CHILEAN JOURNAL OF STATISTICS

Edited by Víctor Leiva and Carolina Marchant

A free open access journal indexed by

Volume 11 Number 2
December 2020

ISSN: 0718-7912 (print)
ISSN: 0718-7920 (online)

Published by the
Chilean Statistical Society
SOCHE
SOCIEDAD CHILENA DE ESTADÍSTICA

Aims

The Chilean Journal of Statistics (ChJS) is an official publication of the Chilean Statistical Society (`www.soche.cl`). The ChJS takes the place of *Revista de la Sociedad Chilena de Estadística*, which was published from 1984 to 2000.

The ChJS covers a broad range of topics in statistics, as well as in artificial intelligence, big data, data science, and machine learning, focused mainly on research articles. However, review, survey, and teaching papers, as well as material for statistical discussion, could be also published exceptionally. Each paper published in the ChJS must consider, in addition to its theoretical and/or methodological novelty, simulations for validating its novel theoretical and/or methodological proposal, as well as an illustration/application with real data.

The ChJS editorial board plans to publish one volume per year, with two issues in each volume. On some occasions, certain events or topics may be published in one or more special issues prepared by a guest editor.

Contents

Chilean Journal of Statistics
Vol. 11, No. 2, December 2020, 73–93

Functional and Nonparametric Statistics
Research Paper

Robust kernel regression estimator of the scale parameter for functional ergodic data with applications

Ibrahim M. Almanjahie[1], Mohammed Kadi Attouch[2], Omar Fetitah[2], and Hayat Louhab[2]

[1]College of Science, Department of Mathematics, King Khalid University, Saudi Arabia
[2]Laboratoire de Probabilités Statistique Processus Stochastiques, Sidi Bel Abbes University, Algeria

(Received: 19 October 2020 · Accepted in final form: 22 December 2020)

Abstract

In this paper, we propose a family of robust nonparametric estimators for a regression function with unknown scale parameter based on the kernel method. We establish the asymptotic normality of the estimators for functional explanatory variables when the observations exhibit some kind of dependence (stationary ergodic process). This approach can be used for predicting and for building confidence regions. A simulation study is conducted to support our theoretical results and to exhibit the good behavior of the proposed estimator for finite samples with different rates of dependency, and particularly in the presence of several outliers in the data set. In addition, a real data study is provided to illustrate the good behavior of our estimator.

Keywords: Confidence bands · Functional data · Lindeberg condition · Nonparametric kernel estimate · Robust equivariant regression.

Mathematics Subject Classification: Primary 62G35 · Secondary 62G20.

1. Introduction

Nonparametric kernel regression estimation is a familiar tool to explore the underlying relation between the response variable and covariates. In the functional data studies, these estimators are largely studied in Ramsay and Silverman (2002), and Ferraty and Vieu (2006). As in parametric regression estimation, the kernel estimator may be affected by outliers and then it is needed to consider robustness estimation.

Recall that robust regression modeling is an old subject in statistics. It was started by Huber (1964) who studied estimation of a location parameter. We cite Collomb and Hardle (1986) and Laïb and Ould Saïd (2000) for some results on multivariate time series (mixing and ergodicity conditions). Robust regression is widely studied in nonparametric functional statistics. Indeed, it was firstly introduced by Azzedine et al. (2008) who proved the almost complete convergence of this model in the independent and identically distributed case. Since their work, several results on nonparametric robust functional regression were considered. Key references on this topic are Crambes et al. (2008), Chen and Zhang (2009), Attouch et

*Corresponding author. Email: attou_kadi@yahoo.fr

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

al. (2009), Attouch et al. (2010), Gheriballah et al. (2013), Boente and Fraiman (1990) and the references therein.

Notice that all these results are obtained when the scale parameter is known. Boente and Vahnovan (2015, 2017) proposed robust equivariant M-estimators for regression and partial linear models. In this paper, we consider the more general case, that is, when the scale is unknown and the data are dependent. Specifically we model ergodic functional time series.

It is well known that ergodicity is a fundamental hypothesis in statistical physics, thermodynamics and signal processing. In all these areas, ergodicity is studied on a continuous path. Thus, it is necessary to develop statistical tools allowing one to treat the continuous ergodic process in its own dimension by exploring its functional character. This is the general framework of the present work.

Note that the ergodicity assumption is less restrictive than the mixing condition usually assumed in functional time series studies. In particular, this is implied by most mixing conditions. The literature on ergodic functional time series data is still limited. The few existing results are in Laïb and Louani (2011, 2010), Gheriballah et al. (2013), Benziadi et al. (2016a,b). Among the extensive literature on functional data analysis, we only refer to the overviews for parametric models given by Bosq (2000), Ramsay and Silverman (2002) and to the monograph of Ferraty and Vieu (2006) for nonparametric models.

The main objective of this paper is to generalize the results of Boente and Vahnovan (2015) from the independent case to the ergodic case. Specifically we prove the asymptotic normality of an estimator constructed by combining the concepts of robustness with those of unknown scale parameter. This result is obtained under standard conditions allowing us to explore the different structural axes of the subject, such as the robustness of the regression function and the correlation between the observations. We point out that, unlike the case of fixed scale, here the scale parameter must be estimated, which makes the establishment of its asymptotic properties more difficult.

The reminder of this paper is organized as follows. Section 2 is dedicated to the presentation of the robust estimator with unknown scale parameter. The needed assumptions and notations are given in Section 3. We state and proof our main results in Section 4. Some simulation results are reported in Section 5 to compare the M-estimator (for known and unknown scale parameter) with the kernel regression estimator. Section 6 deals with a real data application. The proofs of the main results are relegated to the Appendix. In Section 7, the main conclusions of this study and ideas for future research are provided.

2. The robust equivariant estimators and their related functional

Let $(X_i, Y_i)_{i=1,\ldots,n}$ be a sequence of strictly stationary dependent random variables and identically distributed as (X, Y), which is a random pair valued in $\mathcal{F} \times \mathbb{R}$, where $(\mathcal{F}, d)$ is a semi-metric space. We study the nonparametric estimation of the robust regression $\theta(x)$, when the scale parameter is unknown and strong dependencies are present (ergodicity). In fact, for any $x \in \mathcal{F}$, $\theta(x)$ is defined as a zero with respect to the parameter a by means of

$$\Psi\left(x, a, \sigma\right) = \mathrm{E}\left[\psi_x\left(\frac{Y-a}{\sigma}\right)|X=x\right] = 0,$$

where ψ_x is a real valued function which satisfies some regularity conditions, to be stated below, and σ is a robust measure of the conditional scale. In what follows, we assume, for all $x \in \mathcal{F}$, that the robust regression $\theta(x)$ exists and is unique; see, for example, Boente and Fraiman (1989).

Consider a functional stationary ergodic process $Z_i = (X_i, Y_i)_{i=1,\dots,n}$; see Laïb and Louani (2011) for some definitions and examples. When the scale parameter is unknown, a robust estimator may be constructed following two steps. Firstly, we estimate the scale parameter σ by the local median of the absolute deviation from the conditional median (MED), $\widehat{m}_{\mathrm{MED}}(x)$, of the conditional distribution of Y given $X = x$, denoted $F(y|X = x) = \mathrm{E}(\mathbb{1}_{(-\infty,y]}(Y)|X = x)$, for any $y \in \mathrm{R}$, where $\mathbb{1}_A$ denotes the indicator function on the set A. Then, for $x \in \mathcal{F}$, the kernel estimator $\widehat{s}(x)$ of $\sigma(x)$ is the zero of the equation given by $\widehat{F}\,(s|X = x) = 1/2$, with

$$\widehat{F}\,(y|X=x) = \frac{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)\mathbb{1}_{(-\infty,y]}\,(Y_i)}{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)},$$

where K is a kernel function, $d(x, X_i)$ denotes the distance between the fixed point x and the realization of the functional random variable X_i, and the bandwidth parameter $h = h_n$ is a sequence of positive numbers which goes to zero as n goes to infinity. Next, the kernel estimator $\widehat{\theta}(x)$, of the robust regression $\theta(x)$, is the zero, with respect to a, of $\widehat{\Psi}(x, a, \widehat{s}) = 0$, where

$$\widehat{\Psi}\,(x,a,\widehat{s}) = \frac{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)\psi_x\left(\frac{Y_i-a}{\widehat{s}}\right)}{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)}.$$

3. Notations, hypotheses and comments

Throughout the paper, when no confusion is possible, C and $C^{'}$ are some strictly positive generic constants, x is a fixed point in $\mathcal{F}$ and $\mathcal{N}_x$ is a fixed neighborhood of x. For $r > 0$, let $B(x, r) := \{x^{'} \in \mathcal{F}/d(x^{'}, x) < r\}$. Moreover, for $i = 1, \dots, n$, $\mathcal{F}_k$ is the $\sigma-$field generated by $((X_1, Y_1), \dots, (X_k, Y_k))$ and we pose $\mathcal{B}_k$ is the $\sigma-$field generated by $((X_1, Y_1), \dots, (X_k, Y_k), X_{k+1})$.

Our basic assumptions are:

(A1) The function ψ_x is continuous and monotone in the second component.

(A2) The processes $(X_i, Y_i)_{i\in N}$ satisfies: (i) $\phi\,(x, r) = \mathrm{P}\,(X \in B\,(x, r)) > 0$, and $\phi_i\,(x, r) = \mathrm{P}\,(X_i \in B\,(x, r)\,|\mathcal{F}_{i-1}) > 0$, $\forall r > 0$; and (ii) for all $r > 0, 1/(n\phi\,(x, r)) \sum_{i=1}^{n} \phi_i\,(x, r) \xrightarrow{\mathrm{P}} 1$, and $n\phi\,(x, h) \to \infty$ as $h \to 0$, with $\xrightarrow{\mathrm{P}}$ meaning convergence in probability.

(A3) The function Ψ is such that: (i) the function $\Psi\,(x, ., \sigma)$ is of class $\mathcal{C}^1$ in $\mathcal{N}_x$, a fixed neighborhood of $\theta(x)$; (ii) for each fixed t in $\mathcal{N}_x$, the functions $\Psi\,(., t, \sigma)$, and $\lambda_2\,(\cdot, t, \sigma) = \mathrm{E}\left[\psi_x^2\,((Y - t)/\sigma)\,|X = \cdot\right]$ are continuous at x; and (iii) the derivative of $\Phi\,(x, z, \sigma) = \mathrm{E}\,[\Psi\,(X_1, z, \sigma) - \Psi\,(x, z, \sigma)\,|d\,(x, X_1) = s]$ exists at $s = 0$, and is continuous in the second component in $\mathcal{N}_x$.

(A4) For each fixed t in the neighborhood of $\theta(x)$, and $\forall j \geq 2$,

$$\mathrm{E}\left[\psi_x^j\,((Y-t)/\sigma)\,|\mathcal{B}_{i-1}\right] = \mathrm{E}\left[\psi_x^j\,((Y-t)/\sigma)\,|X_i\right] < c < \infty, \text{a.s.},$$

with "a.s." meaning almost sure convergence.

(A5) The kernel K is a positive function with support in $(0, 1)$, its derivative $K^{'}$ exists in $(0, 1)$, and satisfies $K^{'}\,(t) < 0$ for $0 < t < 1$.

