Electronic Journal of Statistics
Vol. 10 (2016) 1973–1999
ISSN: 1935-7524
DOI: 10.1214/16-EJS1156

Bootstrap confidence intervals in functional nonparametric regression under dependence

Paula Raña, Germán Aneiros, Juan Vilar

Departamento de Matemáticas, Universidade da Coruña, A Coruña, Spain
e-mail: paula.rana@udc.es; german.aneiros@udc.es; juan.vilar@udc.es

and

Philippe Vieu

Institut de Mathématiques, Université Paul Sabatier, Toulouse, France
e-mail: philippe.vieu@math.univ-toulouse.fr

Abstract: This paper considers naive and wild bootstrap procedures to construct pointwise confidence intervals for a nonparametric regression function when the predictor is of functional nature and when the data are dependent. Assuming α-mixing conditions on the sample, the asymptotic validity of both procedures is obtained. A simulation study shows promising results when finite sample sizes are used, while an application to electricity demand data illustrates its usefulness in practice.

MSC 2010 subject classifications: 62G08, 62G09, 62G20.
Keywords and phrases: Functional data, bootstrap, nonparametric regression, confidence intervals, α-mixing.

Received December 2015.

Contents

1. Introduction

In time series analysis, autoregressive models are commonly used to model the dependence between past and future values and, then, to provide some predictive methodology. In the very earlier literature, only a finite (usually small) number of past observations of the process have been included in the model in some parametric (i.e. linear) way (see e.g. Box and Jenkins, 1976 [6]). Then, starting with Collomb (1984) [8], nonparametric autoregressive alternative models have been developed (see eg Györfi et al., 1989 [23], or Bosq, 1998 [4], for general monographs on the field). While this nonparametric literature has provided interesting improvements for time series analysis, it was still suffering from the fact of incorporating only a finite (and necessarily small) number of past data for prediction. Roughly in the same moment, along the last twenty years there has been a growth in interest for developing methodologies for continuous data, and a new field of Statistics, called Functional Data Analysis, was borned (see Ramsay and Silverman, 2005 [32], for earlier popularization of this field, Horvath and Kokozka, 2012 [27], and Hsing and Eubank, 2015 [28], for the most recent monographs, and Cuevas, 2014 [9], and Goia and Vieu, 2016 [22], for recent selected reviews). Both fields Time Series and Functional Data Analysis crossed over naturally, and autoregressive models have been extended to incorporate infinite number of past values in the prediction model, through some kind of regression model with continuous covariates:

$$G(\boldsymbol{\chi}_{i+1}) = m(\boldsymbol{\chi}_i) + \varepsilon_i \ (i = 1, \dots, n), \tag{1.1}$$

where $G(\cdot)$ is a known operator and $m(\cdot)$ is an unknown operator to be estimated, while $\boldsymbol{\chi}_i$ are functional random variables and ε_i are mean zero regression errors.

In the last few years many methodological advances for model (1.1) have been proposed, which can be classified into three categories according to the kind of assumptions allowed for the target operator $m(\cdot)$: this goes from parametric modeling as in Bosq (2000) [5] with linear assumptions on $m(\cdot)$, until free modeling approaches through some smooth nonparametric assumptions on $m(\cdot)$ as in Ferraty and Vieu (2006) [20], and with intermediate semiparametric modeling as in Aneiros-Pérez and Vieu (2008) [1]. The practical interest of these new functional time series methodological approaches have been highlighted by means of various applied case studies as in Antoch et al. (2010) [3] for parametric models, in Ferraty, Rabhi, and Vieu (2005) [16] for nonparametric modeling or in Aneiros et al. (2013) [2] for semiparametric models. See also Fernández-de-Castro, Guillas, and González-Manteiga (2005) [12] for a nice applied comparison between predictions of sulfur dioxide levels from functional autoregressive linear and nonparametric models.

Even in the simplest case of finite-dimensional and independent predictors, it is well-known that one of the main problems when stating asymptotic expansions for nonparametric estimates is the fact that rather complicated expressions appear, both on the bias and on the variance of the limit distribution, which are difficult to estimate directly (see the discussion in Härdle and Marron, 1991 [26]). This problem is even more evident in the case of continuous predictors (see Ferraty, Mas, and Vieu, 2007 [15]) in which the bias and variance expressions are dramatically much more intricate. They are, once again, much more complicated when some dependence structure has to be involved (see Masry, 2005 [30]), as it is the case when time series are considered (e.g. in models like (1.1)), because of the additional covariance appearing in the asymptotics. At the end, while such limit distributions in nonparametric functional time series analysis are of high theoretical interest, they still suffer for real practical possibility of applications because of this phenomenon.

The aim of our paper is to investigate the question of practical use of functional nonparametric time series predictions by providing a bootstrapping procedure for overcoming the difficulty related to the estimation of the constants in the limit distribution. More precisely, we are interested on time series prediction from the nonparametric model (1.1) with scalar response (i.e. assuming that the operator $G(\cdot)$ is real valued), and in which the dependence model is controlled by means of some strong dependence condition: namely the covariates $\boldsymbol{\chi}_i$ are identically distributed functional random variables verifying some α-mixing condition. Other less important restrictions on our model are that the ε_i are i.i.d. mean zero random errors, and that $\boldsymbol{\chi}_i$ are valued in some infinite-dimensional space $\mathcal{H}$, which is endowed with a semi-metric $d(\cdot,\cdot)$.

The paper is organized as follows. In Section 2 we present a general nonparametric functional model and the corresponding estimator. Discussion on how bootstrapping procedures can be constructed for this functional dependent framework is shown in Section 3. More precisely, we present two kinds of bootstrapping techniques: one intended to homoscedastic data (this is the so-called naive bootstrapping), while a more sophisticated one (so-called wild bootstrapping) is introduced for heteroscedastic data. A theoretical study of the procedures is carried, in which the validity of the two (naive and wild) bootstrapping is stated by showing how their limit distributions are asymptotically equivalent to the ones of the nonparametric predictors. Thus, by producing various iterations of the bootstrapping procedure one gets empirical approximations of the limit distribution of the predictors (without any need for estimating highly complex bias or variance expressions), making this procedure easily usable in practice (as shown in Section 4 through a simulation study). One of the practical interests of having easy approximations of the predictor errors is for constructing confidence intervals, which is illustrated in Section 5 by providing some real case study linked with electricity demand data. Finally, Section 6 concludes our paper by proposing some challenges to be dealt in a future.

2. The model and the estimator

Although we are interested in time series prediction from the model (1.1) with scalar response, we will obtain asymptotic theory for the more general model

$$Y_i = m(\boldsymbol{\chi}_i) + \varepsilon_i, \tag{2.1}$$

where the process $\{(\boldsymbol{\chi}_i, Y_i)\}$ is α-mixing and identically distributed as $(\boldsymbol{\chi}, Y)$. In this way, our results will be valid even when the response is exogenous. As indicated in Section 1, the response, Y, is scalar while the covariate, $\boldsymbol{\chi}$, is valued in some infinite-dimensional space $\mathcal{H}$, which is endowed with a semi-metric $d(\cdot,\cdot)$. Finally, $m(\cdot)$ is an unknown smooth real-valued operator and the corresponding random errors $\{\varepsilon_i\}$ are i.i.d. as ε, and we assume that $\mathbf{E}(\varepsilon|\boldsymbol{\chi}) = 0$ and $\mathbf{E}(\varepsilon^2|\boldsymbol{\chi}) = \sigma_\varepsilon^2(\boldsymbol{\chi}) < \infty$.

Given a fixed element χ of the space $\mathcal{H}$, the remainder of this paper focuses on inference on $m(\chi)$ in model (2.1) from the sample

$$\mathcal{S} = \{(\boldsymbol{\chi}_1, Y_1), \ldots, (\boldsymbol{\chi}_n, Y_n)\}$$

we have at hand. Specifically, the aim is to construct confidence intervals for $m(\chi)$ based on the estimator

$$\widehat{m}_h(\chi) = \frac{\sum_{i=1}^n K(d(\boldsymbol{\chi}_i, \chi)/h) Y_i}{\sum_{i=1}^n K(d(\boldsymbol{\chi}_i, \chi)/h)}, \tag{2.2}$$

where $K(\cdot)$ is a kernel function and $h > 0$ is a smoothing parameter. On the one hand, in the setting of independent data $\{(\boldsymbol{\chi}_i, Y_i)\}$, the topic of confidence intervals in functional nonparametric regression was dealt in Ferraty et al. (2007) [15], which obtained the asymptotic normality of the properly standardized estimator $\widehat{m}_h(\chi)$. Then, by estimating the constants involved in the standardized estimator, one can construct the corresponding confidence intervals. The main drawback of this procedure is that such constants could be difficult to estimate (for some simple examples, see Proposition 1 in Ferraty et al., 2007 [15]). This drawback was overcome in Ferraty, Van Keilegom, and Vieu (2010) [17] by means of bootstrapping techniques, by approximating directly the distribution of the estimation error without having to estimate the constants involved in the standardized estimator. On the other hand, some studies exist in the case of dependent data $\{(\boldsymbol{\chi}_i, Y_i)\}$. For instance, Masry (2005) [30] and Delsol (2009) [11] obtained the asymptotic normality of the properly standardized estimator $\widehat{m}_h(\chi)$, under α-mixing conditions. The main advantage of the results in Delsol (2009) [11] against the ones in Masry (2005) [30] is the fact that Delsol obtained explicit constants, which is not the case of Masry (2005) [30]. As in the setting of independent data recently referred, there exist situations where the constants given in Delsol (2009) [11] are difficult to estimate, and this drawback could be overcome, again, through implementation of bootstrap techniques. In next Section 3, we will present two bootstrap procedures designed for that.

3. Asymptotic theory

This section proposes two bootstrap procedures and presents our main result: their asymptotic validity. First, the considered assumptions are stated and some comments on them are given.

