, then A3(b), A5 and A9 are fulfilled.

Examples of Gaussian processes such that $-\log(\phi(h)) \approx h^{-2/\alpha_1}(\log(1/h))^{\alpha_2}$, with $\alpha_1 > 1$, can be found in Li and Shao (1999, 2001) and Bogachev (1999). Among others, we will discuss the following two situations: (a) $\mathcal{H}$ is the Banach space of continuous functions $\mathcal{C}(0,1)$ with supremum norm $\|x\| = \sup_{t\in(0,1)} |x(t)|$ and (b) $\mathcal{H} = L^2(0,1)$ with the L^2 norm $\|x\|_2^2 = \int_0^1 x(s)^2 ds$.

(a) When considering the sup norm, the fractional Brownian motion of order $\alpha \in (1,2)$ and the integrated fractional Brownian motion of order $\alpha \in (0,2)$ are examples of exponential-type process with $\alpha_1 > 1$.

For the first one, $\alpha_1 = \alpha$ and $\beta = 0$, i.e., $-\log(\phi(h)) \approx h^{-2/\alpha}$ and $\psi_{S_{\mathcal{H}}}(\epsilon) = \epsilon^{-2/(\alpha+1)}$. Hence, the choice $h = (A \log n)^{-\alpha/2}$, with $A < (\alpha-1)/(\alpha+1)$, which leads to $\phi(h) = \varphi_0(h) = n^{-A}$, fulfills the assumptions. For this process, we have that $\theta_n = (n/\log n)^{1/(\alpha+1)} \sqrt{1/(n\phi(h))}$. Therefore, in Theorems 4.4 and 4.6 the dominating term is related to the bias, i.e., we have that $\sup_{x\in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| = O_{a.co.}(h^\eta) = O_{a.co.}\left((\log n)^{-\alpha\eta/2}\right)$ with $\eta = \eta_{\min}$ and $\eta = \eta_1$ respectively, showing the deterioration in the convergence rate.

On the other hand, for the integrated fractional Brownian motion, $\alpha_1 = 2+\alpha$ and $\beta = 0$, so that $\psi_{S_{\mathcal{H}}}(\log n/n) = (n/\log n)^{2/(\alpha+3)}$ and we can choose $h = (A \log n)^{-1-\alpha/2}$, with $A < (\alpha+1)/(\alpha+3)$.

(b) When $\mathcal{H}$ is the Hilbert space $L^2(0,1)$ and X is the fractional Brownian motion of order $\alpha \in (1,2)$, $-\log(\phi(h)) \approx h^{-2/\alpha}$, so that similar rates to those described in (a) can be derived. On the other hand, if X is the m-fold integrated Brownian motion, Theorem 1.1 in Chen and Li (2003) imply that $-\log(\phi(h)) \approx h^{-2/(2m+1)}$, i.e., the process is an exponential-type process with $\alpha_1 = (2m+1) > 1$ and $\alpha_2 = 0$, so again, the rate of convergence is dominated by the bias term

Denote as $\|x\|_p = (\int_0^1 x(s)^p dt)^{1/p}$ the L^p norm and $\|x\|_{\beta,p,q} = \|x\|_p + \left(\int_0^1 (1/t)\left(\omega_p(t,x)/t^\beta\right)^q dt\right)^{1/q}$ the Besov norm, where $\omega_p(t,x) = \sup_{|\delta|\le t}\left(\int_\delta^1 |x(s-\delta) - x(s)|^p ds\right)^{1/p}$. Furthermore, let $\|x\|_\beta = \sup_{s,t\in(0,1)} |x(t) - x(s)|/|t-s|^\beta$ be the Hölder norm. Similar results to those given above can be obtained for the fractional Brownian motion of order $\alpha \in (1,2)$ taking as norms the L^p, the Hölder or the Besov norms when $p \ge 1$ and $\beta < (\alpha-1)/2$. In those situations, we have that $\alpha_2 = 0$ while $\alpha_1 = \alpha$, for the L^p norm and $\alpha_1 = \alpha - 2\beta > 1$ for the Hölder and Besov norms. For situations in which $\alpha_1 > 1$ and $\alpha_2 > 0$, we refer again to Li and Shao (2001) who considered the Brownian sheet in $[0,1]^2$ with the sup $-L^2$ and the $L^2 -$ sup norms.