(A6) There exists a function τ_x, such that $\forall t \in [0, 1]$, $\lim_{h\to 0} \phi\,(x, th)/\phi\,(x, h) = \tau_x\,(t)$, $K^2\,(1) - \int_0^1 \left(K^2\,(u)\right)^{'} \tau_x\,(u)\,\mathrm{d}u > 0$ and $K\,(1) - \int_0^1 K^{'}\,(u)\,\tau_x\,(u)\,\mathrm{d}u \neq 0$.

(A7) The functions $f_x(x)$ and $p(x)$ are bounded on $\mathcal{S}$ such that $A_p = \inf_{x\in\mathcal{S}} p(x) > 0$, and $A_f = \inf_{x\in\mathcal{S}} f_x(x) > 0$. Moreover, $p(x)$ is a continuous function in a neighborhood of $\mathcal{S}$.

(A8) First, we have that: (i) $F(y|X = x)$ is a continuous function of x in a neighborhood of $\mathcal{S}$ and besides it satisfies the equicontinuity condition $\forall\varepsilon > 0, \exists\delta > 0 : |u - v| < \delta \Longrightarrow \sup_{x\in\mathcal{S}} (|F(U|X = x) - F(v|X = x)|) < \varepsilon$; and second (ii) $F(y|X = x)$ is symmetric around $\theta(x)$ and a continuous function of y for each fixed x.

(A9) The sequence $h = h_n$ is such that $h_n \to 0$, $n\phi(h) \to \infty$ and $(n\phi(h))/n \to \infty$.

(A10) The sequence $k = k_n$ is such that $k_n/n \to 0$, $k_n \to \infty$ and $k_n/\log(n) \to \infty$.

Remark It is well known that a fundamental property of robust M-estimators is the convexity and the boundedness of the score function. Convexity is important for the existence and uniqueness of the estimate, whereas the boundedness is essential for reducing the influence of atypical values. In this work, convexity is controlled by means of the monotonicity condition (A1). However, we opt for a presentation without the boundedness condition to cover, for example, the classical regression, which is studied under the ergodic process framework by Laïb and Louani (2011). Assumptions (A2) and (A3) are the same conditions used in Gheriballah et al. (2013), while conditions (A4), (A5) and (A6) are very similar to those used by Ferraty et al. (2010). In addition, (A7) and (A8) are the regularity conditions on the marginal density of X and on the conditional distribution function which imply that, for any set $\mathcal{S} \in \mathcal{F}$, $0 < \inf_{x\in\mathcal{S}} s(x) \le \sup_{x\in\mathcal{S}} s(x) < \infty$ and that $\theta(x)$ is a continuous function of x. Assumptions (A9) and (A10) are standard conditions imposed for brevity of proofs.

4. Asymptotic results

The result in Proposition 4.1 ensures the uniform consistency on a set $\mathcal{S} \in \mathcal{F}$, for both kernel or nearest neighbor with kernel estimates. Theorem 4.2 deals with the asymptotic normality of the proposed estimator.

Proposition 4.1 Assume that assumptions (A5), (A7) and (A8) holds. Moreover, assume that (A9) hold for kernel weights, and that (A10) holds for nearest neighbor with kernel weights. Then, for any set $\mathcal{S}$, we have that

(a) Under (A1) and (A8-ii), we have that $\sup_{x\in\mathcal{S}} |\widehat{\theta}(x) - \theta(x)| \xrightarrow{\text{a.s.}} 0$.

(b) If $F(y|X = x)$ has a unique median at $\theta(x)$, then we reach $\sup_{x\in\mathcal{S}} |\widehat{m}_{\mathrm{MED}}(x) - \theta(x)| \xrightarrow{\text{a.s.}} 0$.

Theorem 4.2 Assume that (A1)-(A6), and (A8-ii) hold. Then, as $\widehat{\theta}(x) \xrightarrow{\text{p}} \theta(x)$ and $\widehat{s}(x) \xrightarrow{\text{p}} \sigma(x)$, we have that

$$\left(\frac{n\phi(x,h)}{\sigma^2(x,\theta(x))}\right)^{1/2} \left(\widehat{\theta}(x) - \theta(x) - B_n(x)\right) \xrightarrow{\text{d}} \mathrm{N}(0,1) \quad \text{as} \quad n \to \infty,$$

where $\xrightarrow{\text{d}}$ meaning convergence in distribution, $B_n(x) = h\Phi'(0,\theta(x))\beta_0/\beta_1 + o(h)$ and $\sigma^2(x,\theta(x)) = \beta_2\lambda_2(x,\theta(x),\sigma)/(\beta_1^2(\Gamma_1(x,\theta(x),\sigma))^2)$, with $\beta_0 = -\int_0^1 (sK(s))'\beta_x(s)\mathrm{d}s$, $\beta_j = -\int_0^1 (K^j)'(s)\beta_x(s)\mathrm{d}s$, for $j = 1,2$, $\Gamma_1(x,\theta(x),\sigma) = \partial\Psi(x,\theta(x),\sigma)/\partial t$, and $\mathcal{A} = \{z \in \mathcal{F}, \lambda_2(z,\theta(z),\sigma)\Gamma_1(z,\theta(z),\sigma) \neq 0\}$.

In order to remove the bias term B_n, we need an additional condition on the bandwidth parameter h.

Corollary 4.3 Under the assumptions of Theorem 4.2, and if the bandwidth parameter h satisfies $nh^2\phi(x,h) \to 0$ as $n \to \infty$, then

$$\left(\frac{n\phi(x,h)}{\sigma^2(x,\theta(x))}\right)^{1/2}\left(\widehat{\theta}(x)-\theta(x)\right) \xrightarrow{\mathrm{d}} \mathrm{N}(0,1) \quad \text{as} \quad n\to\infty.$$

5. Simulation study

Next, we show the efficiency of the proposed estimator in terms of consistency.

The first direct use of Theorem 4.2 is to predict a functional time series process. Let $(Z_t)_{t\in[0,b[}$ be a continuous-time real-valued random process. From the process Z_t, we may construct N functional random variables $(X_i)_{i=1,\ldots,N}$ defined by $X_i(t) = Z_{N^{-1}((i-1)b+c)}$, $\forall t \in [0,b]$. The predictor estimator of Y is defined by $\widehat{Y} = \widehat{\theta}(X_N)$. Then, by applying the above results, we obtain the following corollary.

Corollary 5.1 Under the assumptions of Corollary 4.3, we have

$$\left(\frac{N\phi(x,h_N)}{\sigma^2(X_N,\theta(X_N))}\right)^{1/2}\left(\widehat{\theta}(X_N)-\theta(X_N)\right) \xrightarrow{\mathrm{d}} \mathrm{N}(0,1) \quad \text{as} \quad N\to\infty.$$

The second direct result obtained in Theorem 4.2 is to build the conditional confidence curve. Note that an important application of the asymptotic normality result is the construction of confidence intervals for the true value of $\theta(x)$ given that $X=x$. However, the latter requires an estimation of the bias $B_n(x)$ term and of the standard deviation $\sigma(x,\theta(x))$. For the sake of shortness, we neglect the bias term and we estimate $\sigma(x,\theta(x))$ by plug-in method as follows. Effectively, if ψ_x is of class $\mathcal{C}^1$, with respect to the second component, the quantities $\lambda_2(x,\theta(x),s)$ and $\Gamma_1(x,\theta(x),s)$ can be estimated by

$$\widehat{\lambda}_2\left(x,\widehat{\theta}(x),\widehat{s}\right) = \frac{\sum_{i=1}^n K\left(\frac{d(x,X_i)}{h}\right)\psi_x^2\left(\frac{Y_i-\widehat{\theta}(x)}{\widehat{s}}\right)}{\sum_{i=1}^n K\left(\frac{d(x,X_i)}{h}\right)},$$

$$\widehat{\Gamma}_1\left(x,\widehat{\theta}(x),\widehat{s}\right) = \frac{\sum_{i=1}^n K\left(\frac{d(x,X_i)}{h}\right)\frac{\partial}{\partial t}\psi_x\left(\frac{Y_i-\widehat{\theta}(x)}{\widehat{s}}\right)}{\sum_{i=1}^n K\left(\frac{d(x,X_i)}{h}\right)}.$$

We estimate β_1 and β_2 by

$$\widehat{\beta}_1 = \frac{1}{n\phi(x,h)}\sum_{i=1}^n K\left(\frac{d(x,X_i)}{h}\right), \quad \widehat{\beta}_2 = \frac{1}{n\phi(x,h)}\sum_{i=1}^n K^2\left(\frac{d(x,X_i)}{h}\right).$$

It follows that $\widehat{\sigma}(x,\widehat{\theta}(x)) = (\widehat{\beta}_2\widehat{\lambda}_2(x,\widehat{\theta}(x),\widehat{s})/(\widehat{\beta}_1)^2\widehat{\Gamma}_1^2(x,\widehat{\theta}(x),\widehat{s}))^{1/2}$.

Then, from the asymptotic normality result in Section 4, we have

$$\Lambda_n = \left(\frac{n\phi(x,h)}{\sigma^2(x,\theta(x))}\right)^{1/2}\left(\widehat{\theta}(x)-\theta(x)\right) \xrightarrow{\mathrm{d}} \mathrm{N}(0,1) \quad \text{as} \quad n\to\infty.$$

Therefore, we get an approximate $(1-\zeta)100\%$ confidence interval for $\theta(x)$ stated as

$$\widehat{\theta}(x) \pm t_{1-\zeta/2} \times \left(\frac{\widehat{\sigma}_n^2 \left(x, \widehat{\theta}(x) \right)}{n\phi\,(x,h)} \right)^{1/2},$$

where $t_{1-\zeta/2}$ denotes the $(1-\zeta/2)100$th standard normal quantile.

To verify the theoretical results, it is possible to visualize the data histogram and then compare its shape to the normal density. The histogram of Λ_n is almost symmetric around zero and to well-shaped like the standard normal density. To do that, we consider the functional nonparametric model given by

$$Y_i = r\,(X_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

where the ϵ_is are generated independently according to a normal distribution with mean 0.

Now, we describe how our functional ergodic data are generated. Firstly, we use an `R` routine named `simul.far` of the `far` package to generate the functional explanatory variables $(X_i)_{i=1,\ldots,n}$. This routine simulates a functional autoregressive process white Wiener noise.

For this simulation experiments, we consider sinusoidal basis, with five functional axis, of the continuous functions from $[0,1]$ to R. Recall that, as it is shown in Laïb and Louani (2011), this kind of process satisfies the ergodicity condition. The curves X_is are discretized in the same grid composed by 100 points and are plotted in Figure 1.