3.1. Assumptions

We start with some notation. For a given fixed element χ of the space $\mathcal{H}$, we denote:

$$\begin{aligned} B(\chi, l) &= \{\chi_1 \in \mathcal{H} \text{ such that } d(\chi_1, \chi) \le l\},\ F_\chi(l) = P(\boldsymbol{\chi} \in B(\chi, l)) \text{ for } l > 0, \\ \varphi_\chi(s) &= \mathbb{E}(m(\boldsymbol{\chi}) - m(\chi) | d(\boldsymbol{\chi}, \chi) = s),\ \tau_{h\chi}(s) = F_\chi(hs)/F_\chi(h) \text{ for } s \in (0,1] \end{aligned}$$

and

$$\tau_{0\chi}(s) = \lim_{h \downarrow 0} \tau_{h\chi}(s).$$

In addition, let

$$\begin{aligned} M_{0\chi} &= K(1) - \int_0^1 (sK(s))' \tau_{0\chi}(s) ds, \\ M_{1\chi} &= K(1) - \int_0^1 K'(s) \tau_{0\chi}(s) ds, \\ M_{2\chi} &= K^2(1) - \int_0^1 (K^2(s))' \tau_{0\chi}(s) ds \end{aligned}$$

and

$$\Theta(s) = \max\{\max_{i \ne j} P(d(\boldsymbol{\chi}_i, \chi) \le s, d(\boldsymbol{\chi}_j, \chi) \le s), F_\chi^2(s)\}.$$

As noted at the beginning of Section 3, we will propose two bootstrap procedures and will prove their asymptotic validity. For that, we will obtain that both the standard estimator, $\widehat{m}_h(\chi)$, and the bootstrap version, $\widehat{m}^*_{hb}(\chi)$, (properly standardized) converge to the same distribution, existing, in addition, a third negligible term (for details, see Section 8). The following set of assumptions guarantees the convergence of $\widehat{m}_h(\chi)$:

- (H1) $m(\cdot)$ and $\sigma^2_\varepsilon(\cdot)$ are continuous on a neighbourhood of χ, and $\sigma^2_\varepsilon(\chi) > 0$.
- (H2) $F_\chi(0) = 0$, $\varphi_\chi(0) = 0$ and $\varphi'_\chi(0)$ exists.
- (H3) $\forall s \in [0,1]$, $\lim_{n\to\infty} \tau_{h\chi}(s) = \tau_{0\chi}(s)$ with $\tau_{0\chi}(s) \ne 1_{[0,1]}(s)$.
- (H4) $\exists p > 2$, $\exists M > 0$ such that $\mathbb{E}(|\varepsilon|^p | \boldsymbol{\chi}) \le M$ *a.s.*
- (H5) $\max\{\mathbb{E}(|Y_i Y_j|^p | \boldsymbol{\chi}_i, \boldsymbol{\chi}_j), \mathbb{E}(|Y_i|^p | \boldsymbol{\chi}_i, \boldsymbol{\chi}_j)\} \le M$ *a.s.* $\forall i, j \in \mathbb{Z}$.
- (H6) $h(nF_\chi(h))^{1/2} = O(1)$ and $\lim_{n\to\infty} nF_\chi(h) = \infty$.
- (H7) $K(\cdot)$ is supported on $[0,1]$ and has a continuous derivative on $[0,1)$. In addition, $K(1) > 0$ and $K'(s) \le 0$ for $s \in [0,1)$.
- (H8) $\{(\boldsymbol{\chi}_i, Y_i)\}_{i=1}^n$ comes from a α-mixing process with α-mixing coefficients $\alpha(n) \le Cn^{-a}$.

(H9) $\exists v > 0$ such that $\Theta(h) = O(F_\chi(h)^{1+v})$ with $a > \frac{(1+v)p-2}{v(p-2)}$ (p and a were defined in (H4) and (H8), respectively).

(H10) $\exists \gamma > 0$ such that $nF_\chi(h)^{1+\gamma} \to \infty$ and $a > \max\left\{\frac{4}{\gamma}, \frac{p}{p-2} + \frac{2(p-1)}{\gamma(p-2)}\right\}$ (p and a were defined in (H4) and (H8), respectively).

Assumptions (H1)–(H10) are standard in the setting of nonparametric regression with functional data. They were justified (from a theoretical nature) and used in Delsol (2009) [11]. The α-mixing structure (H8)–(H10) is quite usual and general to model dependency structure. Some slight possible extension (as indicated in Delsol, 2009 [11] or Masry, 2005 [30]) could be incorporated but the price to pay would be much more tedious calculations, and this would mask the main purpose (bootstraping) of our paper. Finally, from the practical point of view, it is needed to show some processes satisfying our assumptions (in such a way that our methodology can be applied). Focusing on the conditions really linked to the functional nature, the main role is played by the 'small ball probabilities'; that is, by the function $F_\chi(\cdot)$. As can be seen in Ferraty, Laksaci, and Vieu (2006) [14], when the functional space is endowed with a suitable semi-metric, the function $F_\chi(\cdot)$ for standard Ornstein-Uhlenbeck, general diffusion, fractional Brownian motion and general Gaussian processes takes the form

$$F_\chi(h) \sim C_\chi \exp\left(-C/h^\beta\right), \tag{3.1}$$

where C_χ, C and β are positive constants. Therefore, as proven in Ferraty, Mas and Vieu (2007) [15], one has that $\tau_{0\chi}(s) = \delta_1(s)$, where $\delta_1(\cdot)$ stands for the Dirac mass at 1. Another usual example is that of the fractal processes, where $F_\chi(h) \sim C_\chi h^u$ for some $u > 0$ (see Pesin, 1993 [31], for the definition of fractal dimension for a process). In this case, it is obvious that $\tau_{0\chi}(s) = s^u$ (in short, one can find a suitable set of parameters in such a way that our assumptions on $F_\chi(\cdot)$ and $\tau_{0\chi}(\cdot)$ are satisfied by the processes recently named). The last assumption to justify from a practical point of view is (H9), which imposes conditions on the dependence structure in the functional process. Focusing on fractal processes and using again the definition given in Pesin (1993) [31], one has that $\Theta(h) \sim Ch^{u+\min\{u,u_0\}}$ (for some $u, u_0 > 0$). So, one can find a suitable set of parameters in such a way that (H9) is satisfied.

Now, we state a second set of assumptions which should be added to the first one in order to obtain the convergence of $\widehat{m}^*_{hb}(\chi)$ (the bootstrap version of $\widehat{m}_h(\chi)$) and the negligible nature of the third term to which we have referred previously:

(H11) The function $\mathbb{E}(|Y||\boldsymbol{\chi} = \cdot)$ is continuous on a neighbourhood of χ. In addition, for some $\delta > 0$ and for all $q \geq 1$, $\sup_{d(\chi_1,\chi)<\delta} \mathbb{E}(|Y|^q|\boldsymbol{\chi} = \chi_1) < \infty$.

(H12) $\forall(\chi_1, s)$ in a neighbourhood of $(\chi, 0)$, $\varphi_{\chi_1}(0) = 0$, $\varphi'_{\chi_1}(s)$ exists, $\varphi'_{\chi_1}(0) \neq 0$ and $\varphi'_{\chi_1}(s)$ is uniformly Lipschitz continuous of order $0 < \lambda \leq 1$ in (χ_1, s).

(H13) $\forall \chi_1 \in \mathcal{H}$, $F_{\chi_1}(0) = 0$ and $F_{\chi_1}(t)/F_\chi(t)$ is Lipschitz continuous of order λ in χ_1, uniformly in t in a neighbourhood of 0 (λ was defined in (H12))

(H14) $\forall \chi_1 \in \mathcal{H}$ and $\forall s \in [0,1]$, $\tau_{0\chi_1}(s)$ exists, $\sup_{\chi_1 \in \mathcal{H}, s\in[0,1]} |\tau_{h\chi_1}(s) - \tau_{0\chi_1}(s)| = o(1)$, $\inf_{d(\chi_1,\chi)<\epsilon} M_{1\chi} > 0$ for some $\epsilon > 0$, and $M_{k\chi_1}$ is Lipschitz continuous of order λ for $k = 0, 1, 2$ (λ was defined in (H12)). In addition, $M_{0\chi} > 0$ and $M_{2\chi} > 0$.

(H15) $\forall n \; \exists r_n \geq 1$, $l_n > 0$ and curves $\chi_{1n}, \ldots, \chi_{r_n n}$ such that $B(\chi, h) \subset \cup_{k=1}^{r_n} B(\chi_{kn}, l_n)$, with $r_n = O(n^{b/h})$ and $l_n = o(b(nF_\chi(h))^{-1/2})$.

(H16) $\max\{b, h/b, b^{1+\lambda}(nF_\chi(h))^{1/2}, (F_\chi(h)/F_\chi(b)) \log n, n^{1/p} F_\chi(h)^{1/2} \log n\} = o(1)$, $\lim_{n\to\infty} F_\chi(b+h)/F_\chi(b) = 1$ and $\max\{bh^{\lambda-1}, F_\chi(b)^{-1}h/b\} = O(1)$ (p and λ were defined in (H12) and (H4), respectively).

(H17) $a > 4.5$ (a was defined in (H8)).

Note that Assumption (H17) together with $n^{1/p}F_\chi(h)^{1/2} \log n = o(1)$ and $F_\chi(b)^{-1}h/b = O(1)$ in Assumption (H16) allow us to apply the Lemma 3 in Aneiros-Pérez and Vieu (2008) [1] (see the last part in the proof of our Theorem 3.1). Note also that all the other assumptions were used in Ferraty, Van Keilegom, and Vieu (2010) [17] to attain the validity of the bootstrap in the independent case. Finally, a special attention has to be given to the part of Assumption (H15) related to the balls, which is only necessary to make use of the results of uniform consistency of nonparametric regression smoothers. Therefore, it can be weakened by changing it into any other kind of assumptions insuring such uniform consistency.

3.2. The bootstrap procedures

We focus on both naive and wild bootstrap procedures, which have been successfully used in the literature related to regression models (in the setting of scalar variables, see for instance Freedman, 1981 [21], and Mammen, 1993 [29] for linear models, and Cao, 1991 [7] , Härdle and Marron, 1991 [26], and Hall, 1992 [25], for nonparametric models; the setting of functional variables was studied, for instance, in González-Manteiga and Martínez-Calvo, 2011 [24], for linear models while Ferraty, Van Keilegom, and Vieu, 2010 [17], 2012 [18], focused on nonparametric models).

The algorithms for resamplings proceed as follows:

Naive bootstrap. This procedure is designed for the case where the model is homoscedastic; that is, $\sigma_\varepsilon^2(\boldsymbol{\chi}) = \sigma_\varepsilon^2$.

Step 1: Construct the residuals $\widehat{\varepsilon}_{i,b} = Y_i - \widehat{m}_b(\boldsymbol{\chi}_i)$, $i = 1, \ldots, n$.

Step 2: Draw n i.i.d random variables $\varepsilon_1^*, \ldots, \varepsilon_n^*$ from the empirical distribution function of $(\widehat{\varepsilon}_{1,b} - \overline{\widehat{\varepsilon}}_b, \ldots, \widehat{\varepsilon}_{n,b} - \overline{\widehat{\varepsilon}}_b)$, where $\overline{\widehat{\varepsilon}}_b = n^{-1}\sum_{i=1}^n \widehat{\varepsilon}_{i,b}$.

Step 3: Obtain $Y_i^* = \widehat{m}_b(\boldsymbol{\chi}_i) + \varepsilon_i^*$, $i = 1, \ldots, n$.