(iii) Finally, we describe a situation in which the choice of the semi-metric may increase the concentration of $\phi(h)$ around 0, to avoid that the rate of convergence deteriorates with the dimension as in (ii). In this sense, as mentioned in Ferraty et al. (2006, 2010) the choice of the semi-norm is also an important statistical tool. Let us consider the projection semi-metric in a separable Hilbert $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle$ and orthonormal basis $\{e_j : j \geq 1\}$. For a fixed integer $k > 0$, $d_k(x_1, x_2) = \left(\sum_{j=1}^{k} \langle x_1 - x_2, e_j \rangle^2\right)^{1/2}$ defines a semi-metric. 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. (2010) 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 from (i), the ϵ-entropy of $S_{\mathcal{H}}$ has order $\log(1/\epsilon)$. On the other hand, using Lemma 13.6 in Ferraty and Vieu (2006) we have that if the random vector $\mathbf{X} = \chi(X)$ has a density $f_{\mathbf{X}}$ bounded and positive over compact sets, then there exist $C_1, C_2 > 0$ such that $C_1 h^k \leq \mathbb{P}(d_k(X.x) \leq h) \leq C_2 h^k$. Hence, we can take $\phi(h) = h^k$ in A1 which means that the process is fractal of order k with respect to the semi-metric d_k and the same uniform convergence rates as in (i) are obtained.

Acknowledgments

The authors wish to thank two anonymous referees for valuable comments which led to an improved version of the original paper. This research was partially supported by Grants PIP 112-201101-00339 from CONICET, PICT 2011-0397 from ANPCYT and 20120130100241BA from the Universidad de Buenos Aires at Buenos Aires, Argentina.

Appendix A

To prove Theorems 4.1, 4.2 and 4.5, we begin by fixing some notation. For any random variable W_i such that $|W_i| \leq 1$ denote for $j = 0, 1$

$$\widetilde{R}_j(x) = \frac{1}{n} \sum_{i=1}^{n} W_i^j \frac{K_i(x)}{\mathbb{E}K_1(x)} \qquad R_j(x) = \frac{1}{n} \sum_{i=1}^{n} W_i^j \frac{K_i(x)}{\phi(h)}. \tag{A.1}$$

Note that Lemmas 4.3 and 4.4 of Ferraty and Vieu (2006), A1 and A2 imply that there are constants $0 < C < C' < \infty$ such that

$$C\phi(h) < \mathbb{E}K_1(x) < C'\phi(h) \quad \text{for any } x \in S_{\mathcal{H}}. \tag{A.2}$$

Then, if $\widetilde{C} = 1/C$, we have $|\widetilde{R}_j(x) - \mathbb{E}\widetilde{R}_j(x)| \leq \widetilde{C}\, |R_j(x) - \mathbb{E}R_j(x)|$. In order to prove Theorems 4.1 and 4.2, we will show that

$$\sup_{x \in S_{\mathcal{H}}} |R_j(x) - \mathbb{E}R_j(x)| \xrightarrow{a.co.} 0, \tag{A.3}$$

$$\sum_{n \geq 1} \mathbb{P}(\mathcal{A}_n) < \infty \quad \text{with } \mathcal{A}_n = \left\{ \inf_{x \in S_{\mathcal{H}}} \widetilde{R}_0(x) < \frac{1}{2} \right\}. \tag{A.4}$$

The following results will be helpful to derive the desired results. The proof of Lemmas A.1, A.3 and A.4 can be found in the supplementary material available online (see Appendix B).

Lemma A.1. *Let $S_{\mathcal{H}} \subset \mathcal{H}$ be a compact set. For random variables W_i, $1 \leq i \leq n$, such that $|W_i| \leq 1$ let $R_j(x)$, $j = 0, 1$ be defined as in (A.1). Assume that* A1, *to* A3 *hold, then, for $j = 0, 1$, we have that*

(a) *for all $n \geq n_0$ and for any $\epsilon > 0$*

$$\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} \mathbb{P}\{|R_j(x) - \mathbb{E}R_j(x)| > \epsilon\} \leq 2 \exp\left\{ - \frac{\epsilon^2 n \phi(h)}{2C' \|K\|_\infty^2 \left(1 + \frac{\epsilon}{C' \|K\|_\infty}\right)} \right\}.$$