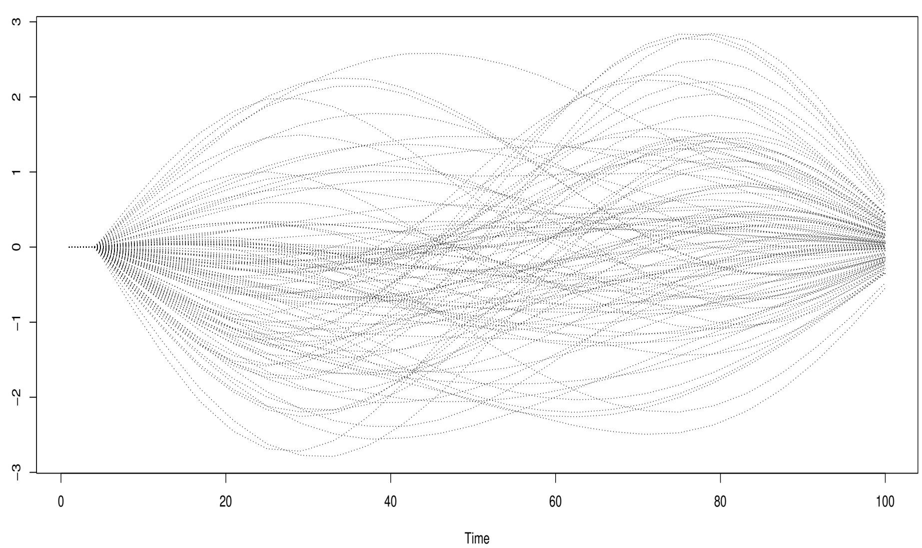


Figure 1. A sample of 100 curves, for $d_\rho = (0.45, 0.90, 0.34, 0.45)$

Secondly, the scalar response Y_i is computed by considering the operator defined as

$$r(x) = 5 \int_0^1 \exp\left\{ x\,(t) \right\} \mathrm{d}t.$$

We compare our estimator (robust equivariant regression –RER–) $\widehat{\theta}(x)$ with the kernel robust regression (KRR) $\tilde{\theta}(x)$ introduced by (Azzedine et al., 2008) and the functional kernel regression (FKR) (Ferraty and Vieu, 2006), where $\widehat{\theta}(x)$, $\tilde{\theta}(x)$ and $\widehat{m}(x)$ are defined as $\widehat{\theta}(x)$ is the zero with respect to a of

$$\frac{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)\psi_x\left(\frac{Y_i-a}{\widehat{s}(x)}\right)}{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)} = 0,$$

and $\widetilde{\theta}(x)$ is the zero with respect to a of

$$\frac{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)\psi_x\left(Y_i-a\right)}{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)} = 0, \quad \widehat{m}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{d(x,X_i)}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)}.$$

The efficiency of the predictors is evaluated by the empirical mean square errord (MSEs) expressed as

$$\mathrm{MSE}_{\widehat{\theta}} = n^{-1}\sum_{i=1}^{n}\left(\theta(X_i)-\widehat{\theta}(X_i)\right)^2,$$

$$\mathrm{MSE}_{\tilde{\theta}} = n^{-1}\sum_{i=1}^{n}\left(\theta(X_i)-\tilde{\theta}(X_i)\right)^2,$$

$$\mathrm{MSE}_{\widehat{m}} = n^{-1}\sum_{i=1}^{n}\left(\theta(X_i)-\widehat{m}(X_i)\right)^2.$$

Through this simulation study, we chose the quadratic kernel K defined as $K(u) = (3/4)(1-u^2)\mathbb{1}_{[0,1]}(u)$. The choice of bandwidth parameter h is a crucial question in nonparametric estimation, we propose to choose the optimal bandwidth by using cross-validation (CV) procedure. We adopt the selection rule proposed by Ferraty and Vieu (2006) and given by $h = \arg\min_h \mathrm{CV}(h)$, where $\mathrm{CV}(h) = \sum_{i=1}^{n}(Y_i - \widehat{\theta}^{-i}(X_i))^2$, with $\widehat{\theta}^{-i}(\cdot)$ being the leave-one-out CV –values of the estimator $\widehat{\theta}(\cdot)$ calculate at X_i–; see Ferraty and Vieu (2006) for more details.

We use the semi-metric given by he first derivative of sample curves stated as

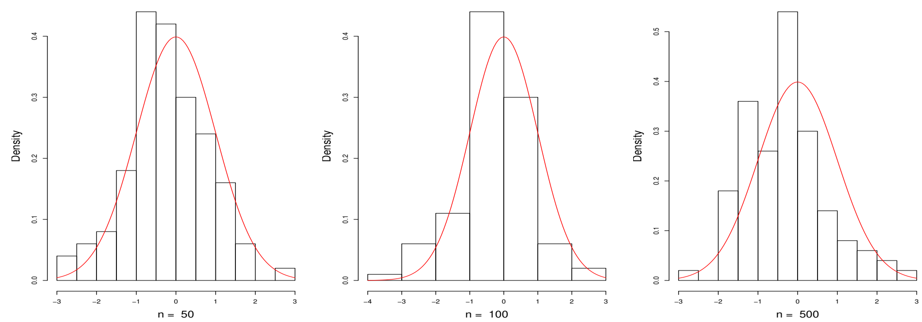
$$d\left(X_i,X_j\right) = \sqrt{\int\left(X_i'(t)-X_j'(t)\right)^2 \mathrm{d}t}.$$

For this comparison study, we treat three estimators in the same conditions.

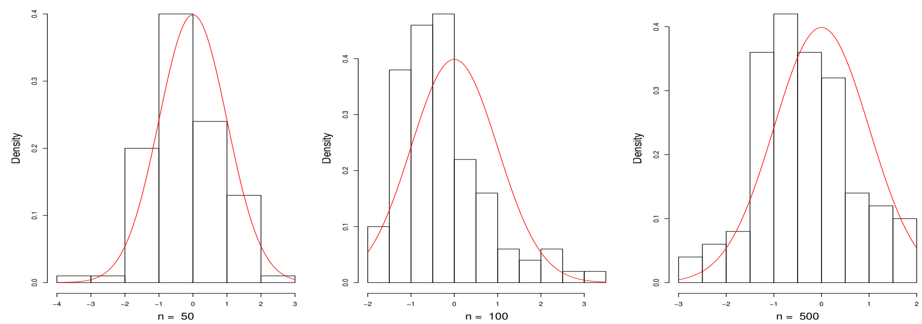
The first illustration concerns the asymptotic normality of $\widehat{\theta}\,(x)$. In order to conduct a Monte Carlo study of the asymptotic normality, we fix one curve, X_0 say, from the previous data. Then, we draw 100 independent replication with samples of size $n = 50, 100, 500$ of the same data and we compute, for each sample a quantity established as

$$\widehat{\Lambda}_n = \left(\frac{\left(\widehat{\beta}_1\right)^2 \widehat{\Gamma}_1^2\left(X_0,\widehat{\theta}\left(X_0\right),\widehat{s}\right)}{\widehat{\beta}_2\widehat{\lambda}_2\left(X_0,\widehat{\theta}\left(X_0\right),\widehat{s}\right)}\right)^{1/2}\left(\widehat{\theta}\left(X_0\right)-\theta\left(X_0\right)\right).$$

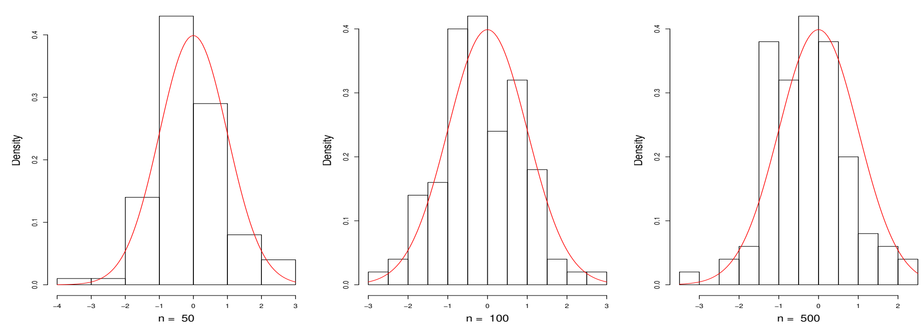
We point out that the functions $\phi(x,h)$ did not intervene in the computation of the normalized deviation by simplification. Thus, the simulation results indicate that $\widehat{\Lambda}_n$ obeys the standard normal law when n is large; see Figure 2 (a)-(c).



(a) $\mathrm{d}_\rho = \mathrm{diag}(0.225, 0.45, 0.17, 0.225)$



(b) $\mathrm{d}_\rho = \mathrm{diag}(0.45, 0.90, 0.34, 0.45)$



(c) $\mathrm{d}_\rho = \mathrm{diag}(0.90, 1.80, 0.68, 0.90)$

Figure 2. Histograms and density curves.

Now, in order to explore the two structural axes of our study, such as the correlation of data and the robustness of the estimate, we compare the performance of our estimator with various values of n and various parameters of the functional autoregressive X_i. Typically, we consider three values of $n = 50, 100, 500$, and three matrix $\mathrm{d}_\rho = \mathrm{diag}(0.225, 0.45, 0.17, 0.225)$, $\mathrm{d}_\rho = \mathrm{diag}(0.45, 0.90, 0.34, 0.45)$ and $\mathrm{d}_\rho = \mathrm{diag}(0.90, 1.80, 0.68, 0.90)$. We emphasize that the results of our simulation study are evaluated over 100 independent replications. The most significant results are gathered in Figure 2 (a)-(c). Note the performance of the estimator is closely related to the degree of correlation expressed by $\|\rho\|$. In sense that the histogram density converge significantly with respect to the value of $\|\rho\|$.

The second result concern the confidence intervals presented in Figures 3 and 4, where three curves corresponding to the predicted interval (green and blue curves) the estimated value (red curve) are drawn. Note that Figure 4 shows the good behavior of our functional forecasting procedure for the robust method in presence of outliers.

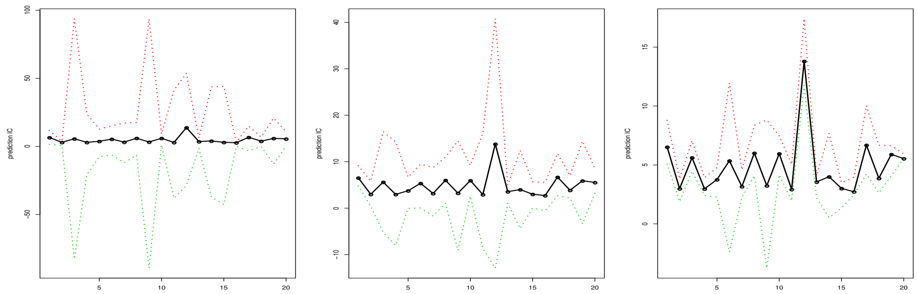


Figure 3. Extremities of the predicted values versus the true values and the confidence bands for the FKR, KRR and RER models respectively (simulation data without outliers).

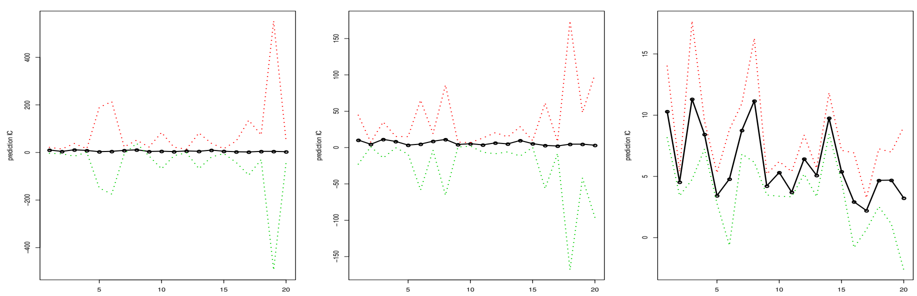


Figure 4. Extremities of the predicted values versus the true values and the confidence bands for the FKR, KRR and RER models respectively (simulation data with 7% of outliers).

Table 1. Comparison between the both methods in the presence of outliers.