Step 4: Define $\widehat{m}_{hb}^*(\chi) = \dfrac{\sum_{i=1}^n K(d(\boldsymbol{\chi}_i, \chi)/h) Y_i^*}{\sum_{i=1}^n K(d(\boldsymbol{\chi}_i, \chi)/h)}$.

Wild bootstrap. This procedure allows the possibility of heteroskedasticity, but is also applicable in the case of homoscedasticity. All one must do is to change Step 2 in the naive bootstrap: define $\varepsilon_i^* = \widehat{\varepsilon}_{i,b} V_i$, $i = 1, \ldots, n$, where $V_1, \ldots, V_n$ are i.i.d. random variables that are independent of the data $\mathcal{S}$ and that satisfy $\mathbb{E}(V_1) = 0$ and $\mathbb{E}(V_1^2) = 1$. Maintain the other three steps.

As usual when one deals with asymptotics related to bootstrap procedures in nonparametric regression, two bandwidths are involved in both algorithms. The first bandwidth, b, is used to construct the residuals to resampling. Then, a second bandwidth, h, is considered to smooth the bootstrap sample. Our assumptions imposed to obtain the asymptotic validity of the proposed bootstrap procedures require that b must be taken to be larger than h (in the same way as in the independent case dealt in Ferraty, Van Keilegom, and Vieu, 2010 [17], 2012 [18]; see also Cao, 1991 [7], and Härdle and Marron, 1991 [26], for the scalar case).

3.3. Asymptotic result

Let $P^{\mathcal{S}}$ denotes probability, conditionally on the sample $\mathcal{S}$, and let us suppose that χ is a fixed element of the space $\mathcal{H}$.

Theorem 3.1. *Under assumptions (H1)–(H17), for the wild bootstrap procedure, we have that*

$$\sup_{y \in \mathbb{R}} \left| P^{\mathcal{S}} \left(\sqrt{nF_\chi(h)} (\widehat{m}_{hb}^*(\chi) - \widehat{m}_b(\chi)) \leq y \right) - \right. \\ \left. P \left(\sqrt{nF_\chi(h)} (\widehat{m}_h(\chi) - m(\chi)) \leq y \right) \right| \to 0 \ a.s.$$

In addition, if the model is homoscedastic (i.e. $\sigma_\varepsilon^2(\cdot) = \sigma_\varepsilon^2$*), then the same result holds for the naive bootstrap.*

Theorem 3.1 extends Theorem 1 in Ferraty, Van Keilegom, and Vieu (2010) [17] from the independent case to the dependent one. Its main practical usefulness is related to the building of confidence intervals for $m(\chi)$ in a context of dependent data. As noted in Section 2, due to the (most of the times) difficulty in estimating the constants involved in the standardization of $\widehat{m}_h(\chi)$, the asymptotic distribution of the true error $\widehat{m}_h(\chi) - m(\chi)$ could be useless to construct the desired confidence interval. Nevertheless, from Theorem 3.1, we can approximate the quantiles of the distribution of $\widehat{m}_h(\chi) - m(\chi)$ by means of the quantiles of the distribution of the bootstrapped error $\widehat{m}_{hb}^*(\chi) - \widehat{m}_b(\chi)$. Then, because we can generate more and more replicates (said B replicates) of such bootstrapped error, Theorem 3.1 together with the percentile method (for instance) allows us to build confidence intervals for $m(\chi)$ without estimating the constants involved in the standardization of $\widehat{m}_h(\chi)$ (see Section 4.1 for details).

4. Simulation study

This section is devoted to illustrate, when finite sample sizes are used, the accuracy of the confidence interval for $m(\chi)$ constructed from the proposed bootstrap methodology. For that, such interval will be compared with the true (and, in practice, unknown) confidence interval. In addition, to show the behavior of our bootstrap interval against that of the interval obtained from the asymptotic distribution of $\widehat{m}_h(\chi)$, some results for the asymptotic interval will be given. Because its generality, we focus on the wild bootstrap procedure.

In a first example we consider smooth curves while in a second one the case of rough curves is dealt.

4.1. Building the confidence intervals

Given a curve χ and a model

$$Y_i = m(\boldsymbol{\chi}_i) + \varepsilon_i \ (i = 1, \dots, n), \tag{4.1}$$

where the process $\{(\boldsymbol{\chi}_i, Y_i)\}$ is α-mixing and identically distributed as $(\boldsymbol{\chi}, Y)$, and χ is observed from $\boldsymbol{\chi}$, the true, bootstrap and asymptotic $(1-\alpha)$-confidence intervals for $m(\chi)$ were constructed as

$$I_{\chi,1-\alpha}^{true} = (\widehat{m}_h(\chi) + q_{\alpha/2}^{true}(\chi), \widehat{m}_h(\chi) + q_{1-\alpha/2}^{true}(\chi)),$$

$$I_{\chi,1-\alpha}^{*} = (\widehat{m}_h(\chi) + q_{\alpha/2}^{*}(\chi), \widehat{m}_h(\chi) + q_{1-\alpha/2}^{*}(\chi))$$

and

$$I_{\chi,1-\alpha}^{asymp} = (\widehat{m}_h(\chi) + q_{\alpha/2}^{asymp}(\chi), \widehat{m}_h(\chi) + q_{1-\alpha/2}^{asymp}(\chi)),$$

respectively, where the quantiles $q_p^{true}(\chi)$, $q_p^{*}(\chi)$ and $q_p^{asymp}(\chi)$ were computed in the following way:

- Theoretical quantiles ($q_p^{true}(\chi)$).
    1. Generate n_{MC} samples $\{(\boldsymbol{\chi}_i^s, Y_i^s), i = 1, \dots, n\}_{s=1}^{n_{MC}}$ from Model (4.1).
    2. Carry out n_{MC} estimates $\{\widehat{m}_h^s(\chi)\}_{s=1}^{n_{MC}}$, where $\widehat{m}_h^s(\cdot)$ is the functional kernel estimator (2.2) derived from the s^{th} sample $\{(\boldsymbol{\chi}_i^s, Y_i^s)\}_{i=1}^{n}$.
    3. Compute the set of approximation errors $ERRORS.MC = \{\widehat{m}_h^s(\chi) - m(\chi)\}_{s=1}^{n_{MC}}$.
    4. Compute the theoretical quantile, $q_p^{true}(\chi)$, from the quantile of order p of $ERRORS.MC$.
- Bootstrap quantiles ($q_p^{*}(\chi)$).
    1. Generate the sample $\mathcal{S} = \{(\boldsymbol{\chi}_1, Y_1), \dots, (\boldsymbol{\chi}_n, Y_n)\}$ from Model (4.1).
    2. Compute $\widehat{m}_b(\chi)$ over the dataset $\mathcal{S}$.

3. Repeat B times the bootstrap algorithm over $\mathcal{S}$ by using i.i.d. random variables V_i drawn from the two Dirac distributions $0.1(5+\sqrt{5})\delta_{(1-\sqrt{5})/2} + 0.1(5-\sqrt{5})\delta_{(1+\sqrt{5})/2}$, giving the B estimates $\left\{\widehat{m}_{hb}^{*,r}(\chi)\right\}_{r=1}^{B}$.
4. Compute set of bootstrap errors $ERRORS.BOOT = \left\{\widehat{m}_{hb}^{*,r}(\chi) - \widehat{m}_b(\chi)\right\}_{r=1}^{B}$.
5. Compute the bootstrap quantile, $q_p^*(\chi)$, from the quantile of order p of $ERRORS.BOOT$.

- Asymptotic quantiles ($q_p^{asymp}(\chi)$).
  1. Generate the sample $\mathcal{S} = \{(\boldsymbol{\chi}_1, Y_1), \ldots, (\boldsymbol{\chi}_n, Y_n)\}$ from Model (4.1).
  2. Use the sample $\mathcal{S}$ to estimate the constants $F_\chi(h)$, $M_{1\chi}$, $M_{2\chi}$ and σ_ε as suggested in Delsol (2009) [11], pages 18 and 20.
  3. Compute the asymptotic quantile, $q_p^{asymp}(\chi)$, from the quantile of order p of the corresponding normal distribution.

Finally, the estimate $\widehat{m}_h(\chi)$ in each of the three intervals was obtained from $\mathcal{S}$.

The quadratic kernel, $K(u) = 1.5(1-u^2)1_{[0,1]}(u)$, was considered in the estimates $\widehat{m}_h$ and $\widehat{m}_{hb}^*$, while the bandwidth $b = b_{CV}$ was selected by means of the cross-validation methodology. Then, $h = b_{CV}$ was set.

4.2. Smooth curves

Our first model, Model 1, is based on the one used in Delsol (2009) [11], where smooth curves $\boldsymbol{\chi}$ were considered. Some modifications were included to adapt his model to our context. Specifically, the discretized functional covariate in Model 1 was

$$\boldsymbol{\chi}_i(t_j) = \cos(a_i + \pi(2t_j - 1)), \tag{4.2}$$

where $\{a_i\}$ comes from AR(1) gaussian process with correlation coefficient $\rho_a = 0.7$ and variance $\sigma_a^2 = 0.05$. Values $0 = t_1 < t_2 < \cdots < t_{99} < t_{100} = 1$ equally spaced were considered. The regression operator was

$$m(\boldsymbol{\chi}) = \frac{1}{2\pi}\int_{1/2}^{3/4} (\boldsymbol{\chi}'(t))^2 dt$$

while the errors $\{\varepsilon_i\}$ were independent centered gaussians of variance equal to 0.1 times the empirical variance of $\{m(\boldsymbol{\chi}_1), \ldots, m(\boldsymbol{\chi}_n)\}$.