(b) *There exist $a_1 > 0$, $a_2 > 0$ such that, for all $n \geq n_0$ and for any $\epsilon > 0$*

$$\mathbb{P}\left\{ \sup_{x \in S_{\mathcal{H}}} |R_j(x) - \mathbb{E}R_j(x)| > \epsilon \right\} \leq 8 N_\rho(S_{\mathcal{H}}) \exp\left\{ - \frac{\epsilon^2 n \phi(h)}{a_1(1 + a_2 \epsilon)} \right\},$$

where ρ is such that $\rho/h \to 0$ and $\rho/\phi(h) \to 0$ when $n \to \infty$.

(c) *Let $\theta_n^2 = \psi_{S_{\mathcal{H}}}(\log(n)/n)/(n\phi(h))$. There exists $c > 2$ such that, for any $\epsilon_0 > c$ and $n \geq n_0$,*

$$\sup_{y \in \mathbb{R}} \mathbb{P}\left\{ \theta_n^{-1} \sup_{x \in S_{\mathcal{H}}} |R_j(x) - \mathbb{E}R_j(x)| > \epsilon_0 \right\} \leq 8 \exp\left\{ \left(1 - \frac{\epsilon_0^2}{8(1 + \epsilon_0)}\right) \psi_{S_{\mathcal{H}}}\left(\frac{\log n}{n}\right) \right\}.$$

Corollary A.2. *Let $S_{\mathcal{H}}$ be a compact set and assume that* A1 *to* A3 *and* A5 *hold. Then, for any random variable W_i such that $|W_i| \leq 1$, (A.3) and (A.4) hold.*

Proof. Using Lemma A.1(b) with $\rho_n = \log(n)/n$ and the fact that A3(b) implies that for $n \geq n_0$, $\psi_{S_{\mathcal{H}}}(\rho_n)/(n\phi(h)) < (1/2)\ \epsilon^2/(a_1(1 + a_2\epsilon))$, we get that for $n \geq n_0$, $\mathbb{P}\left\{\sup_{x \in S_{\mathcal{H}}} |R_j(x) - \mathbb{E}R_j(x)| > \epsilon\right\} \leq 8 \exp\left\{-\epsilon^2\, n\phi(h)/[2\, a_1(1 + a_2\epsilon)]\right\}$.

Therefore, using that $n\phi(h)/\log n \to \infty$ we have that (A.3) holds. On the other hand, using that $\mathbb{E}\widetilde{R}_0(x) = 1$ and $\inf_{x\in S_{\mathcal{H}}} \widetilde{R}_0(x) \geq \inf_{x\in S_{\mathcal{H}}} \mathbb{E}\widetilde{R}_0(x) - \sup_{x\in S_{\mathcal{H}}} |\widetilde{R}_0(x) - \mathbb{E}\widetilde{R}_0(x)|$, together with the fact that $|\widetilde{R}_j(x) - \mathbb{E}\widetilde{R}_j(x)| \leq \widetilde{C}\,|R_j(x) - \mathbb{E}R_j(x)|$, from (A.3) we get that (A.4) holds. □

Lemma A.3. *Let $\widetilde{R}_j(x)$ be defined in (A.1) for $j = 0, 1$ with $W_i = \mathbb{I}_{(-\infty,y]}(Y_i)$. Under* A1, A2 *and* A7*, if $h_n \to 0$, we have that*

(a) $\sup_{y\in\mathbb{R}} \sup_{x\in S_{\mathcal{H}}} |\mathbb{E}\widetilde{R}_1(x) - F(y|X = x)\mathbb{E}\widetilde{R}_0(x)| \to 0.$

(b) *If in addition* A10 *holds, then* $\sup_{y\in\mathbb{R}} \sup_{x\in S_{\mathcal{H}}} |\mathbb{E}\widetilde{R}_1(x) - F(y|X = x)\mathbb{E}\widetilde{R}_0(x)| = O(h^{\eta_1})$.

Note that from Lemma A.3(b), we also have that $\sup_{y\in\mathbb{R}} \sup_{x\in S_{\mathcal{H}}} |\mathbb{E}\widetilde{R}_1(x) - F(y|X = x)\mathbb{E}\widetilde{R}_0(x)| = O\,(h^{\eta_1} + \theta_n)$, where $\theta_n^2 = \psi_{S_{\mathcal{H}}}(\log(n)/n)/(n\phi(h))$.