Number of the perturbed observations by M	$\mathrm{MSE}_{\widehat{\theta}}$	$\mathrm{MSE}_{\widetilde{\theta}}$	$\mathrm{MSE}_{\widehat{m}}$
2	2.3118	2.8112	36.786
14	5.5197	21.8335	2513.116
100	50.716	220.506	331002.4

6. A real data application

Air pollution is one of the most influential factors in human health. Many different chemical substances contribute to the air quality. These substances come from a variety of sources. On the one hand, there are natural sources such as forest fires, volcanic eruptions, wind erosion, pollen dispersal, evaporation of organic compounds, and natural radioactivity. Furthermore, on the other hand, human industrial activity represents the artificial air pollution sources. Ozone (O_3), nitric oxide (NO) and nitrogen dioxide (NO_2) are among the most important contaminants in urban areas, as they have been associated with adverse effects on human health and the natural environment.

We apply the theoretical results obtained in the previous sections to real data. More specifically, in functional prediction context, we examine the performance of the proposed estimator by the robust equivariant approach $\widehat{\theta}(x)$.

In this real data example, we are interested in the prediction of the future O_3, NO and NO_2 concentrations given the curve of it is previous days. For this purpose application, we consider hourly concentrations of the 3 air pollution gases for the year 2018 $(Z_t)_{t\in[0,8760[}$. We consider the data collected from the Leicester University monitoring site in the UK. These observations are available on the following website: https://uk-air.defra.gov.uk. Table 2 gives descriptive statistics of these.

Table 2. Descriptive statistics of the air pollution data.

	O_3	NO	NO_2
Minimum	0.00	0.000	0.00
1st quartile	25.35	1.288	12.26
Median	42.11	3.346	19.56
Mean	42.52	7.027	23.07
3rd quartile	57.73	7.347	30.75
Maximum	149.58	190.765	115.43

We assume that the observations are linked by the model defined as

$$Y_i = r\left(X_i\right) + \epsilon_i, \quad i = 1, \ldots, n-1,$$

where $n = 365$, the functional random variables $\left(X_i\right)_{i=1,\ldots,n}$ defined by $X_i(t) = Z_{(24(i-1)+t)}$, $\forall t \in [0, 24[$, and the scalar response variable Y is defined by $Y_i = \left(Z_{24i+s}\right)_{i=1,\ldots,n-1}$ for a fixed $s \in [0, 24[$. Indeed, Z_t designs the O_3, NO and NO_2 concentrations for 8760 hours between January $01^{st}, 2018$ and 31 December 2018. We cut this functional time series in $n-1 = 364$ pieces X_i of 24 hours (one day). These functionals variables X_i are presented in Figure 5.

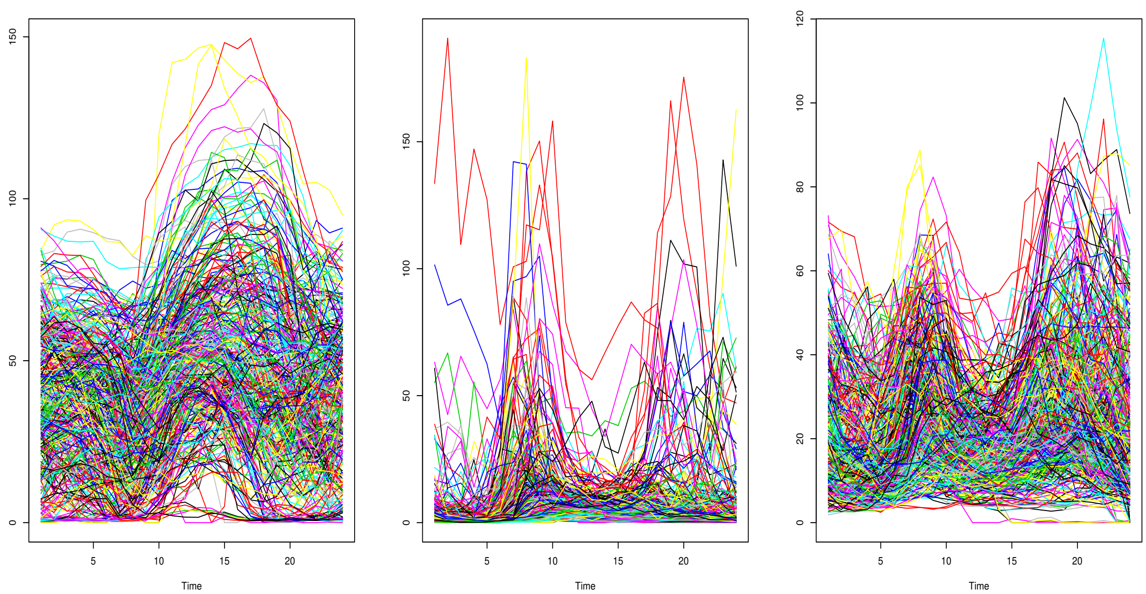


Figure 5. Hourly O_3 (left), NO (center) and NO_2 (right) concentrations of the year 2018.

We want to compare our proposed estimator $\widehat{\theta}(x)$ (RER) with the robust one $\widetilde{\theta}(x)$ (KRR), and the (FKR) $\widehat{m}\left(x\right)$. The kernel K is chosen to be quadratic defined as

$$K(u) = \frac{3}{4}\left(1 - u^2\right) \mathbb{1}_{[0,1]}(u).$$

The choice of bandwidth parameter h is a crucial question in nonparametric estimation. We propose to choose the optimal bandwidth by using the CV procedure. As mentioned, we adopt the selection rule proposed by Ferraty and Vieu (2006). Regarding the shape of the curves X_i, we suggest to use standard functional principal components analysis semi-metrics

(Ferraty and Vieu, 2006), and we adapt it to the data set under analysis obtaining

$$d_q\left(X_i, X_j\right) = \sqrt{\sum_{k=1}^{q}\left(\int \left[X_i(t) - X_j(t)\right] v_k(t)\mathrm{d}t\right)^2}.$$

Here, we take $q = 4$, and v_k is selected among the eigenfunctions of the empirical covariance operator defined as

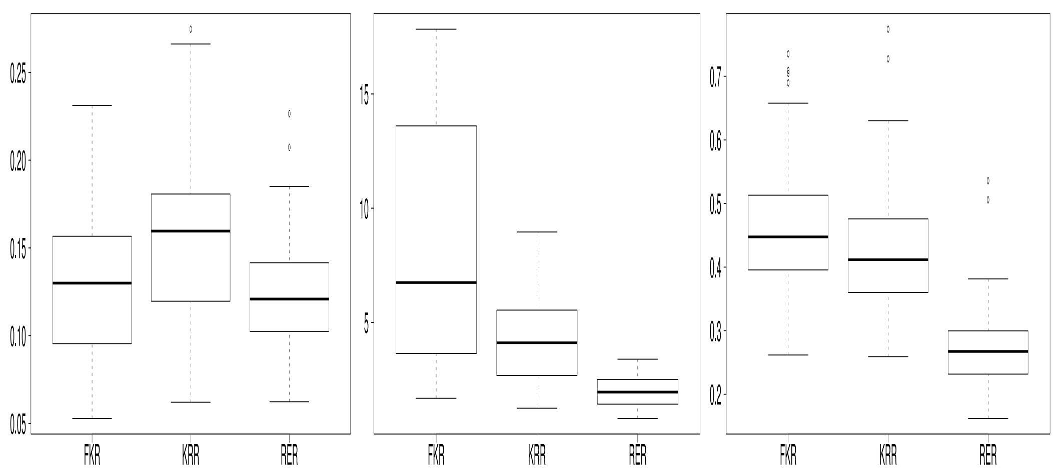
$$\Gamma_X^n(s,t) = \frac{1}{n}\sum_{i=1}^{n} X_i(s)X_i(t).$$

We randomly split our data set $(X_i, Y_i)_{i=1\ldots 364}$ into two subsets, that is, in (i) a learning sample $(T_i, X_i, Y_i)_{i\in I}$ (75% of the observations); and (ii) in a test sample $(X_i, Y_i)_{i\in I'}$, corresponding to a 25% of the observations. We use the relative mean square error RMSE as accuracy measure defined as

$$\mathrm{RMSE} = \frac{1}{\#(I')}\sum_{i\in I'}\left(\frac{Y_i - \widetilde{Y}_i}{Y_i}\right)^2,$$

where $\widetilde{Y}_i$ is the estimator for the three FKR, KRR and RER methods, and $\#(I')$ is the size of I'.

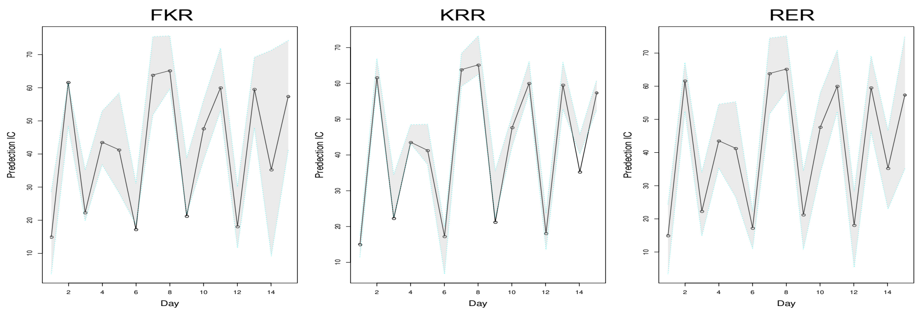
To further explore the performances of our models, we carry out $M = 100$ independent replications which allows us to compute 100 values for RMSE and to display their distribution by means of a scatter-plots. Figures 6 (a)-(c) shows the scatter-plots of the RMSE of the prediction values for the O_3, NO and NO_2, respectively.



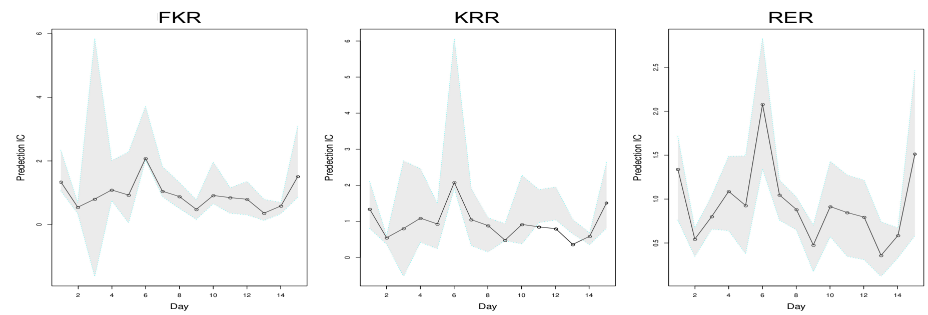
(a) O_3 (b) NO (c) NO_2

Figure 6. Comparison of the RMSE among the FKR, KRR and RER methods for the variable indicated.

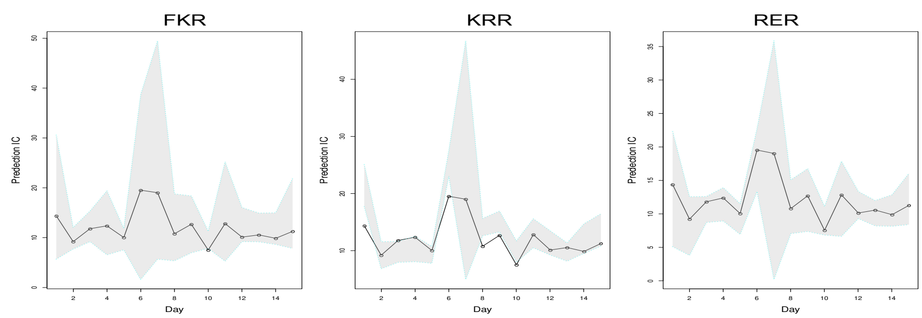
The obtained results of the scatter-plots of the RMSE proves that the Robust equivariant regression gives better results than the Classical and the robust methods. In addition, we give in Figure 7 (a)-(c) the 90% predictive intervals of the concentrations for the three gases of the last 15 values in the sample test by using the three modeles FKR, KRR and RER. The solid black curve the true values. the gray area represents the confidence zone between the dashed Blue curves which represents the lower and upper predicted values.