Note that Model 1 deals with smooth curves (see left panel in Figure 1), this fact suggesting the use of a semi-metric based on some derivative of the curve (for details, see Section 13.6 in Ferraty and Vieu, 2006 [20]). Specifically, as recommended in Delsol (2009) [11], the semi-metric ($d_1^{deriv}(\cdot,\cdot)$) considered in Model 1 was based on the first derivative of the curve:

$$d_1^{deriv}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j) = \sqrt{\int_0^1 (\boldsymbol{\chi}_i'(t) - \boldsymbol{\chi}_j'(t))^2 dt}.$$

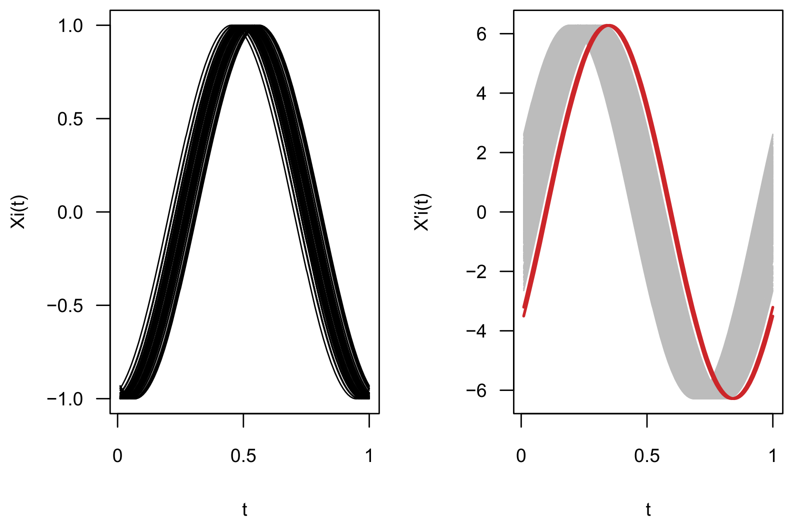

FIG 1. *Left panel: first 50 curves in a training sample $\mathcal{S}$ ($n = 250$ was considered) generated from Model 1. Right panel: first derivative of the curves in $\mathcal{S}$, together with the first derivative of the curves χ_{41} and χ_{94} (in red) in the test sample $\mathcal{C}$.*

TABLE 1
Average over $\mathcal{C}$ of the empirical coverage of the true, bootstrap and asymptotic confidence intervals from Model 1. Standard deviation appears in brackets.

$1-\alpha$	0.95		0.90	
n	100	250	100	250
Coverage (I^{true})	0.946 (0.122)	0.951 (0.012)	0.896 (0.018)	0.903 (0.015)
Coverage (I^{*})	0.890 (0.116)	0.921 (0.076)	0.849 (0.118)	0.877 (0.080)
Coverage (I^{asymp})	0.852 (0.141)	0.898 (0.114)	0.794 (0.143)	0.842 (0.117)

True, bootstrap and asymptotic $(1-\alpha)$-confidence intervals for $m(\chi)$ with $\chi \in \mathcal{C}$ were computed and compared. The test sample $\mathcal{C} = \{\chi_1, \ldots, \chi_{n_\mathcal{C}}\}$, consisting in $n_\mathcal{C}$ independent curves, was generated in the following way: first, $n_\mathcal{C}$ independent functional time series were obtained from the process $\{\boldsymbol{\chi}_i\}$ defined in (4.2); then, a curve χ was selected at random in each of such $n_\mathcal{C}$ functional time series. Note that from the procedure explained in the previous Section 4.1 one obtains one $(1-\alpha)$-confidence interval of each type for $m(\chi)$: true ($I^{true}_{\chi,1-\alpha}$), bootstrap ($I^{*}_{\chi,1-\alpha}$) and asymptotic ($I^{asymp}_{\chi,1-\alpha}$) confidence intervals. To compare the accuracy of each type of interval, we obtained the empirical coverages by repeating the procedure M times and computing the proportion of times that each interval contains the value $m(\chi)$.

Values $n_{MC} = 2000$, $B = 500$, $n_\mathcal{C} = 100$, $M = 500$, $1-\alpha = 0.95, 0.90$ and $n = 100, 250$ were considered.

Table 1 reports the average over $\mathcal{C}$ of the empirical coverage of the three computed confidence intervals. As expected, the accuracy of the coverages improves as the sample size, n, increases. In addition, coverages of the bootstrap intervals

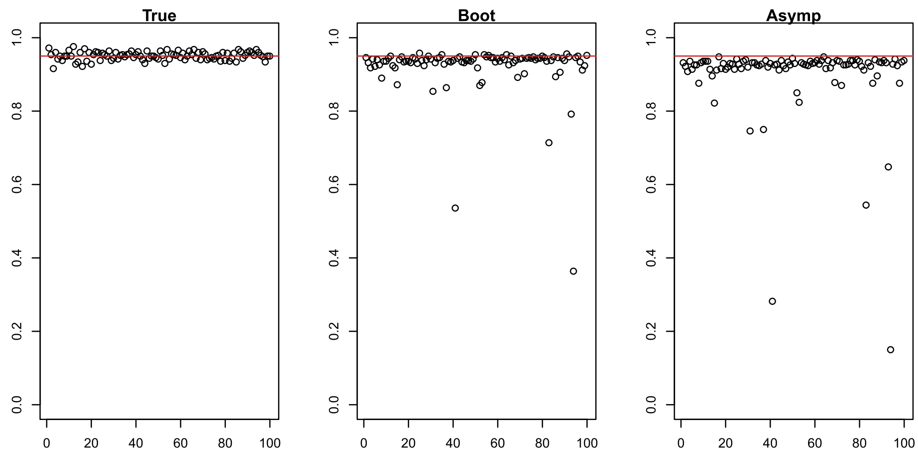

Fig 2. *Empirical coverage of the true, bootstrap and asymptotic confidence intervals from Model 1 for each* $\chi \in \mathcal{C}$ *(values* $1-\alpha = 0.95$ *and* $n = 250$ *are considered). Solid line is located at a height* $1-\alpha$.

are closer to the theoretical coverages than the corresponding to the asymptotic intervals.

Figure 2 shows a comparison of the empirical coverages of $I^{true}_{\chi,1-\alpha}$, $I^{*}_{\chi,1-\alpha}$ and $I^{asymp}_{\chi,1-\alpha}$ for each $\chi \in \mathcal{C}$. On the one hand, this figure clearly reflects the underestimation of the coverage by the asymptotic intervals, this fact being attenuated by the bootstrap ones. Therefore, at least in this example, bootstrap methodology is a nice alternative to asymptotic one. On the other hand, focusing on the empirical coverages by the bootstrap intervals, it is remarkable the presence of two confidence intervals with poor empirical coverages. Specifically, they correspond to $m(\chi_{41})$ and $m(\chi_{94})$ ($\chi_i \in \mathcal{C}$, $i = 41, 94$). In an attempt to find the reasons of those poor behaviors, Figure 1 (right panel) shows the first derivative of the curves in a training sample $\mathcal{S}$, together with the first derivative of the curves χ_{41} and χ_{94} in the test sample $\mathcal{C}$. It seems that χ_{41} and χ_{94} are atypical curves respect to $\mathcal{S}$. As attested from Figure 3 (left panel), this fact causes poor predictions for $m(\chi_{41})$ and $m(\chi_{94})$ and, therefore, poor confidence intervals.

Finally, Figure 3 (right panel) reports, for each $\chi \in \mathcal{C}$, the confidence intervals obtained by means of the bootstrap methodology (using the training sample $\mathcal{S}$ referred in the previous paragraph). True confidence intervals are also shown. Excepting the cases of the atypical curves χ_{41} and χ_{94}, bootstrap intervals are close to the true ones.

4.3. Rough curves

To provide further evidence of the interest of our methodology, we proceeded a second example dealing with rough curves. Specifically, in Model 2 the dis-

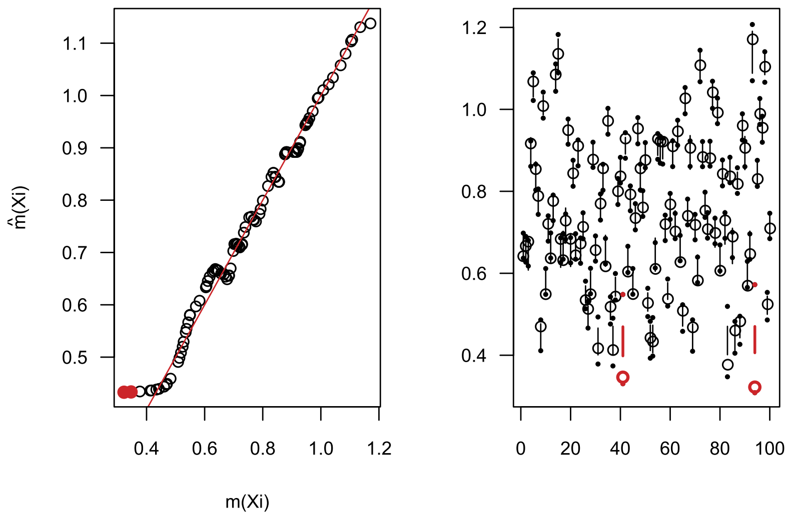

FIG 3. *Left panel: predicted values* ($\widehat{m}_h(\chi)$) *(from a training sample* $\mathcal{S}$ *from Model 1) for each* $\chi \in \mathcal{C}$ *vs true values* ($m(\chi)$)*. Full circles correspond to the atypical curves* χ_{41} *and* $\chi_{94} \in \mathcal{C}$ ($n = 250$ *was considered). Right panel: for each curve* χ *in* $\mathcal{C}$*, the vertical line represents the bootstrap confidence interval for* $m(\chi)$ *obtained from* $\mathcal{S}$*, while the dots delimit the true confidence interval* ($1 - \alpha = 0.95$ *was considered). In addition, the hollow circle locates the regression value* $m(\chi)$*. Outputs for the atypical curves are colored in red.*

cretized functional covariate was

$$\boldsymbol{\chi}_i(t_j) = b_{1i}\cos(b_{2i}t_j) + \sum_{k=1}^{j} B_{ik}/b,$$

where $b = 5$, $\{b_{1i}\}$ and $\{b_{2i}\}$ came from AR(1) and MA(1) gaussian processes with parameters $\rho_{b_1} = 0.9$ and $\theta_{b_2} = -0.5$, respectively, and variances $\sigma^2_{b_1} = \sigma^2_{b_2} = 0.1$. B_{ik} were i.i.d. realizations of $N(0, \sigma)$ with $\sigma = 0.1$ and $0 = t_1 < t_2 < \cdots < t_{99} < t_{100} = \pi$ were 100 equally spaced measurements. The regression operator was

$$m(\boldsymbol{\chi}) = \int_0^{\pi} (\boldsymbol{\chi}(t))^2 dt.$$

Figure 4 (left panel) shows some sequential curves corresponding to a functional time series generated from Model 2.