Lemma A.4. *Let $S_{\mathcal{H}}$ be a compact set and $F_n(y|X = x)$ be a sequence of conditional distribution functions verifying*

$$\sup_{x\in S_{\mathcal{H}}} \sup_{y\in\mathbb{R}} |F_n(y|X = x) - F(y|X = x)| \to 0. \tag{A.5}$$

Then, if F verifies Assumption A7*, there exist positive constants $A \leq B$ such that $s_n(x) = \text{MAD}_c(F_n(\cdot|X = x))$ verifies $A \leq s_n(x) \leq B$ for all $x \in S_{\mathcal{H}}$ and $n \geq n_0$.*

Proof of Theorem 4.1. Using that $\lambda(\cdot, a, \sigma) : \mathcal{H} \to \mathbb{R}$ is a continuous function on $S_{\mathcal{H}}$, it is easy to see that g is also continuous on $S_{\mathcal{H}}$ and that, for each fixed a, $\lambda(u, g(u) + a, \sigma)$ is continuous for $(u, \sigma) \in S_{\mathcal{H}} \times [A, B]$. For details see the proof of Theorem 4.3(a) available online (see Appendix B). Let us begin by showing that

$$\sup_{x\in S_{\mathcal{H}}} |\lambda(x, g(x) + a, \widehat{s}(x)) - \widehat{\lambda}(x, g(x) + a, \widehat{s}(x))| \xrightarrow{a.s.} 0 \tag{A.6}$$

entails that $\sup_{x\in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| \xrightarrow{a.s.} 0$. Effectively, given $\epsilon > 0$, the continuity of $\lambda(x, g(x) \pm \epsilon, \sigma)$, the fact that $g(x)$ is the unique solution of $\lambda(x, a, \sigma) = 0$ and A4 imply that

$$\lambda_1 = \sup_{A\leq\sigma\leq B} \sup_{x\in S_{\mathcal{H}}} \lambda(x, g(x) + \epsilon, \sigma) < 0 < \inf_{A\leq\sigma\leq B} \inf_{x\in S_{\mathcal{H}}} \lambda(x, g(x) - \epsilon, \sigma) = \lambda_2. \tag{A.7}$$

So, if (A.6) holds, from (A.7) we conclude that for n large enough

$$\widehat{\lambda}(x, g(x) + \epsilon, \widehat{s}(x)) < \frac{\lambda_1}{2} < 0 < \frac{\lambda_2}{2} < \widehat{\lambda}(x, g(x) - \epsilon, \widehat{s}(x)),$$

for all $x \in S_{\mathcal{H}}$ almost surely, which entails that $\mathbb{P}(\sup_{x\in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| < \epsilon) = 1$ as desired.

Note that (A.6) follows immediately if we show that for each fixed a

$$A_n = \sup_{x\in S_{\mathcal{H}}} \sup_{A\leq\sigma\leq B} |\widehat{\lambda}(x, g(x) + a, \sigma) - \lambda(x, g(x) + a, \sigma)| \xrightarrow{a.s.} 0.$$

Denote $W_{i,\sigma}(x) = \psi\left((Y_i - g(x) - a)/\sigma\right)/\|\psi\|_\infty$ and define $\widetilde{R}_1(x, \sigma)$ and $R_1(x, \sigma)$ as in (A.1), where we strength the dependence on σ. Then, $\widehat{\lambda}(x, g(x) + a, \sigma) = \|\psi\|_\infty \widetilde{R}_1(x, \sigma)/\widetilde{R}_0(x)$, so

$$\frac{1}{\|\psi\|_\infty} |\widehat{\lambda}(x, g(x) + a, \sigma) - \lambda(x, g(x) + a, \sigma)| \leq \frac{1}{\inf_{x\in S_{\mathcal{H}}} \widetilde{R}_0(x)} \left[\widetilde{B}_{0,n} + \widetilde{B}_{1,n} + \widetilde{B}_{2,n}\right], \tag{A.8}$$