(a) O_3



(b) NO



(c) NO_2

Figure 7. Comparison of the 90% predictive intervals among the listed methods for the variable indicated.

7. Conclusion and future research

We have provided in this work a generalization of the results given in Boente and Vahnovan (2015) to the functional ergodic data. More precisely, we have proven the asymptotic normality of the robust regression function in the case of unknown scale parameter. These results were obtained under sufficient standard conditions that allowed us to explore different structural axes of the subject, such as the functional naturalness of the model and the data as well as the robustness of the regression function and the correlation of the observation.

Based on the results of this paper on robust regression with unknown scale parameter, we guess that most of the techniques using nonparametric functional kernel smothers could be efficiently extended. For instance, challenging open questions in this sense could concern as extensions to other forms of nonparametric predictors (like functional local linear ones, functional kNN ones, and many other ones). Extensions to other kinds of prediction models in which a preliminary kernel stage plays a crucial role. This would include many semiparametric regression models like functional single index models, and partial linear models, and many other ones. In addition, we see the possibility of extending our asymptotic result to other kinds of dependency data, more particularly the data associated positively (Azevedo and Oliveira, 2011).

Acknowledgements

The authors would like to thank the Editors-in-Chief and the anonymous reviewers for their valuable comments and suggestions which improved substantially the quality of this paper.

References

Attouch, M. Laksaci, A., and Ould Saïd E., 2009. Asymptotic distribution of robust estimator for functional nonparametric models. Communications in Statistics: Theory and Methods, 38, 1317–1335.

Attouch, M. Laksaci, A., and Ould Saïd E., 2010. Asymptotic normality of a robust estimator of the regression function for functional time series data. Journal of the Korean Statistical Society, 39, 489–500.

Azevedo, C. and Oliveira, P. E., 2011. On the kernel estimation of a multivariate distribution function under positive dependence. Chilean Journal of Statistics, 2, 99–113.

Azzedine, N., Laksaci, A., and Ould-Saïd, E., 2008. On the robust nonparametric regression estimation for functional regressor. Statistics and Probability Letters, 78, 3216–3221.

Benziadi, F., Gheriballah, A., and Laksaci, A., 2016a. Asymptotic normality of kernel estimator of ψ-regression functional ergodic data. New Trends in Mathematical Sciences, 1, 268–282.

Benziadi, F., Laksaci, A., and Tebboune, F., 2016b. Recursive kernel estimate of the conditional quantile for functional ergodic data. Communications in Statistics: Theory and Methods, 45, 3097–3113.

Boente, G. and Fraiman, R., 1989. Nonparametric regression estimation. J ournal of Multivariate Analysis, 29, 180–198.

Boente, G. and Fraiman, R., 1990. Asymptotic distribution of robust estimators for nonparametric models from mixing processes. Annals of Statistics, 18, 891–906.

Boente, G. and Vahnovan, A., 2015. Strong convergence of robust equivariant nonparametric functional regression estimators. Statistics and Probability Letters, 100, 1–11.

Boente, G. and Vahnovan, A., 2017. Robust estimators in semi-functional partial linear regression models. Journal of Multivariate Analysis, 154, 59–84.

Bosq, D., 2000. Linear Processes in Function Spaces. Springer, New York.

Chen, J. and Zhang, L., 2009. Asymptotic properties of nonparametric M-estimation for mixing functional data. Journal of Statistical Planning and Inference, 139, 533–546.

Collomb, G. and Härdle, W., 1986. Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. Stochastic Processes and their Applications, 23, 77–89.

Crambes, C., Delsol, L., and Laksaci, A., 2008. Robust nonparametric estimation for functional data. Journal of Nonparametric Statistics, 20, 573–598.

Ferraty. F. and Vieu. P., 2006. Nonparametric Functional Data Analysis: Theory and Practice. Springer, New York.

Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. 2010. Rate of uniform consistency for nonparametric estimates with functional variables. Journal of Statistical Planning and Inference, 140, 335–352.

P. Gaenssler, P., Strobel, J., and Stute, W., 1978. On central limit theorems for martingale triangular arrays. Acta Mathematica Academiae Scientiarum Hungarica, 31, 205–216.

Gheriballah, A. Laksaci, A., and Sekkal, S., 2013. Nonparametric M-regression for functional ergodic data. Statistics and Probability Letters, 83, 902–908.

Huber, P.J., 1964. Robust estimation of a location parameter. Annals of Mathematical Statistics, 35, 73–101.

Laïb, N. and Louani, D., 2010. Nonparametric kernel regression estimation for functional stationary ergodic data: Asymptotic properties. Journal of Multivariate Analysis, 101, 2266–2281.

Laïb, N. and Louani D., 2011. Rates of strong consistencies of the regression function estimator for functional stationary ergodic data. Journal of Statistical Planning and Inference, 141, 359–372.

Laïb, N. and Ould Saïd, E., 2000. A robust nonparametric estimation of the autoregression function under ergodic hypothesis. Canadian Journal of Statistics, 28, 817–828.

Ramsay, J.O. and Silverman, B.W., 2002. Applied Functional Data Analysis: Methods and Case Studies. Springer, New York.

Appendix

Proof of Proposition 4.1 In order to prove Proposition 4.1, we begin by fixing some notation. We prove that, for any measurable $A \subset R$, $\widehat{\phi}_A(x) = \widehat{r}_A(x)/\widehat{p}(x)$, where

$$\widehat{r}_A(x) = \sum_{i=1}^{n} W_{i,n}(x) I_A(y_i), \widehat{p}(x) = \sum_{i=1}^{n} W_{i,n}, W_{i,n}(x) = \frac{K\left(\frac{d(x_i,x)}{h_n}\right)}{\sum_{j=1}^{n} K\left(\frac{d(x_j,x)}{h_n}\right)}, \tag{1}$$

denote the kernel weights. Next, we prove (a) and (b). Note that:

(a) Arguing as in Theorem 3.3 in Boente and Fraiman (1990), we only need to prove that

$$\sup_{x\in\mathcal{S}} \sup_{y\in R} \left| \widehat{F}(y|X=x) - F(y|X=x) \right| \overset{\text{a.s.}}{\to} 0.$$

Theorems 3.1 or 3.2 from Boente and Fraiman (1990) entail that

$$\sup_{x\in\mathcal{S}} \sup_{y\in R} |\widehat{r}(y,x) - r(y,x)| \overset{\text{a.s.}}{\to} 0, \quad \sup_{x\in\mathcal{S}} |\widehat{p}(x) - p(x)| \overset{\text{a.s.}}{\to} 0, \tag{2}$$

where $r(y,x) = \phi_{(-\infty,y]}(x) = p(x) F(y|X=x)$ and $\widehat{r}(y,x) = \widehat{\phi}_{(-\infty,y]}(x)$, with $\widehat{\phi}_{(-\infty,y]}(x)$ and $\widehat{p}(x)$ being defined in Equation (1).

Note that Equation (2) can be derived for kernel weights using Proposition 2 in Collomb and Hardle (1986). Now, Equation (2) follows from using (A7) and the inequality

$$\sup_{x\in\mathcal{S}} \sup_{y\in R} \left| \widehat{F}(y|X=x) - F(y|X=x) \right| \le \frac{\sup_{x\in\mathcal{S}} \sup_{y\in R} |\widehat{r}(y,x) - r(y,x)| + \sup_{x\in\mathcal{S}} |\widehat{p}(x) - p(x)|}{A_p \widehat{A}_p},$$

where $A_p = \inf_{x\in\mathcal{S}} p(x)$ and $\widehat{A}_p = \inf_{x\in\mathcal{S}} \widehat{p}(x)$.

(b) The equicontinuity condition given in (A8), and the uniqueness of the conditional median, imply that $\theta(x)$ is a continuous function of x. Thus, for any fixed $a \in$ R, the function $h_a(x) = F(a + \theta(x)|X=x)$ also is continuous with respect to x.

Given $\epsilon > 0$, let $0 < \delta < \epsilon$, such that

$$|u - v| < \delta \Longrightarrow \sup_{x\in\mathcal{S}} (|F(U|X=x) - F(v|X=x)|) < \frac{\epsilon}{2}. \tag{3}$$

Then, from the uniqueness of the conditional median and Equation (3), we get that

$$\frac{1}{2} < F(\theta(x) + \delta|X=x) < \frac{1}{2} + \frac{\epsilon}{2}, \tag{4}$$

$$\frac{1}{2} - \frac{\epsilon}{2} < F(\theta(x) - \delta|X=x) < \frac{1}{2}. \tag{5}$$

Consider $\iota(\delta) = \inf_{x\in\mathcal{S}} F(\theta(x) + \delta|X=x)$ and $\nu(\delta) = \sup_{x\in\mathcal{S}} F(\theta(x) - \delta|X=x)$. The continuity of $h_\delta(x)$ and $h_{-\delta}(x)$ together with Equations (4) and (5), entail that $\nu(\delta) < 1/2 < \iota(\delta)$, and so $\eta = \min(\iota(\delta) - 1/2, 1/2 - \nu(\delta)) > 0$. If Equation (2) holds, $\mathrm{P}(\mathcal{N}) = 0$, and $\sup_{x\in\mathcal{S}} \sup_{y\in R} |\widehat{F}(y|X=x) - F(y|X=x)| \to 0$, then, for n large enough, we have that $\sup_{x\in\mathcal{S}} \sup_{y\in R} |\widehat{F}(y|X=x) - F(y|X=x)| < \min(\eta/2, \epsilon/2) = \epsilon_1$. Then, for $x \in \mathcal{S}$, we get

$$F(\theta(x) + \delta|X=x) - \epsilon_1 < \widehat{F}(\theta(x) + \delta|X=x) < F(\theta(x) + \delta|X=x) + \epsilon_1,$$

$$F(\theta(x) - \delta|X=x) - \epsilon_1 < \widehat{F}(\theta(x) - \delta|X=x) < F(\theta(x) - \delta|X=x) + \epsilon_1,$$

which entails that

$$\frac{1}{2} < \widehat{F}(\theta(x) + \delta|X=x) < \frac{1}{2} + \epsilon, \quad \frac{1}{2} - \epsilon < \widehat{F}(\theta(x) - \delta|X=x) < \frac{1}{2},$$

and hence, $\sup_{x\in\mathcal{S}} |\widehat{m}_{\text{MED}}(x) - \theta(x)| \le \delta < \epsilon$, which concludes the proof. □