Note that Model 2 adapts the model considered in Ferraty, Van Keilegom and Vieu (2012) [18] to a setting of both scalar response and dependent curves. As recommended in that paper, the semi-metric ($d_4^{proj}(\cdot,\cdot)$) was based on the projection on the four eigenvectors, $v_1(\cdot), \ldots, v_4(\cdot)$, associated with the four largest eigenvalues of the empirical covariance operator of the functional predictor $\boldsymbol{\chi}$:

$$d_4^{proj}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j) = \sqrt{\sum_{k=1}^{4} \Big(\int_0^{\pi} (\boldsymbol{\chi}_i(t) - \boldsymbol{\chi}_j(t)) v_k(t) dt\Big)^2}. \tag{4.3}$$

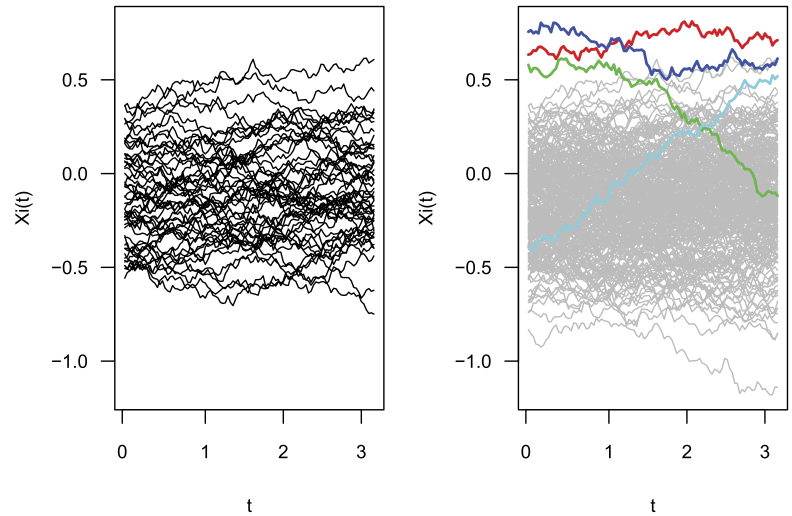

Fig 4. *Left panel: first 50 curves in a training sample $\mathcal{S}$ ($n = 250$ was considered) generated from Model 2. Right panel: curves in $\mathcal{S}$ together with the curves χ_{20}, χ_{28}, χ_{39} and χ_{75} in the test sample $\mathcal{C}$.*

Table 2
Average over $\mathcal{C}$ of the empirical coverage of the true, bootstrap and asymptotic confidence intervals from Model 2. Standard deviation appears in brackets.

$1-\alpha$	0.95		0.90	
n	100	250	100	250
Coverage (I^{true})	0.950 (0.013)	0.949 (0.011)	0.903 (0.020)	0.897 (0.015)
Coverage (I^{*})	0.804 (0.180)	0.889 (0.067)	0.774 (0.178)	0.861 (0.069)
Coverage (I^{asymp})	0.755 (0.171)	0.818 (0.056)	0.693 (0.164)	0.750 (0.056)

Table 2 reports the average over $\mathcal{C}$ of the empirical coverage of the three computed confidence intervals. The accuracy of the coverages improves as the sample size, n, increases. Coverages of both the bootstrap and the asymptotic intervals are worse than the ones obtained in the previous example of smooth curves, this fact showing the difficulties of the inference when dealing with curves with higher variability. In any case, bootstrap intervals continue to be better than the asymptotic ones (at least in this example).

Figure 5 compares the empirical coverages of $I^{true}_{\chi,1-\alpha}$, $I^{*}_{\chi,1-\alpha}$ and $I^{asymp}_{\chi,1-\alpha}$ for each $\chi \in \mathcal{C}$. The underestimation of the coverage by the asymptotic intervals is clearly shown in this figure, this fact being attenuated by the bootstrap ones (as in the case of Model 1). Therefore, at least in this example, bootstrap methodology is a nice alternative to asymptotic one. Focusing on the empirical coverages by the bootstrap intervals, it is noted, again, the presence of some (four) confidence intervals with poor empirical coverages. They correspond to $m(\chi_{20})$, $m(\chi_{28})$, $m(\chi_{39})$ and $m(\chi_{75})$ ($\chi_i \in \mathcal{C}$, $i = 20, 28, 39, 75$).

Figure 6 shows the scores of the first (left panel) and second (right panel) principal components of the curves in a training sample $\mathcal{S}$. The scores corre-

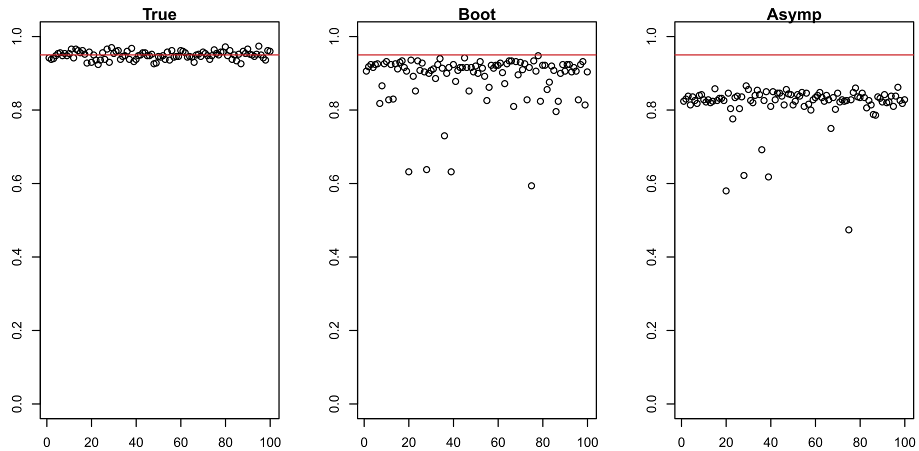

FIG 5. *Empirical coverage of the true, bootstrap and asymptotic confidence intervals from Model 2 when values* $1-\alpha = 0.95$ *and* $n = 250$ *are considered. Solid line is located at a height* $1-\alpha$.

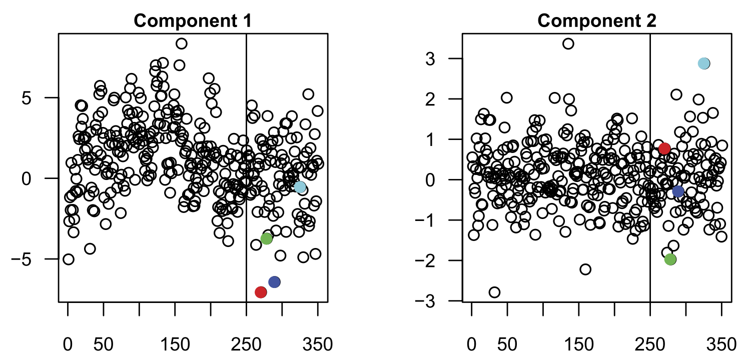

FIG 6. *Left of the vertical line: scores of the first (left panel) and second (right panel) principal component of the curves in a training sample* $\mathcal{S}$ *(sample size* $n = 250$*). Right of the vertical line: scores of the curves in the test sample* $\mathcal{S}$. *Full circles correspond to the curves* χ_{20}, χ_{28}, χ_{39} *and* $\chi_{75} \in \mathcal{C}$.

sponding to the curves in the test sample $\mathcal{C}$ are also included. This figure shows that the scores of the first principal component of χ_{20} and χ_{39} are atypical with respect to the scores of the curves in the training sample. The same occurs for the scores of the second principal component of χ_{28} and χ_{75}. Note that the atypical behavior of these four curves is supported by Figure 4 (right panel), which shows the curves in $\mathcal{S}$ together with χ_{20}, χ_{28}, χ_{39} and χ_{75}.

Figure 7 (left panel) displays the points $(m(\chi), \widehat{m}_h(\chi))$ for $\chi \in \mathcal{C}$. The expected poor estimation of $m(\chi)$ in (three of the four) atypical curves χ_i, $i = 20, 28, 39, 75$ is attested from such figure, this fact causing the poor behavior of the confidence intervals associated to those curves.

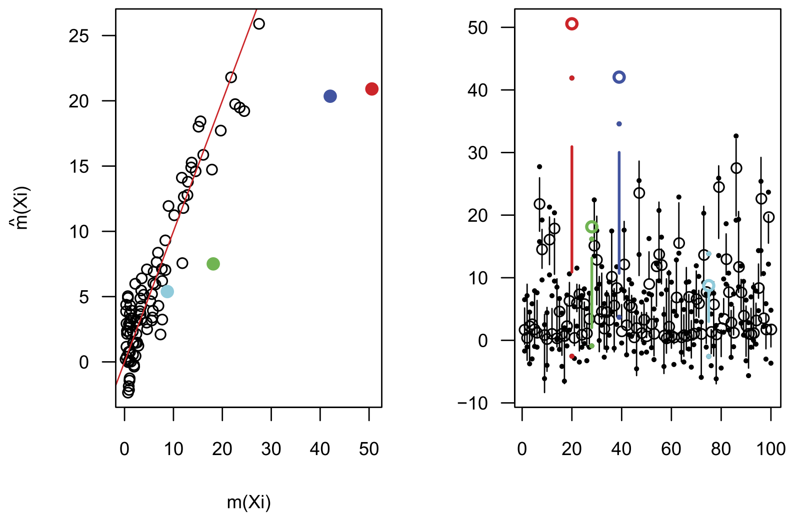

Fig 7. *Left panel: predicted values* ($\widehat{m}_h(\chi)$) *(from a training sample* $\mathcal{S}$ *from Model 2) for each* $\chi \in \mathcal{C}$ *vs observed values* ($m(\chi)$)*. Full circles correspond to the atypical curves* χ_{20}, χ_{28}, χ_{39} *and* $\chi_{75} \in \mathcal{C}$ ($n = 250$ *was considered). Right panel: for each curve* χ *in* $\mathcal{C}$*, the vertical line represents the bootstrap confidence interval obtained from* $\mathcal{S}$*, while the dots delimit the true confidence interval* ($1 - \alpha = 0.95$ *was considered). In addition, the hollow circle locates the regression value* $m(\chi)$*. Outputs in color other than black correspond to the atypical curves.*

Finally, Figure 7 (right panel) reports, for each $\chi \in \mathcal{C}$, the confidence intervals obtained by means of the bootstrap methodology (using the training sample $\mathcal{S}$ referred in a previous paragraph). True confidence intervals are also shown. Excepting the cases of the atypical curves, bootstrap intervals are close to the true ones.