where $\widetilde{B}_{0,n} = \sup_{x\in S_{\mathcal{H}}} |\widetilde{R}_0(x) - \mathbb{E}\widetilde{R}_0(x)|$, $\widetilde{B}_{1,n} = \sup_{x\in S_{\mathcal{H}}} \sup_{A\leq\sigma\leq B} |\widetilde{R}_1(x, \sigma) - \mathbb{E}\widetilde{R}_1(x, \sigma)|$ and

$$\widetilde{B}_{2,n} = \sup_{x\in S_{\mathcal{H}}} \sup_{A\leq\sigma\leq B} \frac{1}{\|\psi\|_\infty} \left|\frac{\mathbb{E}\left[\lambda(X_1, g(x) + a, \sigma) - \lambda(x, g(x) + a, \sigma)\right] K_1(x)}{\mathbb{E}K_1(x)}\right|.$$

Using that, for each fixed a, $\lambda(x, g(x) + a, \sigma)$ is continuous for $(x, \sigma) \in S_{\mathcal{H}} \times [A, B]$ and that $S_{\mathcal{H}}$ is a compact set, we obtain easily that given $\eta > 0$ there exists $\delta > 0$ such that for any $x \in S_{\mathcal{H}}$ and for any u such that $d(u, x) < \delta$ we have that $\sup_{\sigma\in[A,B]} |\lambda(u, g(u) + a, \sigma) - \lambda(x, g(x) + a, \sigma)| < \eta/2$ and $|g(u) - g(x)| < (\eta/2)A\|\psi'\|_\infty$. Hence, $\sup_{\sigma\in[A,B]} |\lambda(u, g(x) + a, \sigma) - \lambda(x, g(x) + a, \sigma)| < \eta$. This bound and similar arguments to those considered in the proof of Lemma A.3(a) allow to show that $\widetilde{B}_{2,n} \to 0$. Hence, for any fixed $\epsilon > 0$, we have that $\widetilde{B}_{2,n} < \epsilon$, for $n \geq n_0$. Using (A.8), we conclude that for $n \geq n_0$, $\mathbb{P}(A_n > 6\,\epsilon\|\psi\|_\infty) \leq \mathbb{P}(\mathcal{A}_n) + \mathbb{P}(\widetilde{B}_{0,n} > \epsilon) + \mathbb{P}(\widetilde{B}_{1,n} > \epsilon)$, where $\mathcal{A}_n$, defined in (A.4), is such that $\sum_{n\geq1} \mathbb{P}(\mathcal{A}_n) < \infty$. On the other hand, Corollary A.2 entails that $\sum_{n\geq1} \mathbb{P}(\widetilde{B}_{0,n} > \epsilon) < \infty$. Hence, it only remains to show that $\sum_{n\geq1} \mathbb{P}(\widetilde{B}_{1,n} > \epsilon) < \infty$ for any $\epsilon > 0$.

Recall that $|\widetilde{R}_1(x,\sigma) - \mathbb{E}\widetilde{R}_1(x,\sigma)| \le |R_1(x,\sigma) - \mathbb{E}R_1(x,\sigma)|/C$ with C given in (A.2). Therefore, it is enough to show that $\sum_{n\ge 1}\mathbb{P}(B_{1,n} > \epsilon) < \infty$ where $B_{1,n} = \sup_{x\in S_{\mathcal{H}}}\sup_{A\le\sigma\le B}|R_1(x,\sigma) - \mathbb{E}R_1(x,\sigma)|$. Since $W_{i,\sigma}(x) \le 1$, Lemma A.1(b) entail that for $\rho = \log(n)/n$

$$\sup_{A\le\sigma\le B}\mathbb{P}\left\{\sup_{x\in S_{\mathcal{H}}}|R_1(x,\sigma) - \mathbb{E}R_1(x,\sigma)| > \epsilon\right\} \le 8N_\rho(S_{\mathcal{H}})\exp\left\{-\frac{\epsilon^2 n\phi(h)}{a_1(1+a_2\epsilon)}\right\}.$$