Proof of Theorem 4.2 and Corollary 4.3 We give the proof for the case of increasing ψ_x, with the decreasing case being obtained by considering $-\psi_x$. Thus, we define, for all $u \in \mathrm{R}$, $z = \theta(x) - B_n(x) + u\left[n\phi(x,h)\right]^{-1/2}\sigma(x,\theta(x))$. Notice that

$$\mathrm{P}\left(\left(\frac{n\phi(x,h)}{\sigma^2(x,\theta(x))}\right)^{1/2}\left(\widehat{\theta}(x) - \theta(x) + B_n(x)\right) < u\right) = \mathrm{P}\left(\widehat{\theta}(x) < \theta(x) - B_n(x) + u\left[n\phi(x,h)\right]^{-1/2}\sigma(x,\theta(x))\right)$$
$$= \mathrm{P}\left(0 < \widehat{\Psi}(x,z,\widehat{s})\right).$$

In addition, we have that

$$\widehat{\Psi}(x,t,\widehat{s}) = B_n(x,t,\widehat{s}) + \frac{R_n(x,t,\widehat{s})}{\widehat{\Psi}_D(x)} + \frac{Q_n(x,t,\widehat{s})}{\widehat{\Psi}_D(x)},$$

where

$$Q_n(x,t,\widehat{s}) = \left(\widehat{\Psi}_N(x,t,\widehat{s}) - \bar{\Psi}_N(x,t,\widehat{s})\right) - \Psi(x,t,\widehat{s})\left(\widehat{\Psi}_D(x) - \bar{\Psi}_D(x)\right),$$
$$R_n(x,t,\widehat{s}) = -\left(\frac{\bar{\Psi}_N(x,t,\widehat{s})}{\bar{\Psi}_D(x)} - \Psi(x,t,\widehat{s})\right)\left(\widehat{\Psi}_N(x,t,\widehat{s}) - \bar{\Psi}_N(x,t,\widehat{s})\right),$$
$$B_n(x,t,\widehat{s}) = \frac{\bar{\Psi}_N(x,t,\widehat{s})}{\bar{\Psi}_D(x)},$$

with

$$\widehat{\Psi}_N(x,a,\widehat{s}) = \frac{1}{n\mathrm{E}\left[K\left(h^{-1}d(x,X_1)\right)\right]}\sum_{i=1}^{n} K\left(h^{-1}d(x,X_i)\right)\psi_x\left(\frac{Y_i - a}{\widehat{s}}\right),$$
$$\bar{\Psi}_N(x,a,\widehat{s}) = \frac{1}{n\mathrm{E}\left[K\left(h^{-1}d(x,X_1)\right)\right]}\sum_{i=1}^{n} \mathrm{E}\left[K\left(h^{-1}d(x,X_i)\right)\psi_x\left(\frac{Y_i - a}{\widehat{s}}\right)/\mathcal{F}_{i-1}\right],$$
$$\widehat{\Psi}_D(x) = \frac{1}{n\mathrm{E}\left[K\left(h^{-1}d(x,X_1)\right)\right]}\sum_{i=1}^{n} K\left(h^{-1}d(x,X_i)\right),$$
$$\bar{\Psi}_D(x) = \frac{1}{n\mathrm{E}\left[K\left(h^{-1}d(x,X_1)\right)\right]}\sum_{i=1}^{n} \mathrm{E}\left[K\left(h^{-1}d(x,X_i)\right)|\mathcal{F}_{i-1}\right].$$

Then, it follows that

$$\mathrm{P}\left(\left(\frac{n\phi(x,h)}{\sigma^2(x,\theta_x)}\right)^{1/2}\left(\widehat{\theta}(x) - \theta(x) + B_n(x)\right) < u\right) =$$
$$\mathrm{P}\left(-\widehat{\Psi}_D(x)B_n(x,z,\widehat{s}) - R_n(x,z,\widehat{s}) < Q_n(x,z,\widehat{s})\right).$$

Therefore, our main result is a consequence of the following intermediate results. □

Lemma 7.1 Under the assumptions of Theorem 4.2, we have, for any $x \in \mathcal{A}$,

$$\left(\frac{n\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta(x),\widehat{s})}\right)^{1/2} Q_n(x,z,\widehat{s}) \xrightarrow{\mathrm{d}} \mathrm{N}(0,1), \quad \text{as} \quad n \to \infty.$$

Proof of Lemma 7.1 For all $i = 1,\ldots,n$, we denote by $K_i(x) = K\left(h^{-1}d(x,X_i)\right)$,

$$\eta_{ni} = \left(\frac{\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta(x),\widehat{s})}\right)^{1/2}\left(\psi_x\left(\frac{Y_i - z}{\widehat{s}}\right) - \Psi(x,z,\widehat{s})\right)\frac{K_i(x)}{\mathrm{E}\left[K_1(x)\right]},$$

and we define $\zeta_{ni} = \eta_{ni} - \mathrm{E}\left[\eta_{ni}|\mathcal{F}_{i-1}\right]$. Then, we obtain

$$\left(\frac{n\phi\left(x,h\right)\beta_1^2}{\beta_2\lambda_2\left(x,\theta(x),\widehat{s}\right)}\right)^{1/2} Q_n\left(x,z,\widehat{s}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\zeta_{ni}.$$

Since ζ_{ni} is a triangular array of martingale differences according the $\sigma-$field $\left(\mathcal{F}_{i-1}\right)$, we can apply the Central Limit Theorem based on the unconditional Lindeberg condition (Gaenssler et al., 1978). More precisely, we must verify conditions:

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\zeta_{ni}^2|\mathcal{F}_{i-1}\right] \xrightarrow{\mathrm{P}} 1, \tag{6}$$

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\zeta_{ni}^2 I_{\zeta_{ni}^2>\epsilon n}\right] \xrightarrow{\mathrm{P}} 0, \forall\epsilon>0, \tag{7}$$

We begin by proving Equation (6). In order to do that, we write

$$\mathrm{E}\left[\zeta_{ni}^2|\mathcal{F}_{i-1}\right] = \mathrm{E}\left[\eta_{ni}^2|\mathcal{F}_{i-1}\right] - \mathrm{E}^2\left[\eta_{ni}|\mathcal{F}_{i-1}\right].$$

Therefore, it suffices to prove that

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}^2\left[\eta_{ni}|\mathcal{F}_{i-1}\right] \xrightarrow{\mathrm{P}} 0,$$

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\eta_{ni}^2|\mathcal{F}_{i-1}\right] \xrightarrow{\mathrm{P}} 1. \tag{8}$$

For the first convergence, we have

$$\begin{aligned}
\left|\mathrm{E}\left[\eta_{ni}|\mathcal{F}_{i-1}\right]\right| &= \frac{1}{\mathrm{E}K_1\left(x\right)}\left(\frac{\phi\left(x,h\right)\beta_1^2}{\beta_2\lambda_2\left(x,\theta(x),\widehat{s}\right)}\right)^{1/2}\left|\mathrm{E}\left[\left(\Psi\left(X_i,t,\widehat{s}\right)-\Psi\left(x,t,\widehat{s}\right)K_i\left(x\right)\right)|\mathcal{F}_{i-1}\right]\right| \\
&\le \frac{1}{\mathrm{E}\left[K_1\left(x\right)\right]}\left(\frac{\phi\left(x,h\right)\beta_1^2}{\beta_2\lambda_2\left(x,\theta(x),\widehat{s}\right)}\right)^{1/2}\sup_{u\in B(x,h)}\left|\Psi\left(u,t,\widehat{s}\right)-\Psi\left(x,t,\widehat{s}\right)\right|\mathrm{E}\left[K_i\left(x\right)|\mathcal{F}_{i-1}\right].
\end{aligned}$$

Obviously, under (A2) and (A5), we have $C\phi_i\left(x,h\right) \le \mathrm{E}\left[K_i|\mathcal{F}_{i-1}\right] \le C^{'}\phi_i\left(x,h\right)$ and $C\phi\left(x,h\right) \le \mathrm{E}\left[\Delta_i\left(x\right)\right] \le C^{'}\phi\left(x,h\right)$. In addition, condition (A3-ii) implies that

$$\sup_{u\in B(x,h)}\left|\Psi\left(u,t,\widehat{s}\right)-\Psi\left(x,t,\widehat{s}\right)\right| = o\left(1\right).$$

Combining the lasts three results, we obtain

$$\begin{aligned}
\left(\left|\mathrm{E}\left[\eta_{ni}|\mathcal{F}_{i-1}\right]\right|\right)^2 &\le \sup_{u\in B(x,h)}\left|\Psi\left(u,t,\widehat{s}\right)-\Psi\left(x,t,\widehat{s}\right)\left(\frac{\beta_1^2}{\beta_2\lambda_2\left(x,\theta(x),\widehat{s}\right)}\right)\right|\frac{1}{\phi\left(x,h\right)}\phi_i^2\left(x,h\right) \\
&\le \sup_{u\in B(x,h)}\left|\Psi\left(u,t,\widehat{s}\right)-\Psi\left(x,t,\widehat{s}\right)\left(\frac{\beta_1^2}{\beta_2\lambda_2\left(x,\theta(x),\widehat{s}\right)}\right)\right|\frac{1}{\phi\left(x,h\right)}\phi_i\left(x,h\right).
\end{aligned}$$

Thus, by using the fact that

$$\frac{1}{n\phi\left(x,h\right)}\sum_{i=1}^{n}\phi_i\left(x,h\right) \xrightarrow{\mathrm{P}} 1,$$

we obtain

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{E}\left[\eta_{ni}|\mathcal{F}_{i-1}\right]\right)^2 &= \sup_{u\in B(x,h)}\left|\Psi\left(u,t,\widehat{s}\right)-\Psi\left(x,t,\widehat{s}\right)\right|\left(\frac{\beta_1^2}{\beta_2\lambda_2\left(x,\theta_x,\widehat{s}\right)}\right)\left(\frac{1}{n\phi\left(x,h\right)}\sum_{i=1}^{n}\phi_i\left(x,h\right)\right) \\
&= o_p\left(1\right).
\end{aligned}$$

Now, we analyze to the convergence in Equation (8). Consider

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\eta_{ni}^2|\mathcal{F}_{i-1}\right] &= \frac{1}{n\left(\mathrm{E}K_1(x)\right)^2}\left(\frac{\phi\left(x,h\right)\beta_1^2}{\beta_2\lambda_2\left(x,\theta(x),\widehat{s}\right)}\right)\\
&\quad\times\sum_{i=1}^{n}\mathrm{E}\left[\left(\psi_x\left(\frac{Y_i-z}{\widehat{s}}\right)-\Psi\left(x,z,\widehat{s}\right)\right)^2K_i^2\left(x\right)|\mathcal{F}_{i-1}\right]\\
&= \frac{1}{n\left(\mathrm{E}K_1\left(x\right)\right)^2}\left(\frac{\phi\left(x,h\right)\beta_1^2}{\beta_2\lambda_2\left(x,\theta_x,\widehat{s}\right)}\right)\left(\sum_{i=1}^{n}\mathrm{E}\left[\psi_x^2\left(\frac{Y_i-z}{\widehat{s}}\right)\Delta_i^2\left(x\right)|\mathcal{F}_{i-1}\right]\right)\\
&\quad-\frac{2\Psi\left(x,z,\widehat{s}\right)}{n\left(\mathrm{E}K_1\left(x\right)\right)^2}\left(\frac{\phi\left(x,h\right)\beta_1^2}{\beta_2\lambda_2\left(x,\theta_x,\widehat{s}\right)}\right)\sum_{i=1}^{n}\mathrm{E}\left[\psi_x\left(\frac{Y_i-z}{\widehat{s}}\right)\Delta_i^2\left(x\right)|\mathcal{F}_{i-1}\right]\\
&\quad+\frac{1}{n\left(\mathrm{E}K_1\left(x\right)\right)^2}\left(\frac{\phi\left(x,h\right)\beta_1^2}{\beta_2\lambda_2\left(x,\theta_x,\widehat{s}\right)}\right)\Psi^2\left(x,z,\widehat{s}\right)\sum_{i=1}^{n}\mathrm{E}\left[\Delta_i^2\left(x\right)|\mathcal{F}_{i-1}\right].
\end{aligned}$$