5. Application to real data

Modeling and forecasting of electricity demand and price is of main interest for agents involved in the electricity markets, and the statistical/engineering literature in this field is quite abundant. See for instance the book by Weron (2006) [35] for a nice monograph on electricity demand and price forecasting, and also Suganthi and Samuel (2012) [34] and Weron (2014) [36] for reviews on electricity demand forecasting and electricity price forecasting, respectively. Despite the importance of inference through confidence intervals, most publications in this field have focused on pointwise prediction. This section applies the methodology proposed in this paper to the construction of confidence intervals for the mean hourly electricity demand in Spain given the daily curve of electricity demand in the previous day. As in the simulation study presented in Section 4, we focus on the wild bootstrap procedure. 1000 bootstrap replicates were drawn, the quadratic kernel was used and equal smoothing parameters $h = b = b_{CV}$ were considered, where b_{CV} was selected from a cross-validation method. In addition,

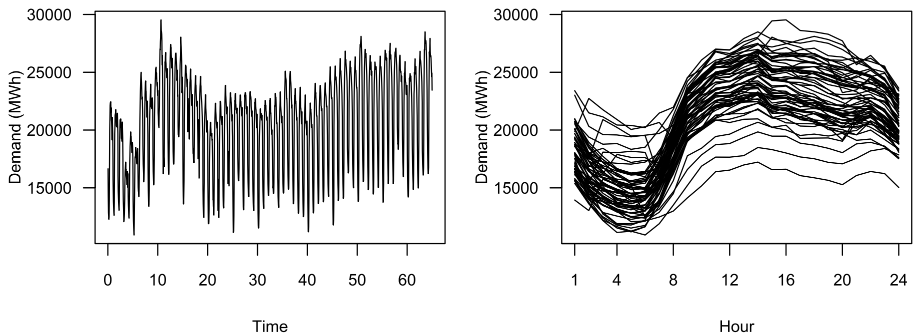

FIG 8. *Functional time series (left panel) and daily curves (right panel) for the electricity demand corresponding to the workdays during the second quarter of the year 2012.*

the class of projection-based semi-metrics $\{d_v^{proj}(\cdot,\cdot)\}_v$ (see (4.3) for the case of $v = 4$) was considered, the quantity of eigenvectors v being also chosen from cross-validation. The confidence level considered was $1-\alpha = 0.95$.

It is known that electricity demand shows vastly different patterns on workdays, public holidays and weekend. Thus, in order to accommodate this fact to our model (1.1), we have focused on workdays; in addition, to avoid (or attenuate) the effect that abrupt changes in temperature exert on electricity demand, our database was reduced to the second quarter of the year 2012. In summary, our database, $\mathcal{B}$, consists in the workdays of the second quarter of the year 2012. Each daily functional datum, χ_i, comes from the 24 hourly observations of electricity demand in Spain for each day in our database. Such curves can be seen in Figure 8, as a functional time series as well as a set of curves (our data source was OMIE, 'Operador del Mercado Ibérico de Energía', which is the Market Operator).

In the following, we present two applications. In the first one, we estimate mean hourly electricity demand for each hour in a fixed day, while in the second application we estimate for a fixed hour in different days.

In our first application, we focus on bootstrap confidence intervals for each mean hourly electricity demand corresponding to the last day in our database (Friday, June 29, 2012) given the daily curve of electricity demand in the previous day in our database. Therefore, 24 confidence intervals need to be computed. The interval corresponding to the hour $t = 1,\dots,24$ was based on the regression model

$$\boldsymbol{\chi}_{i+1}(t) = m_t^{(1)}(\boldsymbol{\chi}_i) + \varepsilon_{i,t}^{(1)} \ (i = 1,\dots,n);$$

that is, to do inference on the mean hourly electricity demand at hour t, the functional $G(\cdot)$ in model (1.1) is defined as $G_t(\boldsymbol{\chi}_{i+1}) = \boldsymbol{\chi}_{i+1}(t)$. Historical curves consisted in the days in our database, $\mathcal{B}$, previous to Friday, June 29, 2012 (fixed historical curves, not dependent on the prediction horizon t; equivalently, not dependent on the model). Figure 9 (left panel) displays the corresponding bootstrap confidence intervals. Note that the small sizes of such intervals respect

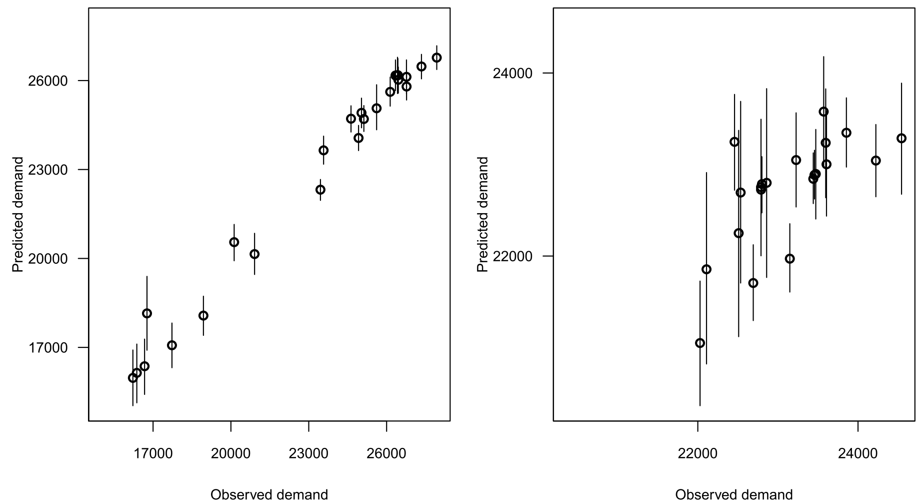

FIG 9. *Left panel: Bootstrap confidence intervals computed for the 24 hours of Friday, June 29, 2012. Right panel: Bootstrap confidence intervals computed for the workdays in June, 2012 (fixed hour: 09:00 a.m.). The confidence level was* $1 - \alpha = 0.95$.

to the big magnitude of the observed demand suggest a good accuracy of the bootstrap confidence intervals.

In order to maintain the prediction horizon (note that in the previous application 24 prediction horizons were considered), a second application was implemented. Specifically, a confidence interval was constructed for each mean electricity demand at fixed hour 9:00 a.m. corresponding to each of the $d = 1, \dots, 21$ workdays in June 2012, given the daily curve of electricity demand in the previous day in our database. Therefore, 21 confidence intervals need to be computed. In this second application, historical data consisted in the workdays included in the 61 previous days (two previous months) to the day to predict, while the modeling to obtain the confidence interval corresponding to the day d was done by means of the regression model

$$\boldsymbol{\chi}_{i+1,d}(9) = m_d^{(2)}(\boldsymbol{\chi}_{i,d}) + \varepsilon_{i,d}^{(2)} \ (i = 1, \dots, n_d);$$

that is, $G(\cdot)$ in model (1.1) was defined as $G(\boldsymbol{\chi}_{i+1}) = \boldsymbol{\chi}_{i+1}(9)$, and the historical curves were changing as d (equivalently, the model) does (in opposite at what occurred in the previous application). Figure 9 (right panel) shows the corresponding bootstrap confidence intervals. In this case, it can be observed that the ratio 'length of the interval/magnitude of the observed demand' is slightly greater than in the first application. This fact could be a consequence of decreased sample size in (roughly) a 33 percent. In any case, it seems that the accuracy of the bootstrap confidence intervals continue to be sufficiently good.

6. Concluding remarks

This paper has revisited the work by Ferraty, Van Keilegom, and Vieu (2010) [17] to extend its asymptotic results, established for independent samples, to the case of dependent ones. Based on both naive and wild bootstrap procedures, pointwise confidence intervals for the regression function in a nonparametric model with functional predictor have been built, and their asymptotic validity has been established. Examples on finite sample sizes (via simulations and applications to real data) have shown that such results are useful in practice.

Interesting challenges remain as open problems to be dealt in a future. They include the topic of bandwidths selection: based on the simulation study presented in Ferraty, Van Keilegom, and Vieu (2010) [17], equal bandwidths chosen by cross-validation were considered in our applications; despite it is out of the scope of this paper, theoretical results on bandwidth selection for the proposed methodology would contribute greatly to the statistical literature (although we are aware of the difficulty of obtaining them). In addition, researches based on bootstraping pairs (instead of on bootstrapping residuals, as was done here) would give rise to new tools in the setting of nonparametric functional data analysis. Finally, to expand the range of possible applications, it would be very interesting to obtain results under dependence conditions in the random errors of the regression model.

7. Some technical lemmas

Before proving our main result (Theorem 3.1), we state some preliminary lemmas to be used in that proof. First, let us denote

$$J_\chi = \frac{\int_0^1 tK(t)dP^{d(\boldsymbol{\chi},\chi)/h}(t)}{\int_0^1 K(t)dP^{d(\boldsymbol{\chi},\chi)/h}(t)} \text{ and } \widehat{m}_h(\chi) = \frac{\widehat{g}_h(\chi)}{\widehat{f}_h(\chi)},$$

where

$$\widehat{g}_h(\chi) = \frac{\sum_{i=1}^n K(d(\boldsymbol{\chi}_i,\chi)/h)Y_i}{nF_\chi(h)} \text{ and } \widehat{f}_h(\chi) = \frac{\sum_{i=1}^n K(d(\boldsymbol{\chi}_i,\chi)/h)}{nF_\chi(h)}.$$

Furthermore, we will use the notation

$$A_1 = -\mathbb{E}\left(\widehat{g}_h(\chi)(\widehat{f}_h(\chi) - \mathbb{E}(\widehat{f}_h(\chi)))\right) \text{ and } A_2 = \mathbb{E}\left((\widehat{f}_h(\chi) - \mathbb{E}(\widehat{f}_h(\chi))^2\widehat{m}_h(\chi)\right).$$

Lemma 7.1 (Lemma 1, Ferraty et al. 2007 [15])**.** *Under assumptions (H1), (H2), (H3), (H6) and (H7) we have that*

$$\frac{\mathbb{E}\left(\widehat{g}_h(\chi)\right)}{\mathbb{E}\left(\widehat{f}_h(\chi)\right)} - m(\chi) = h\varphi'_\chi(0)J_\chi + o(h).$$

Lemma 7.2 (Lemma 2, Ferraty et al. 2007 [15])**.** *Under assumptions (H3), (H6) and (H7) we have that*

$$J_{\chi} = \frac{K(1) - \int_0^1 (sK(s))' \tau_{h\chi}(s) ds}{K(1) - \int_0^1 K'(s) \tau_{h\chi}(s) ds} \longrightarrow \frac{M_{0\chi}}{M_{1\chi}} \text{ as } n \longrightarrow \infty.$$

Lemma 7.3 (Lemma 4, Ferraty et al., 2007 [15])**.** *Under assumptions of Lemma 7.1 we have that*

$$\mathbb{E}\left(\widehat{f}_h(\chi)\right) \longrightarrow M_{1\chi} \text{ and } \mathbb{E}\left(\widehat{g}_h(\chi)\right) \longrightarrow m(\chi) M_{1\chi} \text{ as } n \longrightarrow \infty.$$

Actually, lemmas 7.1 and 7.3 above were established in Ferraty et al. (2007) [15] under independence conditions. Noting that dependence does not influence on the expectation of $\widehat{g}_h(\chi)$ nor $\widehat{f}_h(\chi)$, one has that they remain valid in our setting of dependent samples. Of course, the validity of Lemma 2 in Ferraty et al. (2007) [15] (Lemma 7.2 above) also remains because J_χ is not random.