Let $M_n = \max(N_\rho(S_{\mathcal{H}}), n)$ and $\nu_n = (B-A)/M_n \to 0$. Let us consider a finite covering of $[A, B]$ with M_n intervals $\mathcal{I}_j = [\sigma_j, \sigma_{j+1}]$ such that $|\sigma_{j+1} - \sigma_j| \le \nu_n$. Using that $\zeta(t) = t\psi'(t)$ is bounded, we get that for $\sigma \in \mathcal{I}_j$, $|R_1(x,\sigma) - S_1(x,a,\sigma_j)| \le (\|\zeta\|_\infty/A)|\sigma - \sigma_j|\sum_{i=1}^n K_i(x)/n\phi(h) \le (\|\zeta\|_\infty/A)|\sigma-\sigma_j|R_0(x)$. Therefore, using that $\mathbb{E}R_0(x) = \mathbb{E}K_1(x)/\phi(h) \le C'$, we obtain the bound

$$B_{1,n} \le \max_{1\le j\le N}\sup_{x\in S_{\mathcal{H}}}|R_1(x,\sigma_j) - \mathbb{E}R_1(x,\sigma_j)| + \frac{\|\zeta\|_\infty}{A}\nu_n\sup_{x\in S_{\mathcal{H}}}|R_0(x) - \mathbb{E}R_0(x)| + 2C'\frac{\|\zeta\|_\infty}{A}\nu_n. \tag{A.9}$$

The fact that $\nu_n \to 0$, (A.9) and Lemma A.1(b) entail that

$$\mathbb{P}\left(B_{1,n} > 3\epsilon\right) \le 8\left(1+M_n\right)N_\rho(S_{\mathcal{H}})\exp\left\{-\frac{\epsilon^2 n\phi(h)}{a_1(1+a_2\epsilon)}\right\} \le 16\exp\left\{-\frac{\epsilon^2 n\phi(h)}{a_1(1+a_2\epsilon)} + 2\psi_{S_{\mathcal{H}}}(\rho_n) + \log n\right\}.$$

Note that A3(b) and A5 imply that $(2\psi_{S_{\mathcal{H}}}(\rho_n) + \log n)/(n\phi(h)) < \epsilon^2/[4a_1(1+a_2\epsilon)]$, for $n \ge n_0$. Thus, $\mathbb{P}\left(B_{1,n} > 3\epsilon\right) \le 16\exp\left\{-\epsilon^2 n\phi(h)/[2a_1(1+a_2\epsilon)]\right\}$ and the proof is concluded since that $n\phi(h)/\log n \to \infty$ entails that $\sum_{n\ge 1}\mathbb{P}\left\{B_{1,n} > 3\epsilon\right\} < \infty$. □

Proof of Theorem 4.2. Let $y \in \mathbb{R}$ be fixed and take $W_i = \mathbb{I}_{(-\infty,y]}(Y_i)$ in (A.1). Then, $\widehat{F}(y|X=x) = \widetilde{R}_1(x)/\widetilde{R}_0(x)$ and as in Collomb (1982), we have the following bound

$$\begin{aligned}&\sup_{x\in S_{\mathcal{H}}}|\widehat{F}(y|X=x) - F(y|X=x)| \\ &\quad\le \frac{1}{\inf_{x\in S_{\mathcal{H}}}\widetilde{R}_0(x)}\left[\sup_{x\in S_{\mathcal{H}}}|\widetilde{R}_1(x) - \mathbb{E}\widetilde{R}_1(x)| + \sup_{x\in S_{\mathcal{H}}}|\widetilde{R}_0(x) - \mathbb{E}\widetilde{R}_0(x)| + \sup_{x\in S_{\mathcal{H}}}|\mathbb{E}\widetilde{R}_1(x) - F(y|X=x)\mathbb{E}\widetilde{R}_0(x)|\right].\end{aligned} \tag{A.10}$$

Corollary A.2, Lemma A.3(a) and (A.10) entail that $\sup_{x\in S_{\mathcal{H}}}|\widehat{F}(y|X=x) - F(y|X=x)| \xrightarrow{a.s.} 0$. The proof follows now using similar arguments to those considered in the proof of Theorem 3.1 in Boente and Fraiman (1991). □

Proof of Theorem 4.3. The proof of part (a) can be found in the supplement available online (see Appendix B). To prove (b), we begin by showing that $m(x)$ is continuous at any $x_0 \in S_{\mathcal{H}}$. The fact that there is only one median entails that for any $\eta > 0$ $F(m(x_0)+\eta|X=x_0) > F(m(x_0)|X=x_0) > F(m(x_0)-\eta|X=x_0)$. Moreover, A7(b) implies that given $\epsilon > 0$ there exists $\delta > 0$ such that $d(x_0,x) < \delta$ implies $|F(m(x_0)\pm\eta|X=x) - F(m(x_0)\pm\eta|X=x_0)| < \epsilon$.