Let $D_1=\sum_{i=1}^{n}\mathrm{E}\left[\psi_x^2\left((Y_i-z)/\widehat{s}\right)\Delta_i^2\left(x\right)|\mathcal{F}_{i-1}\right]$, $D_2=\sum_{i=1}^{n}\mathrm{E}\left[\psi_x\left((Y_i-z)/\widehat{s}\right)\Delta_i^2\left(x\right)|\mathcal{F}_{i-1}\right]$, and $D_3=\sum_{i=1}^{n}\mathrm{E}\left[\Delta_i^2\left(x\right)|\mathcal{F}_{i-1}\right]$. Observe that

$$\begin{aligned}
D_1 &= \lambda_2\left(x,z,\widehat{s}\right)\sum_{i=1}^{n}\mathrm{E}\left[K_i^2\left(x\right)|\mathcal{F}_{i-1}\right]+\sum_{i=1}^{n}\left[\mathrm{E}\left[K_i^2\left(x\right)\mathrm{E}\left[\psi_x^2\left(\frac{Y_i-z}{\widehat{s}}\right)|\mathcal{B}_{i-1}\right]|\mathcal{F}_{i-1}\right]\right]\\
&\quad-\sum_{i=1}^{n}\lambda_2\left(x,z,\widehat{s}\right)\mathrm{E}\left[K_i^2\left(x\right)|\mathcal{F}_{i-1}\right]\\
&= \lambda_2\left(x,z,\widehat{s}\right)\sum_{i=1}^{n}\mathrm{E}\left[K_i^2\left(x\right)|\mathcal{F}_{i-1}\right]+\sum_{i=1}^{n}\mathrm{E}\left[K_i^2\left(x\right)\mathrm{E}\left[\psi_x^2\left(\frac{Y_i-z}{\widehat{s}}\right)|X_i\right]|\mathcal{F}_{i-1}\right]\\
&\quad-\sum_{i=1}^{n}\left[\lambda_2\left(x,z,\widehat{s}\right)\mathrm{E}\left[K_i^2\left(x\right)|\mathcal{F}_{i-1}\right]\right].
\end{aligned}$$

To evaluate the second term, we have

$$\begin{aligned}
&\frac{1}{n\mathrm{E}\left[K_1\left(x\right)\right]}\sum_{i=1}^{n}\left[\mathrm{E}\left[K_i^2\left(x\right)\mathrm{E}\left[\psi_x^2\left(\frac{Y_i-z}{\widehat{s}}\right)|X_i\right]|\mathcal{F}_{i-1}\right]-\lambda_2\left(x,z,\widehat{s}\right)\mathrm{E}\left[K_i^2\left(x\right)|\mathcal{F}_{i-1}\right]\right]\\
&\qquad\le\sup_{u\in B(x,h)}\left|\lambda_2\left(x,u,\widehat{s}\right)-\lambda_2\left(x,z,\widehat{s}\right)\right|\left(\frac{1}{n\phi\left(x,h\right)}\sum_{i=1}^{n}\mathrm{P}\left(X_i\in B\left(x,h\right)|\mathcal{F}_{i-1}\right)\right).
\end{aligned}$$

Moreover, we use the continuity of $\lambda_2\left(x,.,\widehat{s}\right)$ to write

$$\lambda_2\left(x,z,\widehat{s}\right)=\lambda_2\left(x,\theta(x),\widehat{s}\right)+o\left(1\right).$$

Thus, we get

$$\frac{1}{n\mathrm{E}\left[K_1\left(x\right)\right]}D_1=\lambda_2\left(x,\theta(x),\widehat{s}\right)\frac{1}{n\mathrm{E}\left[K_1\left(x\right)\right]}\sum_{i=1}^{n}\mathrm{E}\left[K_i^2\left(x\right)|\mathcal{F}_{i-1}\right]+o\left(1\right),$$

and similarly, we can obtain

$$\frac{1}{n\mathrm{E}\left[K_1\left(x\right)\right]}D_2=\Psi\left(x,\theta(x),\widehat{s}\right)\frac{1}{n\mathrm{E}\left[K_1\left(x\right)\right]}\sum_{i=1}^{n}\mathrm{E}\left[K_i^2\left(x\right)|\mathcal{F}_{i-1}\right]+o\left(1\right)=o\left(1\right).$$

Hence, we have

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\eta_{ni}^{2}|\mathcal{F}_{i-1}\right]=\frac{1}{n\left(\mathrm{E}\left[K_{1}\left(x\right)\right]\right)^{2}}\left(\frac{\phi\left(x,h\right)\beta_{1}^{2}}{\beta_{2}}\right)\sum_{i=1}^{n}\mathrm{E}\left[K_{i}^{2}\left(x\right)|\mathcal{F}_{i-1}\right]+o\left(1\right).$$

In what follows, we employ the same ideas used in Ferraty et al. (2010) to reach

$$\mathrm{E}\left[K_{i}^{2}\left(x\right)|\mathcal{F}_{i-1}\right]=K^{2}\left(1\right)\phi_{i}\left(x,h\right)-\int_{0}^{1}\left(K^{2}(u)\right)^{\prime}\phi_{i}\left(x,uh\right)\mathrm{d}u,$$

and $\mathrm{E}\left[K_{1}\left(x\right)\right]=K\left(1\right)\phi\left(x,h\right)-\int_{0}^{1}\left(K(u)\right)^{\prime}\phi\left(x,uh\right)\mathrm{d}u$. Then, it follows that

$$\begin{aligned}\frac{1}{n\phi\left(x,h\right)}\sum_{i=1}^{n}\mathrm{E}\left[K_{i}^{2}\left(x\right)|\mathcal{F}_{i-1}\right]&=\frac{K^{2}\left(1\right)}{n\phi\left(x,h\right)}\sum_{i=1}^{n}\phi_{i}\left(x,h\right)\\&\quad-\int_{0}^{1}\left(K^{2}(u)\right)^{\prime}\frac{\phi\left(x,uh\right)}{n\phi\left(x,h\right)\phi\left(x,uh\right)}\sum_{i=1}^{n}\phi_{i}\left(x,uh\right)\mathrm{d}u\\&=K^{2}\left(1\right)-\int_{0}^{1}\left(K^{2}(u)\right)^{\prime}\tau_{x}\left(u\right)\mathrm{d}u+o_{p}\left(1\right)=\beta_{2}+o_{p}\left(1\right),\end{aligned}$$

and

$$\frac{1}{n\phi\left(x,h\right)}\mathrm{E}\left[K_{1}\left(x\right)\right]=\beta_{1}+o\left(1\right).$$

We deduce that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\eta_{ni}^{2}|\mathcal{F}_{i-1}\right]=1,$$

which completes the proof of Equation (6).

Concerning Equation (7), we write

$$\zeta_{ni}^{2}I_{\zeta_{ni}^{2}>\varepsilon n}\leq\frac{\left|\zeta_{ni}\right|^{2+\delta}}{\sqrt{\left(\varepsilon n\right)^{\delta}}},\quad\forall\delta>0.$$

Observe that

$$\begin{aligned}\mathrm{E}\left[\zeta_{ni}^{2+\delta}\right]&=\mathrm{E}\left[\left|\eta_{ni}\left(x\right)-\mathrm{E}\left[\eta_{ni}\left(x\right)|\mathcal{F}_{i-1}\right]\right|^{2+\delta}\right]\\&\leq2^{1+\delta}\mathrm{E}\left[\left|\eta_{ni}\left(x\right)\right|^{2+\delta}\right]+2^{1+\delta}\left|\mathrm{E}\left[\mathrm{E}\left[\eta_{ni}|\mathcal{F}_{i-1}\right]^{2+\delta}\right]\right|.\end{aligned}$$

Using the Jensen inequality, we obtain $\mathrm{E}[\zeta_{ni}^{2+\delta}]\leq C\mathrm{E}[|\eta_{ni}(x)|^{2+\delta}]$. Thus, it remains to evaluate $\mathrm{E}[|\eta_{ni}(x)|^{2+\delta}]$. To that end, we once again use the C_r-inequality obtaining

$$\begin{aligned}\mathrm{E}\left[\left|\eta_{ni}\left(x\right)\right|^{2+k}\right]\leq C\left(\frac{\phi\left(x,h\right)\beta_{1}^{2}}{\beta_{2}\lambda_{2}\left(x,\theta(x),\widehat{s}\right)\mathrm{E}^{2}\left[K_{1}\left(x\right)\right]}\right)^{1+\delta/2}\mathrm{E}\left[K_{i}^{2+\delta}\left(x\right)\psi_{x}^{2+\delta}\left(\frac{Y_{i}-t}{\widehat{s}}\right)\right]\\+\quad\Psi^{2+\delta}\left(x,z,\widehat{s}\right)\mathrm{E}\left[K_{i}^{2+\delta}\left(x\right)\right].\end{aligned}$$

We condition on X_i and use the fact that

$$\mathrm{E}\left[\psi_{x}^{2+\delta}\left(\frac{Y_{i}-t}{\widehat{s}}\right)|X_{i}\right]<\infty,$$

to obtain

$$\mathrm{E}\left[\left|\eta_{ni}\left(x\right)\right|^{2+\delta}\right]\leq C\left(\frac{1}{\phi\left(x,h\right)}\right)^{1+\delta/2}\mathrm{E}\left(\left[K_{i}\left(x\right)\right]^{2+\delta}\right)\leq C\left(\frac{1}{\phi\left(x,h\right)}\right)^{\delta/2}.$$

Consequently, we get

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\zeta_{ni}^{2}I_{\zeta_{ni}^{2}>\varepsilon n}\right]\leq C\left(\frac{1}{n\phi\left(x,h\right)}\right)^{\delta/2}\to 0,$$

and the proof is complete. □

Lemma 7.2 (Laïb and Louani, 2010) Under assumptions (A1), (A2), (A5), and (A6), we have $\widehat{\Psi}_{D}\left(x\right)-1=o_{p}\left(1\right)$.