Lemma 7.4 (Lemma 2.5, Delsol, 2009 [11])**.** *Under assumptions (H1)–(H9) we have that*

$$Var\left(\widehat{f}_h(\chi)\right) = \frac{M_{2\chi}}{nF_\chi(h)}(1 + o(1)),$$

$$Var\left(\widehat{g}_h(\chi)\right) = (\sigma_\varepsilon^2 + m^2(\chi)) \frac{M_{2\chi}}{nF_\chi(h)}(1 + o(1))$$

and

$$Cov\left(\widehat{g}_h(\chi), \widehat{f}_h(\chi)\right) = m(\chi) \frac{M_{2\chi}}{nF_\chi(h)}(1 + o(1)).$$

Lemma 7.5. *Under assumptions of Lemma 7.4 we have that*

$$A_1 = O((nF_\chi(h))^{-1}) \text{ and } A_2 = O((nF_\chi(h))^{-1}).$$

Proof. On the one hand, we have that $A_1 = -Cov\left(\widehat{g}_h(\chi), \widehat{f}_h(\chi)\right)$; so, from Lemma 7.4 one obtains that $A_1 = O((nF_\chi(h))^{-1})$. On the other hand,

$$\begin{aligned}
|A_2| &= \left|\mathbb{E}\left((\widehat{f}_h(\chi) - \mathbb{E}(\widehat{f}_h(\chi))^2 \widehat{m}_h(\chi)\right)\right| \\
&= \left|\mathbb{E}\left(\mathbb{E}\left((\widehat{f}_h(\chi) - \mathbb{E}(\widehat{f}_h(\chi))^2 \widehat{m}_h(\chi) | \boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_n\right)\right)\right| \\
&= \left|\mathbb{E}\left((\widehat{f}_h(\chi) - \mathbb{E}(\widehat{f}_h(\chi))^2 \mathbb{E}\left(\widehat{m}_h(\chi) | \boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_n\right)\right)\right| \\
&= \left|\mathbb{E}\left((\widehat{f}_h(\chi) - \mathbb{E}(\widehat{f}_h(\chi))^2 \frac{\sum_{i=1}^n K(d(\boldsymbol{\chi}_i, \chi)/h) m(\boldsymbol{\chi}_i)}{\sum_{i=1}^n K(d(\boldsymbol{\chi}_i, \chi)/h)}\right)\right| \\
&\leq CVar\left(\widehat{f}_h(\chi)\right),
\end{aligned}$$

where the inequality is a consequence of assumptions (H1) and (H7). The proof concludes by using, again, Lemma 7.4. ☐

Lemma 7.6. *Under assumptions of Lemma 7.4 we have that*

$$\mathbb{E}\left(\widehat{m}_h(\chi)\right) - m(\chi) = \varphi'_\chi(0)\frac{M_{0\chi}}{M_{1\chi}}h + O\left(\frac{1}{nF_\chi(h)}\right) + o(h). \tag{7.1}$$

If in addition Assumption (H10) holds, then

$$Var\left(\widehat{m}_h(\chi)\right) = \frac{1}{nF_\chi(h)}\frac{M_{2\chi}}{M_{1\chi}^2}\sigma_\varepsilon^2 + o\left(\frac{1}{nF_\chi(h)}\right). \tag{7.2}$$

Proof of (7.1). In the same way as for the case of independent data (see Ferraty et al., 2007 [15]), this proof is based on the decomposition

$$\mathbb{E}\left(\widehat{m}_h(\chi)\right) = \frac{\mathbb{E}\left(\widehat{g}_h(\chi)\right)}{\mathbb{E}\left(\widehat{f}_h(\chi)\right)} + \frac{A_1}{\left(\mathbb{E}\left(\widehat{f}_h(\chi)\right)\right)^2} + \frac{A_2}{\left(\mathbb{E}\left(\widehat{f}_h(\chi)\right)\right)^2}.$$

In fact, using Lemmas 7.1, 7.2 and 7.5 above, the proof of (7.1) is easily obtained following the same steps as those in the proof of (2) in Ferraty et al. (2007) [15]. □

Proof of (7.2). See Theorem 7.3.1 in Delsol (2008) [10]. □

8. Proof of Theorem 3.1

The proof of our Theorem 3.1 follows the same steps as those of Theorem 1 in Ferraty, Van Keilegom, and Vieu (2010) [17], where the case of an independent sample $\mathcal{S}$ was dealt. Thus, for the sake of brevity, we will focus on the issues where the dependence affects. In addition, our proof covers the case of the naive bootstrap as well as the case of the wild bootstrap (actually, as can be seen in Ferraty, Van Keilegom, and Vieu, 2010 [17], there exists a difference in the proof of Theorem 3.1 under each bootstrap procedure; specifically, regarding the proof of (8.3). However, the way of to show our proof avoids the need of indicating such difference).

First, let $\mathbb{E}^{\mathcal{S}}$ and $Var^{\mathcal{S}}$ denote expectation and variance, respectively, conditionally on the sample $\mathcal{S}$, while Φ is the standard normal distribution function. Let us write

$$\begin{aligned} P^{\mathcal{S}}\left(\sqrt{nF_\chi(h)}(\widehat{m}^*_{hb}(\chi) - \widehat{m}_b(\chi)) \le y\right) - P\left(\sqrt{nF_\chi(h)}(\widehat{m}_h(\chi) - m(\chi)) \le y\right) \\ = T_1(y) + T_2(y) + T_3(y), \end{aligned} \tag{8.1}$$

where

$$\begin{aligned} T_1(y) = P^{\mathcal{S}}\left(\sqrt{nF_\chi(h)}(\widehat{m}^*_{hb}(\chi) - \widehat{m}_b(\chi)) \le y\right) - \\ \Phi\left(\frac{y - \sqrt{nF_\chi(h)}\left(\mathbb{E}^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right) - \widehat{m}_b(\chi)\right)}{\sqrt{nF_\chi(h)Var^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right)}}\right), \end{aligned}$$

$$T_2(y) = \Phi\left(\frac{y - \sqrt{nF_\chi(h)}\left(\mathbb{E}^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right) - \widehat{m}_b(\chi)\right)}{\sqrt{nF_\chi(h)Var^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right)}}\right) - \Phi\left(\frac{y - \sqrt{nF_\chi(h)}\left(\mathbb{E}\left(\widehat{m}_h(\chi)\right) - m(\chi)\right)}{\sqrt{nF_\chi(h)Var\left(\widehat{m}_h(\chi)\right)}}\right)$$

and

$$T_3(y) = \Phi\left(\frac{y - \sqrt{nF_\chi(h)}\left(\mathbb{E}\left(\widehat{m}_h(\chi)\right) - m(\chi)\right)}{\sqrt{nF_\chi(h)Var\left(\widehat{m}_h(\chi)\right)}}\right) - P\left(\sqrt{nF_\chi(h)}(\widehat{m}_h(\chi) - m(\chi)) \leq y\right).$$

On the one hand, from Theorem 2.7 in Delsol (2009) [11] together with Lemma 7.6 and Assumption (H6), we obtain that

$$T_3(y) \longrightarrow 0 \text{ a.s. for any fixed value of } y. \tag{8.2}$$

On the other hand, in order to obtain the same convergence for $T_1(y)$, it is sufficient to prove that

$$\frac{\widehat{m}^*_{hb}(\chi) - \mathbb{E}^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right)}{\sqrt{Var^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right)}} \stackrel{d}{\longrightarrow} N(0,1), \text{ a.s., conditionally on the sample } \mathcal{S}.$$

For that, following the same steps as in the proof of the second statement in Lemma 2 in Ferraty, Van Keilegom, and Vieu (2010) [17] (established in the independent case), it is sufficient to prove that

$$\frac{Var^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right)}{Var\left(\widehat{m}_h(\chi)\right)} \longrightarrow 1 \text{ a.s.} \tag{8.3}$$

As can be seen in Ferraty, Van Keilegom, and Vieu (2010) [17], when data in $\mathcal{S}$ are independent, the proof of (8.3) is based on both the type of bootstrap procedure used and Lemmas 4 and 5 and Theorem 1 in Ferraty et al. (2007) [15]. On the one hand, because the random errors ε_i in our model (2.1) are independent, we are considering the same bootstrap procedures as in Ferraty et al. (2007) [15]. On the other hand, our Lemmas 7.3, 7.4 and 7.6 give the same results as Lemmas 4 and 5 and Theorem 1 in Ferraty et al. (2007) [15], respectively, but under dependence conditions on $\mathcal{S}$. These facts allow to follow step by step the proof of (8.3) given in Ferraty et al. (2007) [15] (for the independent case), and to conclude that (8.3) holds under our dependence conditions on $\mathcal{S}$. As a consequence, we have obtained that

$$T_1(y) \longrightarrow 0 \text{ a.s. for any fixed value of } y. \tag{8.4}$$

Now, from (8.2), (8.4) and Polya's theorem (see, e.g., Serfling, 1980, p. 18 [33]) together with the continuity of the function Φ, we have that

$$\sup_{y\in\mathbb{R}} |T_1(y)| + \sup_{y\in\mathbb{R}} |T_3(y)| \to 0 \text{ a.s.} \tag{8.5}$$

Finally, we are going to study the term $T_2(y)$. Using the fact that, for any $a \in \mathbb{R}$ and $c > 0$,

$$\sup_{y\in\mathbb{R}} |\Phi(a+cy) - \Phi(y)| \leq |a| + \max\{c, c^{-1}\} - 1,$$

and considering

$$a = \frac{\sqrt{nF_\chi(h)}\left(\mathbb{E}\left(\widehat{m}_h(\chi)\right) - m(\chi) - \mathbb{E}^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right) + \widehat{m}_b(\chi)\right)}{\sqrt{nF_\chi(h)Var^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right)}}$$

and

$$c = \sqrt{\frac{Var\left(\widehat{m}_h(\chi)\right)}{Var^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right)}},$$

we have that

$$\begin{aligned}\sup_{y\in\mathbb{R}}|T_2(y)| \quad &\leq \quad \left|\frac{\sqrt{nF_\chi(h)}\left(\mathbb{E}\left(\widehat{m}_h(\chi)\right) - m(\chi) - \mathbb{E}^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right) + \widehat{m}_b(\chi)\right)}{\sqrt{nF_\chi(h)Var^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right)}}\right| \\ &\quad + \max\left\{\sqrt{\frac{Var\left(\widehat{m}_h(\chi)\right)}{Var^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right)}}, \sqrt{\frac{Var^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right)}{Var\left(\widehat{m}_h(\chi)\right)}}\right\} - 1. \qquad (8.6)\end{aligned}$$

From (8.6), and taking into account (8.3) together with the statement (7.2) in Lemma 7.6, we have that, in order to obtain the following convergence:

$$\sup_{y\in\mathbb{R}} |T_2(y)| \to 0 \text{ a.s.}, \qquad (8.7)$$

it is sufficient to prove that

$$\left|\sqrt{nF_\chi(h)}\left(\mathbb{E}\left(\widehat{m}_h(\chi)\right) - m(\chi) - \mathbb{E}^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right) + \widehat{m}_b(\chi)\right)\right| \to 0 \text{ a.s.} \qquad (8.8)$$