Let $\epsilon_\eta = \min\{F(m(x_0)|X=x_0) - F(m(x_0)-\eta|X=x_0), F(m(x_0)+\eta|X=x_0) - F(m(x_0)|X=x_0)\}$. Following the same ideas used in the proof of (a), we obtain that if $d(x_0,x) < \delta_1$ then $F(m(x_0)-\eta|X=x) < F(m(x_0)-\eta|X=x_0)+\epsilon_\eta < F(m(x_0)|X=x_0) = 1/2$, so that $m(x_0) - m(x) < \eta$. Similarly we have that $m(x_0) - m(x) > -\eta$ if $d(x_0,x) < \delta_2$ for some $\delta_2 > 0$, which leads to the continuity of m at x_0.

Let $\epsilon > 0$, A7(b) entails that there exists $\delta > 0$ such that $|u - v| < \delta$ implies that $\sup_{x\in S_{\mathcal{H}}}|F(u|X=x) - F(v|X=x)| < \epsilon/2$. Define $\delta_1 = \min(\delta, \epsilon/4)$ and $\delta_2 = \delta_1/4 < \epsilon/8$. Therefore, for any $x \in S_{\mathcal{H}}$, we have that $|F(m(x)+\delta_2|X=x) - F(m(x)-\delta_2|X=x)| < \epsilon/2$. Using that $F(m(x)|X=x) = 1/2$, we obtain that $1/2 < F(m(x)+\delta_2|X=x) < 1/2+\epsilon/2$ and $1/2-\epsilon/2 < F(m(x)-\delta_2|X=x) < 1/2$. The continuity of m and $F(y|X=\cdot)$ implies that $i_{\delta_2} = \inf_{x\in S_{\mathcal{H}}}F(m(x)+\delta_2|X=x) > 1/2$ and $s_{\delta_2} = \sup_{x\in S_{\mathcal{H}}}F(m(x)-\delta_2|X=x) < 1/2$. Let $\nu = \min\{i_{\delta_2} - 1/2, 1/2 - s_{\delta_2}\} > 0$ and $\epsilon_1 = \min\{\nu/2, \epsilon/2\}$. Theorem 4.2 implies that $\mathbb{P}(\mathcal{N}) = 0$, where $\mathcal{N} = \{\omega \in \Omega : \sup_{x\in S_{\mathcal{H}}}\sup_{y\in\mathbb{R}}|\widehat{F}(y|X=x) - F(y|X=x)| \not\to 0\}$. Let $\omega \notin \mathcal{N}$, then there exists n_0 such that if $n \ge n_0$, for all $x \in S_{\mathcal{H}}$ and for all $y \in \mathbb{R}$, we have that $|\widehat{F}(y|X=x) - F(y|X=x)| < \epsilon_1$. In particular, for $j = 0, 1$, $|\widehat{F}\left(m(x)+(-1)^j\delta_2|X=x\right) - F\left(m(x)+(-1)^j\delta_2|X=x\right)| < \epsilon_1$, that is, we have that for $j = 0, 1$

$$\left(m(x)+(-1)^j\delta_2|X=x\right) - \epsilon_1 < \widehat{F}\left(m(x)+(-1)^j\delta_2|X=x\right) < F\left(m(x)+(-1)^j\delta_2|X=x\right)+\epsilon_1. \tag{A.11}$$

On the other hand, using that $F(m(x)-\delta_2|X=x) < s_\delta < 1/2$ and $\epsilon_1 < \nu/2$, we get that for all $x \in S_{\mathcal{H}}F(m(x)-\delta_2|X=x)+\epsilon_1 < 1/2$. Likewise, we have $F(m(x)+\delta_2|X=x)-\epsilon_1 > 1/2$. Hence, from (A.11), we obtain that $\widehat{F}(m(x)-\delta_2|X=x) < 1/2$ and $\widehat{F}(m(x)+\delta_2|X=x) > 1/2$. As $\widehat{F}(\widehat{m}(x)|X=x) = 1/2$, we conclude that $m(x)-\delta_2 < \widehat{m}(x) < m(x)+\delta_2$, i.e., $\sup_{x\in S_{\mathcal{H}}}|\widehat{m}(x) - m(x)| < \epsilon/8$. □

Appendix B. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.spl.2015.01.028.