Lemma 7.3 Under assumptions (A1)-(A3), (A5), and (A6), we have

$$\left(\frac{n\phi\left(x,h\right)\beta_{1}^{2}}{\beta_{2}\lambda_{2}\left(x,\theta(x),\widehat{s}\right)}\right)^{1/2}B_{n}\left(x,z,\widehat{s}\right)=u+o\left(1\right),\quad\text{as}\quad n\to\infty.$$

Proof of Lemma 7.3 From a simple manipulation, we obtain

$$\begin{aligned}\frac{\bar{\Psi}_{N}\left(x,z,\widehat{s}\right)}{\bar{\Psi}_{D}\left(x\right)}&=\frac{1}{\sum_{i=1}^{n}\mathrm{E}\left[K_{i}\left(x\right)|\mathcal{F}_{i-1}\right]}\sum_{i=1}^{n}\mathrm{E}\left[\ K_{i}\left[\mathrm{E}\left[\psi_{x}\left(\frac{Y-z}{\widehat{s}}\right)|X_{1}\right]\right.\right.\\&\quad\left.\left.-\ \mathrm{E}\left[\psi_{x}\left(\frac{Y-z}{\widehat{s}}\right)|X=x\right]\right]|\mathcal{F}_{i-1}\ \right]+\mathrm{E}\left[\psi_{x}\left(\frac{Y-z}{\widehat{s}}\right)|X=x\right]\\&\quad-\mathrm{E}\left[\psi_{x}\left(\frac{Y-\theta(x)}{\widehat{s}}\right)|X=x\right]=D_{1}\left(x\right)+D_{2}\left(x\right).\end{aligned}$$

For $D_1\left(x\right)$, the main idea of the proof follows from Ferraty et al. (2010). Under (A3-iii), obtaining

$$\begin{aligned}A_{i}&=\mathrm{E}\left[K_{i}\left[\mathrm{E}\left[\psi_{x}\left(\frac{Y-z}{\widehat{s}}\right)|X_{i}\right]-\mathrm{E}\left[\psi_{x}\left(\frac{Y-z}{\widehat{s}}\right)|X=x\right]\right]|\mathcal{F}_{i-1}\right]\\&=\mathrm{E}\left[K_{i}\left[\mathrm{E}\left[\Psi\left(X_{i},z,\widehat{s}\right)-\Psi\left(x,z,\widehat{s}\right)|d\left(x,X_{i}\right)|\mathcal{F}_{i-1}\right]\right]\right]\\&=\mathrm{E}\left[K_{i}\Phi\left(d\left(x,X_{i}\right),z\right)|\mathcal{F}_{i-1}\right]\\&=\int\Phi\left(th,z\right)K\left(t\right)\mathrm{dP}^{\mathcal{F}_{i-1}}\left(th\right)=h\Phi^{'}\left(0,z\right)\int tK\left(t\right)\mathrm{dP}^{\mathcal{F}_{i-1}}\left(th\right).\end{aligned}$$

We use the continuity of $\Phi^{'}\left(0,\cdot\right)$, and the fact that

$$\int tK\left(t\right)d\mathrm{P}^{\mathcal{F}_{i-1}}\left(th\right)=K\left(1\right)\phi_{i}\left(x,h\right)-\int_{0}^{1}\left(sK\left(s\right)\right)^{'}\phi_{i}\left(x,sh\right)\mathrm{d}s,$$

to obtain

$$\frac{1}{n}\sum_{i=1}^{n}A_{i}=h\Phi^{'}\left(0,\theta(x)\right)\left(K\left(1\right)-\int_{0}^{1}\left(sK\left(s\right)\right)^{'}\tau_{x}\left(s\right)\mathrm{d}s\right)+o_{p}\left(h\right).$$

In similar way, we have

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[K_{i}\left(x\right)|\mathcal{F}_{i-1}\right]=\left(K\left(1\right)-\int_{0}^{1}K^{'}\left(s\right)\tau_{x}\left(s\right)\mathrm{d}s\right)+o_{p}\left(1\right).$$

Thus, we have $D_1=B_n\left(x\right)+o\left(h\right)$. Concerning D_2, we use a Taylor expansion to get, under (A3),

$$D_{2}=-B_{n}\left(x\right)+u\left[n\phi\left(x,h\right)\right]^{-1/2}\sigma\left(x,\theta(x)\right)\frac{\partial}{\partial t}\Psi\left(x,\theta(x),\widehat{s}\right)+o\left(\left[n\phi\left(x,h\right)\right]^{-1/2}\right).$$

This completes the proof. □

Lemma 7.4 Under assumptions (A1)-(A3), (A5), and (A6), we have

$$\left(\frac{n\phi\left(x,h\right)\beta_{1}^{2}}{\beta_{2}\lambda_{2}\left(x,\theta(x),\widehat{s}\right)}\right)^{1/2}R_{n}\left(x,z,\widehat{s}\right)=o\left(1\right).$$

Proof of Lemma 7.4 Here, it suffices to prove that

$$\frac{\bar{\Psi}_N(x,t,\widehat{s})-\Psi(x,t,\widehat{s})\,\bar{\Psi}_D(x)}{\bar{\Psi}_D(x)}=o_p(1)$$

and

$$\left|\widehat{\Psi}_N(x,t,\widehat{s})-\bar{\Psi}_N(x,t,\widehat{s})\right|=o_p(1).$$

In addition, we have that

$$\begin{aligned}
\frac{\bar{\Psi}_N(x,t,\widehat{s})-\Psi(x,t,\widehat{s})\,\bar{\Psi}_D(x)}{\bar{\Psi}_D(x)} &= \frac{1}{n\mathrm{E}\left[K_1(x)\right]\bar{\Psi}_D(x)}\sum_{i=1}^{n}\mathrm{E}\left[K_i(x)\,\mathrm{E}\left[\psi_x\left(\frac{Y_i-t}{\widehat{s}}\right)|\mathcal{B}_{i-1}\right]|\mathcal{F}_{i-1}\right]\\
&\quad -\Psi(x,t,\widehat{s})\,\mathrm{E}\left[K_i(x)\,|\mathcal{F}_{i-1}\right]\\
&= \frac{1}{n\mathrm{E}\left[K_1(x)\right]\bar{\Psi}_D(x)}\sum_{i=1}^{n}\mathrm{E}\left[K_i(x)\,\mathrm{E}\left[\psi_x\left(\frac{Y_i-t}{\widehat{s}}\right)|X_i\right]|\mathcal{F}_{i-1}\right]\\
&\quad -\Psi(x,t,\widehat{s})\,\mathrm{E}\left[K_i(x)\,|\mathcal{F}_{i-1}\right]\\
&\le \frac{1}{n\mathrm{E}\left[K_1(x)\right]\bar{\Psi}_D(x)}\sum_{i=1}^{n}\mathrm{E}\left[K_i(x)\,|\Psi(X_i,t,\widehat{s})-\Psi(x,t,\widehat{s})|\,|\mathcal{F}_{i-1}\right].
\end{aligned}$$

Using (A2-ii), we deduce that

$$\left|\frac{\bar{\Psi}_N(x,t,\widehat{s})-\Psi(x,t,\widehat{s})\,\bar{\Psi}_D(x)}{\bar{\Psi}_D(x)}\right|\le\sup_{x^{'}\in B(x,h)}\left|\Psi\left(x^{'},t,\widehat{s}\right)-\Psi(x,t,\widehat{s})\right|\to 0.$$

Furthermore, we get $\widehat{\Psi}_N(x,z,\widehat{s})-\bar{\Psi}_N(x,z,\widehat{s})=o_p(1)$. Now, we must prove that $\mathrm{E}[\widehat{\Psi}_N(x,z,\widehat{s})-\bar{\Psi}_N(x,z,\widehat{s})]\to 0$ and $\mathrm{Var}[\widehat{\Psi}_N(x,z,\widehat{s})-\bar{\Psi}_N(x,z,\widehat{s})]\to 0$. The first one is a consequence of the definitions of $\widehat{\Psi}_N(x,z,\widehat{s})$, and $\bar{\Psi}_N(x,z,\widehat{s})$. For the second one, we obtain $\widehat{\Psi}_N(x,z,\widehat{s})-\bar{\Psi}_N(x,z,\widehat{s})=\sum_{i=1}^{n}\delta_i(x,z,\widehat{s})$, where

$$\delta_i(x,z,\widehat{s})=\frac{1}{n\mathrm{E}\left[K_1\right]}K_i\psi_x\left(\frac{Y_i-z}{\widehat{s}}\right)-\mathrm{E}\left[K_i\psi_x\left(\frac{Y_i-z}{\widehat{s}}\right)|\mathcal{F}_{i-1}\right].$$

By the Burkholder inequality, we have

$$\mathrm{E}\left[\sum_{i=1}^{n}\delta_i(x,z,\widehat{s})\right]^2\le\sum_{i=1}^{n}\mathrm{E}\left[\delta_i(x,z,\widehat{s})\right]^2.$$

In addition, by the Jensen inequality, we arrive at

$$\mathrm{E}^2\left[\delta_i(x,z,\widehat{s})\right]\le\frac{1}{n^2\mathrm{E}^2\left[K_1\right]}\mathrm{E}\left[K_i^2\psi_x^2\left(\frac{Y_i-z}{s}\right)\right]\le\frac{1}{n^2\mathrm{E}^2\left[K_1\right]}\mathrm{E}\left[K_i^2\right]\le\frac{1}{n\phi^2(x,h)}\phi_i(x,h).$$

Now, (A2) yields $\mathrm{Var}\left[\widehat{\Psi}_N(x,z,\widehat{s})-\bar{\Psi}_N(x,z,\widehat{s})\right]\to 0.$ □

Lemma 7.5 Under assumptions (A1), (A2), (A5), and (A6), $\widehat{\theta}(x)$ exists a.s. for all sufficiently large n.

Proof of Lemma 7.5
From the monotonicity of $\psi_x(Y-./\widehat{s})$, for all $\varepsilon>0$, we have

$$\Psi(x,\theta(x)-\varepsilon,\widehat{s})\le\Psi(x,\theta(x),\widehat{s})\le\Psi(x,\theta(x)+\varepsilon,\widehat{s}).$$

By using a similar argument as those used in the previous Lemmas, we prove that

$$\widehat{\Psi}(x,t,\widehat{s})\xrightarrow{\mathrm{P}}\Psi(x,t,\widehat{s})\,m\quad\forall t\in N_x.$$

Thus, for sufficiently large n and for all ε small enough, we reach $\widehat{\Psi}\left(x, \theta(x)-\varepsilon, \widehat{s}\right) \leq 0 \leq \widehat{\Psi}\left(x, \theta(x)+\varepsilon, \widehat{s}\right)$, which holds with probability tending to one.

Since ψ_x is a continuous function, it follows that $\widehat{\Psi}\left(x, t, \widehat{s}\right)$ is a continuous function of t and, there exists $\widehat{\theta}\left(x\right) \in [\theta(x)-\varepsilon, \theta(x)+\varepsilon]$ such that $\widehat{\Psi}(x, \widehat{\theta}(x), \widehat{s}) = 0$. Hence, the uniqueness of $\widehat{\theta}\left(x\right)$ is a direct consequence of the strict monotonicity of ψ_x in the second component and the fact that

$$\mathrm{P}\left(\sum_{i=1}^{n} K_i = 0\right) = \mathrm{P}\left(\widehat{\Psi}_D\left(x\right) = 0\right) \to 0 \quad \text{as} \quad n \to \infty,$$

which implies $\sum_{i=1}^{n} K_i \neq 0$ with probability tending to 1. Moreover, since $\widehat{\theta}(x) \in [\theta(x)-\varepsilon, \theta(x)+\varepsilon]$ in probability, it follows that $\widehat{\theta}(x) \xrightarrow{\mathrm{P}} \theta(x)$, as $n \to \infty$. $\square$

Information for authors

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to the Editors-in-Chief of the ChJS (E-mail: `chilean.journal.of.statistics@gmail.com`). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a"cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

Preparation of accepted manuscripts

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at `http://chjs.mat.utfsm.cl/files/ChJS.zip`. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at `http://www.ams.org/mathscinet/msc/`. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., 1994. Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., 1997. Local influence. In Kotz, S., Read, C.B., and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 1., Wiley, New York, pp. 380-385.

Rukhin, A.L., 2009. Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters, 79, 1004-1007.

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Peña, D., and Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika, 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

Copyright