For that, we will follow the same steps as those used in Ferraty, Van Keilegom, and Vieu (2010) [17] to establish (8.8) under independence conditions (see Lemma 4 in Ferraty, Van Keilegom, and Vieu, 2010 [17]). Specifically, let us consider the decomposition

$$\mathbb{E}^{\mathcal{S}}\left(\widehat{m}^*_{hb}(\chi)\right) - \widehat{m}_b(\chi) = U_1 + U_2 + U_3, \qquad (8.9)$$

where

$$U_1 = \frac{(nF_\chi(h))^{-1}}{\widehat{f}_h(\chi)} \sum_{i=1}^{n} \left(\widehat{m}_b(\boldsymbol{\chi}_i) - \widehat{m}_b(\chi) - \mathbb{E}\left(\widehat{m}_b(\boldsymbol{\chi}_i)\right) + \mathbb{E}\left(\widehat{m}_b(\chi)\right)\right) K(d(\boldsymbol{\chi}_i, \chi)/h),$$

$$U_2 = \frac{(nF_\chi(h))^{-1}}{\widehat{f}_h(\chi)} \sum_{i=1}^{n} \left(\mathbb{E}\left(\widehat{m}_b(\boldsymbol{\chi}_i)\right) - \mathbb{E}\left(\widehat{m}_b(\chi)\right) - m(\boldsymbol{\chi}_i) + m(\chi)\right) K(d(\boldsymbol{\chi}_i, \chi)/h)$$

and

$$U_3 = \frac{(nF_\chi(h))^{-1}}{\widehat{f}_h(\chi)} \sum_{i=1}^{n} (m(\boldsymbol{\chi}_i) - m(\chi))\, K(d(\boldsymbol{\chi}_i, \chi)/h).$$

Using our Lemma 7.6 instead of Theorem 1 in Ferraty et al. (2007) [15], we can follow the lines of the proof of Lemma 5 in Ferraty, Van Keilegom, and Vieu (2010) [17] to obtain that

$$U_1 = o\left((nF_\chi(h))\right)^{-1/2} \text{ a.s.} \tag{8.10}$$

(remember that our Lemma 7.6 extends the Theorem 1 in Ferraty et al., 2007 [15], to the case of dependent data). The term U_2 can be studied in a similar way as in Lemma 6 in Ferraty, Van Keilegom, and Vieu (2010) [17], where the key tools were the Lemma 3 in Ferraty et al. (2007) [15] and a Bernstein's inequality for independent variables. Specifically, all what we must to do to obtain that

$$U_2 = o\left((nF_\chi(h))\right)^{-1/2} \text{ a.s.} \tag{8.11}$$

is to follow the same steps used to prove Lemma 6 in Ferraty, Van Keilegom, and Vieu (2010) [17], but using our Lemma 7.5 instead of Lemma 3 in Ferraty et al. (2007) [15] and, in addition, considering the Lemma 3 in Aneiros-Pérez and Vieu (2008) [1] instead of the mentioned Bernstein's inequality. On the one hand, our Lemma 7.5 allows to maintain in our dependent case the different decompositions used in the proof of Lemma 6 in Ferraty, Van Keilegom, and Vieu (2010) [17]. On the other hand, Lemma 3 in Aneiros-Pérez and Vieu (2008) [1] allows to obtain the orders of the terms of such decompositions that in Ferraty, Van Keilegom, and Vieu (2010) [17] were obtained by means of a Bernstein's inequality for independent variables. For instance, in the case of the term

$$V_1 = \max_{1 \le k \le r_n} \left| \sum_{i=1}^{n} n^{-1}(Z_{ik} - \mathbb{E}(Z_{ik})) \right|$$

where $Z_{ik} = F_\chi(b)^{-1} Y_i \left(K(d(\boldsymbol{\chi}_i, \chi)/h) - K(d(\boldsymbol{\chi}_i, \chi_{kn})/h)\right)$ with $\mathbb{E}(|Z_{ik}|^p) = O\left(\left(F_\chi(b)^{-1}h/b\right)^{p-1}\right)$ (for details, see the proof of Lemma 6 in Ferraty, Van Keilegom, and Vieu, 2010 [17]), from Lemma 3 in Aneiros-Pérez and Vieu (2008) [1] one obtains that

$$V_1 = O\left(n^{-1/2+1/p} \log n\right) = o\left((nF_\chi(h))^{-1/2}\right) \text{ a.s.}$$

(Note that to check that the assumptions in Lemma 3 in Aneiros-Pérez and Vieu, 2008 [1], holds we have used that $F_\chi(b)^{-1}h/b < C$, while the last equality above is a consequence of the fact that $n^{1/p}F(h)^{1/2} \log n = o(1)$) Finally, by means of similar techniques as those used in Ferraty and Vieu (2004) [19] to obtain the rate of convergence of $\widehat{m}_h(\chi)$, it is easy to show that

$$U_3 = \mathbb{E}(\widehat{m}_h(\chi)) - m(\chi) + o\left((nF(h))\right)^{-1/2} \text{ a.s.} \tag{8.12}$$

Now, from (8.9)–(8.12), we obtain that (8.8) holds. Finally, (8.1), (8.5) and (8.7) conclude the proof. □

Acknowledgements

The authors greatly appreciate the helpful suggestions coming from an Associate Editor and two anonymous referees. This research is partly supported by Grants MTM2014-52876-R and CN2012/130 from Spanish Ministerio de Economía y Competitividad and Xunta de Galicia, respectively.

References

[1] Aneiros-Pérez, G. and Vieu, P. (2008). Nonparametric time series prediction: A semi-functional partial linear modeling. *J Multivariate Anal*, **99** 834–857. MR2405094

[2] Aneiros, G., Vilar, J.M., Cao, R., and Muñoz-San-Roque, A. (2013). Functional prediction for the residual demand in electricity spot markets. *IEEE Trans Power Syst*, **28** 4201–4208.

[3] Antoch, J., Prchal, L., De Rosa, M.R., and Sarda, P. (2010). Electricity comsuption prediction with functional linear regression using spline estimators. *J Appl Statist*, **37** 2027–2041. MR2740138

[4] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction* (2nd edition). Lecture Notes in Statistics, 110. Springer-Verlag. MR1640691

[5] Bosq, D. (2000). *Linear Processes in Function Spaces. Theory and Applications*. Lecture Notes in Statistics, 149. Springer-Verlag. MR1783138

[6] Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: forecasting and control*. Prentice Hall. MR0436499

[7] Cao, R. (1991). Rate of convergence for the wild bootstrap in nonparametric regression. *Annals of Statistics*, **19** 2226–2231. MR1135172

[8] Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. (French) [Almost complete convergence properties of kernel predictors] *Z. Wahrsch. Verw. Gebiete*, **66(3)** 441–460. MR0751581

[9] Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *J Statist Plann Inference*, **147** 1–23. MR3151843

[10] Delsol, L. (2008). *Régression sur Variable Fonctionnelle: Estimation, Tests de Structure et Applications*. PhD thesis.

[11] Delsol, L. (2009). Advances on asymptotic normality in non-parametric functional time series analysis. *Statistics*, **43** 13–33. MR2499359

[12] Fernández-de-Castro, B., Guillas, S., and González-Manteiga, W. (2005). Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics*, **47** 212–222. MR2188081

[13] Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, **140** 335–352. MR2558367

[14] Ferraty, F., Laksaci, A., and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statistical Inference for Stochastic Processes*, **9** 47–76. MR2224840

[15] Ferraty, F., Mas, A., and Vieu, P. (2007). Nonparametric regression on functional data: inference and practical apects. *Aust. N. Z. J. Stat.*, **49** 267–286. MR2396496

[16] Ferraty, F., Rabhi, A., and Vieu, P. (2005). Conditional quantiles for dependent functional data with application to the climatic El Niño phenomenon. *Sankhya*, **67(2)** 378–398. MR2208895

[17] Ferraty, F., Van Keilegom, I., and Vieu, P. (2010). On the validity of the bootstrap in non-parametric functional regression. *Scand. J. Statist.*, **37** 286–306. MR2682301

[18] Ferraty, F., Van Keilegom, I., and Vieu, P. (2012). Regression when both response and predictor are functions. *J Multivariate Anal*, **109** 10–28. MR2922850

[19] Ferraty, F. and Vieu P. (2004). Nonparametric models for functional data, with application in regression, time–series prediction and curve discrimination. *Nonparameric Statistics*, **16**: 111–125. MR2053065

[20] Ferraty, F. and Vieu P. (2006). *Nonparametric Functional Data Analysis.* Springer Series in Statistics. Springer-Verlag, New York. MR2229687

[21] Freedman, D.A. (1981). Bootstrapping regression models. *Annals of Statistics*, **9** 1218–1228. MR0630104

[22] Goia, A. and Vieu, P. (2016). An introduction to recent advances in high/infinite dimensional Statistics. *J. Multivariate Anal*, **146** 1–6. MR3477644

[23] Györfi, L., Härdle, W., Sarda, P., and Vieu, P. (1989). *Nonparametric curve estimation from time series.* Lecture Notes in Statistics, 60. Springer-Verlag, Berlin. MR1027837

[24] González-Manteiga, W. and Martínez-Calvo, A. (2011). Bootstrap in functional linear regression. *J. Statist. Plann. Inference*, **141** 453–461. MR2719509

[25] Hall, P. (1992). On bootstrap confidence intervals in nonparametric regression. *Annals of Statistics*, **20** 695–711. MR1165588

[26] Härdle, W. and Marron, J.S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Annals of Statistics*, **19** 778–796. MR1105844

[27] Horváth L. and Kokoszka P. (2012). *Inference for Functional Data with Applications* (2nd ed.). Springer. MR2920735

[28] Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators.* Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester. MR3379106

[29] Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics*, **21** 255–285. MR1212176

[30] Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Process. Appl.*, **115** 155–177. MR2105373

[31] Pesin, Y.B. (1993). On rigourous mathematicak definitions of correlation dimension and generalized spectrum for dimensions. *Journal of Statistical Physic*, **71** 529–547. MR1219021

[32] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis.* Springer-Verlag, New York. MR2168993

[33] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

[34] Suganthi, L. and Samuel, A.A. (2012). Energy models for demand forecasting – A review. *Renew Sustain Energy Rev*, **16** 1223–1240.

[35] Weron, R. (2006). *Modeling and Forecasting Electricity Loads and Prices. A Statistical Approach.* Wiley.

[36] Weron, R. (2014). Electricity price forecasting: a review of the state-of-the-art with a look into the future. *Int J Forecast*, **30** 1030–1081.