References

Attouch, M., Laksaci, A., Ould-Said, E., 2009. Asymptotic distribution of robust estimators for functional nonparametric models. Comm. Statist. Theory Methods 38, 1317–1335.
Azzedine, N., Laksaci, A., Ould-Said, E., 2008. On robust nonparametric regression estimation for a functional regressor. Statist. Probab. Lett. 78, 3216–3221.
Boente, G., Fraiman, R., 1989. Robust nonparametric regression estimation for dependent observations. Ann. Statist. 17, 1242–1256.
Boente, G., Fraiman, R., 1991. Strong uniform convergence rates for some robust equivariant nonparametric regression estimates for mixing processes. Int. Stat. Inst. 59, 355–372.
Bogachev, V.I., 1999. Gaussian Measures. In: Mathematical Surveys and Monographs, vol. 62. American Mathematical Society.
Burba, F., Ferraty, F., Vieu, Ph., 2009. k-nearest neighbour method in functional nonparametric regression. J. Nonparametr. Stat. 21, 453–469.
Cadre, B., 2001. Convergent estimators for the L1-median of a Banach valued random variable. Statistics 35, 509–521.
Chen, X., Li, W.V., 2003. Quadratic functionals and small ball probabilities for the m-fold integrated Brownian motion. Ann. Probab. 31, 1052–1077.
Collomb, G., 1982. Prédiction non paramétrique: étude de l'erreur quadratique du prédictogramme. C. R. Acad. Sci., Paris I 294, 59–62.
Collomb, G., Härdle, W., 1986. Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations. Stochastic Process. Appl. 23, 77–89.
Crambes, C., Delsol, L., Laksaci, A., 2008. Robust nonparametric estimation for functional data. In: Dabo-Niang, , Ferraty, (Eds.), International Workshop on Functional and Operatorial Statistics 2008 Proceedings, Functional and Operatorial Statistics. Physica-Verlag, Springer.
Ferraty, F., Laksaci, A., Vieu, Ph., 2006. Estimating some characteristics of the conditional distribution in nonparametric functional models. Stat. Inference Stoch. Process. 9, 47–76.
Ferraty, F., Laksaci, A., Vieu, Ph., 2010. Rate of uniform consistency for nonparametric estimates with functional variables. J. Statist. Plann. Inference 140, 335–352.
Ferraty, F., Romain, Y., 2011. The Oxford Handbook of Functional Data Analysis. Oxford University Press.
Ferraty, F., Vieu, P., 2004. Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. J. Nonparametr. Stat. 16, 111–125.
Ferraty, F., Vieu, Ph., 2006. Nonparametric Functional Data Analysis: Theory and Practice. In: Springer Series in Statistics, Springer, New York.
Härdle, W., 1990. Applied Nonparametric Regression. Cambridge University Press.
Huber, P.J., 1964. Robust estimation of a location parameter. Ann. Math. Statist. 35, 73–101.
Kolmogorov, A.N., 1956. Asymptotic characteristics of some completely bounded metric spaces. Dokl. Akad. Nauk SSSR 108, 585–589.
Li, W.V., Shao, Q.M., 1999. Small ball estimates for Gaussian processes under Sobolev type norms. J. Theoret. Probab. 12, 699–720.
Li, W.V., Shao, Q.M., 2001. Gaussian processes: inequalities, small ball probabilities and applications. In: Rao, C.R., Shanbhag, D. (Eds.), Stochastic Processes: Theory and Methods, Handbook of Statistics, vol. 19. North-Holland, Amsterdam, pp. 533–597.
Maronna, R., Martin, R., Yohai, V., 2006. Robust Statistics, Theory and Methods. John Wiley & Sons, Ltd..
Ramsay, J., Silverman, B., 2002. Applied Functional Data Analysis: Methods and Case Studies. Springer.
Ramsay, J., Silverman, B., 2005. Functional Data Analysis. In: Springer Series in Statistics.
van de Geer, S., 2000. Empirical Processes in M-Estimation. Cambridge University Press.